Quantitative Assessments from the Clinical Examination

How Should Clinicians Integrate the Numerous Results?

Donald R. Holleman, Jr., MD, David L. Simel, MD, MHS

OBJECTIVE: To describe strategies for using multiple clinical examination items to estimate disease probabilities; and to evaluate the diagnostic accuracy of each strategy.

DESIGN: Prospective observational study.

SETTING: Medical preoperative evaluation clinic at a university-affiliated Veterans Affairs Medical Center.

PATIENTS: Previously reported consecutive series of patients referred for outpatient medical preoperative risk assessment.

MEASUREMENTS AND MAIN RESULTS: Pulmonary clinical examination and spirometry were the measurements. A strategy of using likelihood ratios (LRs) from seven clinical examination items was least accurate ($p < .0001$). Three alternative strategies were equivalent in diagnostic accuracy ($p \geq .2$): (1) using the single best clinical examination item and its LR, (2) using the LRs from three clinical examination items chosen by logistic regression, and (3) using the adjusted LRs chosen in strategy 2. When compared with using LRs from all seven items, the strategies of using three LRs chosen by logistic regression or using adjusted likelihood ratios better discriminated patients with airflow limitation from those without (receiver operating characteristic [ROC] areas 0.79 vs 0.69; $p = .02$). Using the single best clinical finding did not statistically degrade the clinical examination's discriminative ability (ROC areas 0.79 vs 0.75; $p = .20$).

CONCLUSIONS: Describing the rational clinical examination requires evaluating conditional independence of examination components. Conditional independence assumptions were violated when seven clinical examination items were used to estimate posterior probability of airflow limitation. Focusing on clinical examination items identified through logistic models overcame violations of independence; further statistical adjustment did not improve diagnostic accuracy. Clinicians can use the single most predictive clinical examination finding to avoid inaccuracy from violating the independence assumption.

KEY WORDS: quantitative assessment; clinical examination; logistic models; estimating probability of disease.

J GEN INTERN MED 1997;12:165-171.

The clinical examination comprises the medical history and the physical examination. Each component of the clinical examination can be considered a separate diagnostic test. As such, clinicians must choose how to incorporate the numerous results obtained from eliciting the patient's history and performing a physical examination into their clinical decision making. Some of these results will be highly relevant, whereas others ought to have little impact on clinical decisions. This process of determining the relevance of clinical examination results is two-fold: first, is the result statistically associated with the target disorder; and second, how much does the result

contribute to the diagnosis above other clinical examination results? To make these decisions about relevance, clinicians need a summary measure of the impact of a given clinical examination result on the likelihood of the target disorder.

Likelihood ratios (LRs) are convenient summary measures of diagnostic test performance. They describe the prevalence of a diagnostic test result in patients with a condition of interest versus the prevalence of the same test result among people without the condition.¹ Whether the test yields dichotomous, ordinal, or continuous results, LRs convey a similar type of diagnostic information. Clinicians can use LRs to modify their estimated prior odds of a target disorder to arrive at the posterior odds of the disorder. When there is a single diagnostic test, posterior odds is simply the product of prior odds and the LR for the test result.^{1,2} When there are multiple diagnostic tests, the posterior odds calculated from the first test's LR becomes the prior odds for the second test, and so on.^{1,2} However, this serial multiplication of LRs assumes that the tests are conditionally independent,^{1,2} that is, that the probability of the outcome of one test is not affected by the outcome of any of the other tests. If two or more of the tests are not conditionally independent, then the estimate of posterior odds may be inaccurate.¹⁻³ As multiple testing occurs frequently in clinical diagnosis, clinicians need to know how to use LRs from multiple tests to efficiently arrive at an accurate estimate of posterior odds.

When reading original research articles, review articles, or textbooks, clinicians may collect a long list of clinical examination LRs. Using these results in clinical encounters may improve a clinician's ability to accurately diagnose. Clinicians have several options when choosing among LRs to use in a clinical encounter. They can use LRs from all pertinent clinical examination items, LRs

Received from the Medical Service, Lexington Veterans Affairs Medical Center and Division of General Internal Medicine, University of Kentucky, Lexington (DRH), and Center for Health Services Research in Primary Care, Durham Veterans Affairs Medical Center, and Department of Medicine and Center for Health Care Policy Research and Education, Duke University, Durham, NC (DLS).

Presented in part at the 15th annual meeting, Society of General Internal Medicine, Washington, DC, April 29-May 1, 1992.

This material is based on work supported by the Department of Veterans Affairs and funded in part by the A. W. Mellon Foundation.

Address correspondence and reprint requests to Dr. Holleman: Medical Service (111K), Lexington VAMC, 2250 Leestown Rd., Lexington, KY 40511.

from a single clinical examination item, or LRs from a subset of findings. If they choose to use LRs from all findings, the calculations may become complex and the risk of violating the independence assumption is greatest. If they choose only a single finding, then there is no risk of violating independence assumptions, but important diagnostic information contained in other variables may be lost. Using a subset of findings that minimizes the violations of independence assumptions and uses all important diagnostic information solves the problems of the other options, but requires knowing which items are conditionally independent predictors of the target disorder.

Bias in posterior probability estimates from Bayes' theorem has been explored with regard to literature versus locally derived data,⁴ verification bias,⁵ and lack of conditional independence.⁶⁻¹⁶ Bayes' theorem can work well when independence is assumed,^{6,7,12,15,16} but violations of the independence assumption can degrade diagnostic accuracy.⁷⁻¹⁰ Proposed solutions include rank-ordering variables according to degree of intercorrelation,⁸ selecting variables based on whether they increase diagnostic accuracy,^{13,14} or using discriminant function analysis.¹⁰ However, clinicians are urged to use LRs published in medical literature, particularly with regard to the clinical examination, to calculate disease probabilities at the bedside.¹¹ We used data from a previously reported study to highlight four strategies of applying multiple LRs, each reflecting options potentially available to clinicians. These four strategies sequentially investigate the conditional independence assumption for components of the clinical examination, using the example of chronic airflow limitation. We illustrate how these strategies can be generalized to medical history and physical examination data to facilitate selecting the most computationally efficient and sensible combination of clinical diagnostic findings for clinicians to use when examining their patients. These strategies can also help clinicians to understand how erroneous conditional independence assumptions might affect their estimates of disease probabilities.

METHODS

Calculating Likelihood Ratios for the Clinical Examination of Airflow Limitation

Consecutive patients ($n = 164$) at a medical preoperative evaluation clinic of the Durham Veterans Affairs Medical Center were independently evaluated by an internist and an anesthesiologist.¹⁷ Each clinician recorded a focused pulmonary history and physical examination on a standard data collection form containing seven history and five physical examination items.¹⁷ Following the clinician's evaluations, each patient performed reference standard spirometry administered by a trained respiratory technician blinded to the results of the clinical examination. Details of our methods and the patient population were reported previously.¹⁷ We defined airflow limitation

as a ratio of forced expiratory volume in 1 second to forced vital capacity ratio (FEV_1/FVC) less than the fifth percentile predicted for gender, age, and height,¹⁸ according to American Thoracic Society Guidelines.¹⁹

We calculated LRs from contingency tables (traditional LRs) for each of the 12 history and physical examination items. As LRs of 1 are diagnostically indeterminate,²⁰ a history or physical examination item was considered useful only if the 95% confidence interval (CI)²¹ about its LRs excluded 1. Useful items from this bivariate analysis (Table 2) were used as predictor variables in a binary logistic regression model. Logistic regression was used to determine which history and physical examination items independently predicted airflow limitation. Logistic regression was performed on SAS software (SAS Institute, Cary, NC).

Strategies of Applying Multiple Likelihood Ratios

Strategy 1: All Seven Traditional Likelihood Ratios

The first strategy was to use all clinical history and physical examination features found useful by univariate analysis. The traditional LRs are calculated in the conventional manner:

$$LR^{\text{result}} = \frac{P(\text{Result} \mid \text{Disease present})}{P(\text{Result} \mid \text{Disease absent})}, \quad (1)$$

that is, the probability of the result given disease present divided by the probability of the same result given disease absent. For the case of a positive test result, equation 1 can be rewritten:

$$LR^+ = \frac{P(\text{Test} + \mid \text{Disease present})}{P(\text{Test} + \mid \text{Disease absent})}.$$

In this strategy, multiplying the prior odds by the appropriate LR for every diagnostic test result yields the posterior odds. For the case of k diagnostic tests;

$$\text{Posterior odds} = (\text{Prior odds})(LR_1)(LR_2) \dots (LR_{k-1})(LR_k). \quad (2)$$

For example, if we have three tests with test A positive, test B negative, and test C positive, then the posterior odds would be calculated by equation 2:

$$\text{Posterior odds} = (\text{Prior odds})(LR_A^+)(LR_B^-)(LR_C^+).$$

Although using all available diagnostic information seems advantageous, as the number of possible test outcomes increases, the likelihood of violating the independence assumption also increases. If all tests are conditionally independent, then this strategy would be analogous to using a logistic regression model with all univariate predictors included in the model as main effects.

Strategy 2: The Single Best Likelihood Ratio

The second strategy was to use only the single clinical history or physical examination feature that has the greatest effect on prior odds. One sure way to avoid violating the independence assumption is to use only a single LR when calculating posterior odds. Although one could choose a single examination item to use in every instance, no single item is always the best choice. Therefore, we recommend a strategy of choosing the single best LR for each patient being examined. In this strategy, all possible traditional LRs (equation 1) are ranked in order of decreasing ability to alter prior odds. Because all diagnostic tests have at least two possible outcomes (e.g., positive and negative), k tests will yield at least $2k$ LRs with at least k LRs ≤ 1 and at least k LRs ≥ 1 . Likelihood ratios that increase the odds of a target disorder range from 1 (diagnostically unhelpful) to ∞ (absolutely rules in disease). Likelihood ratios that decrease the odds of a target disorder range from 1 (diagnostically unhelpful) to 0 (absolutely rules out disease). Therefore, LRs are increasingly potent in altering prior odds as they diverge from unity in either direction. To rank order LRs < 1 with LRs > 1 , they must first be transformed to the same scale. This is most easily done by calculating an LR' where LR' is the absolute value of the natural logarithm of LR. Likelihood ratios are then ranked from largest to smallest. Alternatively, the reciprocal of LRs < 1 can be ranked with LRs > 1 . Once every LR is ranked, the single best LR can be chosen for any patient by identifying the patient's finding appearing highest on the list.

For example, we have three tests with binary outcomes (positive or negative) and test A has an $LR^+ = 15$ and $LR^- = 0.9$, test B has $LR^+ = 5$ and $LR^- = 0.6$, and test C has $LR^+ = 2$ and $LR^- = 0.1$. These outcomes can be ranked:

Rank	Finding	LR	$ \ln(LR) $
1	A present	15.0	2.71
2	C absent	0.1	2.30
3	B present	5.0	1.61
4	C present	2.0	0.69
5	B absent	0.6	0.51
6	A absent	0.9	0.11

For any given patient we would record the results from tests A, B, and C. If finding A were present, we would use $LR = 15$ as this LR has the greatest impact on prior odds. If finding A were absent, we would then look at the result of test C. If finding C were absent, then we would use $LR = 0.1$ as this LR ranks second in its ability to alter prior odds. This process continues until the patient's $LR_{\text{single best}}$ is determined. Of course, if a single examination item has the largest LR^+ and the smallest LR^- , then that item would always be chosen.

In this strategy, multiplying prior odds by the chosen $LR_{\text{single best}}$ yields posterior odds.

$$\text{Posterior odds} = (\text{Prior odds})(LR_{\text{single best}}) \quad (3)$$

This strategy does not require the independence assumption, and thus may be preferred to other strategies. However, because it does not take advantage of all available diagnostic information, the single best LR strategy may lose diagnostic accuracy. This strategy differs from using a single variable logistic regression model in that it does not always use the same test for each patient. Also, this strategy should not be interpreted as resulting in a more time-efficient clinical examination. Most, if not all examination items will need to be performed even though only a single finding is used to calculate posterior odds.

Strategy 3: Logistic Regression-Selected Traditional Likelihood Ratios

The third strategy was to use only the medical history and physical examination items chosen by logistic regression analysis. For this strategy, logistic regression is used solely as a variable reduction technique. The diagnostic tests used in strategy 1 that do not contribute additional diagnostic information are eliminated from the model. For k diagnostic tests, logistic regression will choose some j tests as independent predictors of the outcome of interest where $j \leq k$. Having selected the j independent predictors, j traditional LRs (equation 1) are then used to calculate posterior odds using equation 2:

$$\text{Posterior odds} = (\text{Prior odds})(LR_1)(LR_2) \dots (LR_{j-1})(LR_j).$$

By eliminating redundant diagnostic tests, this strategy reduces the risk and potential error of violating the independence assumption. However, this strategy still requires the assumption of independence because the j traditional LRs have not been corrected for interdependence. When all variables chosen by the logistic regression model are conditionally independent, this strategy will give the same results as the logistic regression model.

Strategy 4: Adjusted Likelihood Ratios

The fourth strategy was to use only the medical history and physical examination items chosen by logistic regression analysis and adjust their LRs to conform to the independence assumption. Although we used logistic regression to select variables in strategy 3, the utility of logistic regression can be exploited further to yield adjusted LRs. The logistic regression equation can be used to calculate adjusted LRs for different permutations of the j diagnostic tests appearing in the model derived in strategy 3, providing that data were collected from a series of unselected patients. First, we derive the logistic regression equation and record the β coefficient for each diagnostic test. The sum of the products of the patient's results for each diagnostic test (X) with its regression coefficient (β) yields the natural logarithm of the posterior odds:

$$\ln(\text{Posterior odds}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{j-1} X_{j-1} + \beta_j X_j. \quad (4)$$

The patient's posterior odds of disease can be calculated:

$$\text{Posterior odds} = e^{\ln(\text{Posterior odds})} \quad (5)$$

This simplified model can be used if it fits the data well—that is, if the predicted and observed probabilities agree. We recommend dividing the data into quantiles of predicted probability and comparing the mean predicted probability to the observed probability within each quantile. This comparison can be made with a Hosmer-Lemeshow χ^2 goodness-of-fit test.²² If the observed and predicted probabilities do not agree, then interaction terms must be included in the model to achieve an accurate estimate of the adjusted LRs.

The LR for any combination of test results can then be determined by solving equation 2 for LR:

$$\text{LR} = \frac{\text{Posterior odds}}{\text{Prior odds}}, \quad (6)$$

where prior odds is estimated from the sample prevalence and posterior odds is the odds of the disorder in patients with a defined combination of test results.

For example, we consider three diagnostic tests (A, B, and C) with binary outcomes ($X_i = 1$ when finding present, $X_i = 0$ when finding absent). The logistic model is defined according to equation 4:

$$\ln(\text{Posterior odds}) = \beta_0 + \beta_A X_A + \beta_B X_B + \beta_C X_C.$$

Suppose $\beta_0 = -5$, $\beta_A = 2.8$, $\beta_B = 2.1$, and $\beta_C = 3$, then:

$$\ln(\text{Posterior odds}) = -5 + 2.8X_A + 2.1X_B + 3X_C$$

If we find A present, B absent, and C present, then

$$\ln(\text{Posterior odds}) = -5 + 2.8 + 0 + 3 = 0.8,$$

and from equation 5,

$$\text{Posterior odds}_{(A^+ B^- C^+)} = e^{0.8} = 2.2.$$

If the prior odds in the population were 0.5, then from equation 6:

$$\text{LR}_{(A^+ B^- C^+)} = \frac{2.2}{0.5} = 4.4.$$

Thus, we do not obtain an LR for finding A, but an LR for a given permutation of results for the cluster of findings A, B, and C.

This strategy has the advantages of meeting the independence assumption and of using all useful test information. However, the complexity of this method increases as the number of variables and the number of possible outcomes for each variable increases. The number of ad-

justed LRs for any group of j tests and m possible results for each test is determined by

$$\text{Number of adjusted LRs} = \prod_{i=1}^j m_i. \quad (7)$$

For each added test, the number of LRs doubles if the test has a dichotomous outcome, triples if it has a trichotomous outcome, and so forth. Therefore, this strategy may be too complex for routine clinical use except in diagnostic strategies with few possible diagnostic test results. Because this strategy yields the same results as the logistic regression model, the regression equation may be the preferred method of calculating a posterior probability when the number of variables is large.

Comparing the Four Strategies in Airflow Limitation

We calculated LRs for clinical examination items for airflow limitation using each of the four strategies outlined above and used those LRs to calculate posterior odds of airflow limitation in an external validation sample of 70 patients. This validation sample comprised a convenience sample of patients attending medical preoperative evaluation clinic or general internal medicine clinics. Each of the 70 patients in the validation series had four posterior odds estimates corresponding to the four strategies outlined above. The predicted posterior odds were converted to posterior probabilities, and the four strategies were compared in their ability to discriminate between patients with versus those without airflow limitation, and in their accuracy at predicting observed probabilities of airflow limitation.

The ability of each strategy to discriminate between patients with and those without airflow limitation was compared using the area under the ROC curve.²³ The area under the ROC curve is the probability that a patient with airflow limitation will have a greater predicted probability of airflow limitation than a patient without this disorder. Therefore, the greater the ROC curve area, the greater the ability of the strategy to discriminate between patients with airflow limitation and those without. ROC curve areas were compared using methods accounting for repeated testing on the same patients.²⁴

We compared the accuracy of each strategy with the Hosmer-Lemeshow test,²² a χ^2 goodness-of-fit test that contrasted the number of predicted cases of airflow limitation with the number of observed cases. Predicted probabilities of airflow limitation were generated for each patient in the validation set using each of the four strategies. The validation set was divided into quantiles of predicted probability to obtain g groups of at least 20 patients per group. This was performed for each strategy. Within each of the groups the mean predicted probability was multiplied by the number of patients in the group to arrive at the expected number of cases of airflow limitation for the

Table 1. Patient Characteristics*

Characteristic	Study Sample	Validation Sample
Male, n (%)	164 (100)	70 (100)
White, n (%)	132 (81)	59 (84)
Black, n (%)	30 (18)	11 (16)
Airflow limitation, n (%)	72 (44)	28 (40)
Age, mean years ($\pm$ SD)	66 ($\pm$ 7.8)	65 ($\pm$ 7.8)
Years smoked, mean ($\pm$ SD)	34 ($\pm$ 20.6)	31 ($\pm$ 20.8)
FEV ₁ , mean L/s ($\pm$ SD)	2.08 ($\pm$ 0.83)	2.16 ($\pm$ 0.71)
FEV ₁ /FVC, mean ($\pm$ SD)	66.4 ($\pm$ 15.2)	69.4 ($\pm$ 12.7)

*Adapted from Holleman et al.¹⁷ Validation sample represents consecutive patients with complete data for all variables.

group. The expected number of cases was contrasted with the observed number, and a χ^2 value was calculated for each of the four strategies. The degrees of freedom for each χ^2 test was g rather than $g - 1$ because we estimated the expected distribution from the original sample, not the validation sample.²²

RESULTS

The characteristics of the 164 patients in the original study sample and the 70 patients in the validation sample are shown in Table 1. Six items from the pulmonary history and one item from the pulmonary physical examination had LRs with 95% CIs excluding 1 for predicting airflow limitation (Table 2).¹⁷ These seven items were, therefore, useful in predicting airflow limitation according to the bivariate analysis. Three distinct cutpoints for the number of years the patient had smoked were determined

by inspection of the ROC curve (<30, 30–55, and >55 years). The seven useful clinical evaluation items were candidate predictor variables in a logistic regression model predicting airflow limitation. Items were excluded from the model, using backward elimination, if they were not significant predictors of airflow limitation ($p > .05$). Three items remained as independent predictors of airflow limitation: smoking history (<30, 30–55, and >55 years), patient reported wheezing, and clinician auscultated wheezing. The same model was chosen whether backward, forward, or stepwise selection methods were used.

Strategy 1 (all seven traditional LRs), strategy 2 (the single best LR), and strategy 3 (logistic regression–selected traditional LRs) used the LRs shown in Table 2. Strategy 4 (adjusted LRs) used LRs shown in Table 3. The ordered LRs used in Strategy 2 are shown in Table 4.

Comparing the Four Strategies

We calculated posterior probabilities of airflow limitation for each patient in the validation series using each of the four strategies. We constructed ROC curves for each of the strategies to compare overall diagnostic accuracy. The strategies of “logistic regression–selected traditional LRs” and “adjusted LRs” were significantly better at discriminating between patients with and without airflow limitation than the “all traditional LR” strategy as measured by the area under the ROC curve (areas = 0.79 and 0.79 vs 0.69; $p = .02$). The “single best LR” strategy was not significantly different from any of the other three strategies in discriminating between those with and those without airflow limitation (area = 0.75; $p = .20$). However, ROC area differences of 4% can be important in some clinical decision making; larger studies would be needed to exclude this magnitude of difference.

Table 2. Likelihood Ratios for Components from the Bedside Clinical Evaluation*

Component	LR ⁺	LR [−]
History		
Smoking duration		
<30 years	0.31	
30–55 years	1.5	
>55 years	3.1	
Dyspnea (any)	1.2	0.55
Reported cough	1.8	0.69
Reported wheezing	3.1	0.58
Chronic bronchitis symptoms	3.8	0.66
Exertional dyspnea		
Grade I	0.56	
Grade II	1.2	
Grade III–IV	2.0	
Physical Examination		
Auscultated wheezing	12	0.87

*Adapted from Holleman et al.¹⁷ Only statistically significant predictors of airflow limitation are included. LR⁺ indicates likelihood ratio for finding present; LR[−], likelihood ratio for finding absent.

Table 3. Adjusted Likelihood Ratios from the Logistic Regression Model

Smoking Duration (years)	Reported Wheezing	Auscultated Wheezing	Adjusted LR
<30	No	No	0.23
30–55	No	No	0.81
<30	Yes	No	0.87
<30	No	Yes	0.94
>55	No	No	2.3
30–55	Yes	No	3.1
30–55	No	Yes	3.3
<30	Yes	Yes	3.6
>55	Yes	No	8.8
>55	No	Yes	9.4
30–55	Yes	Yes	13
>55	Yes	Yes	36

Table 4. Likelihood Ratios in Descending Order of Ability to Alter Prior Odds

Component	LR
Auscultated wheezing	12
Chronic bronchitis symptoms	3.8
Smoked <30 years	0.31
Reported wheezing	3.1
Smoked >55 years	3.1
Exertional dyspnea, grade III-IV	2.0
Reported cough	1.8
No reported dyspnea	0.55
Exertional dyspnea, grade I	0.56
No reported wheezing	0.58
No chronic bronchitis symptoms	0.66
Smoked 30–55 years	1.5
No reported cough	0.69
Reported dyspnea	1.2
Exertional dyspnea, grade II	1.2
No auscultated wheezing	0.87

Table 5 summarizes the differences in ROC curve areas between strategies and the 95% CIs about the difference. Potentially important differences between strategies 4 or 3 and strategy 2; or between strategies 2 and 1 are not excluded (Table 5). Therefore, strategies 4 and 3 are at least as good as strategy 2; and strategy 2 is at least as good as strategy 1. Observed probabilities of airflow limitation did not deviate significantly from predicted probabilities (Table 6) for the strategies of “adjusted LRs” ($p = .56$), “logistic regression–selected traditional LRs” ($p = .48$), or “single best LR” ($p = .60$). However, observed probabilities differed significantly from predicted probabilities (Table 6) for the “all seven traditional LRs” strategy ($p < .0001$).

DISCUSSION

We have described an approach to identifying the conditionally independent findings from clinical examina-

tion studies, using the example of airflow limitation. Although the approaches are quantitative, they detail an important paradigm for defining the rational clinical examination. The rational clinical examination should contain only the smallest number of most useful and efficient history and physical examinations findings, eliminating redundant or inaccurate items. We found that using all seven clinical examination items that were significant on univariate analysis yielded inaccurate posttest probabilities of disease ($p < .0001$). This diagnostic inaccuracy was most likely a result of violating the conditional independence assumption. Furthermore, this strategy was less able than the logistic regression model to discriminate between patients with airflow limitation and those without ($p = .02$). This loss of predictive ability was primarily one of decreased sensitivity, with patients at the lowest predicted probabilities (4%) having a fairly high observed probability of disease (33%) (Table 6).

The strategy of choosing the clinical finding with the single best LR yielded accurate posttest probabilities for airflow limitation. However, it was less accurate than using findings identified in a logistic regression model. Although this decrement in discriminating ability was not statistically significant ($p = .2$), it was expected because the use of a single clinical examination item ignored diagnostic information contained in the results of other items. However, even this single clinical finding approach compared favorably with physicians' subjective clinical assessments in discriminating between patients with airflow limitation and those without (ROC curve areas of 0.72 vs 0.75).¹⁷ This strategy will yield valid estimates of conditional posterior odds for any target disorder; therefore, it is generalizable. However, predictions for individual patients will always be superior when the results of all truly independent clinical examination items are used to calculate posterior odds.

The other two strategies used logistic regression to identify three independent predictors of airflow limitation (smoking history, auscultated wheezing, and subjective wheezing). These two strategies had two advantages: they used all useful diagnostic information, unlike the single best clinical finding approach; and they yielded accurate and discriminating probabilities of airflow limitation, unlike the approach of using all seven items. Therefore, these strategies would be preferred over the preceding two strategies. In situations in which adjusting the LRs to be independent improved diagnostic accuracy, this approach (strategy 4) might be preferred; otherwise, the unadjusted approach (strategy 3) might be preferred because it may be simpler and more intuitive, and it allows for one or more unobtainable items to be excluded.

Our results have implications for clinicians and clinical researchers concerned with the rational clinical examination. The case of airflow limitation supports the concept that obtaining every relevant historical item and performing every physical examination maneuver may not only cause inefficiency, but also create inaccurate assess-

Table 5. Differences in ROC Curve Areas Between the Four Strategies*

	Strategy 4	Strategy 3	Strategy 2
Strategy 3	0 (−0.05–0.05)		
Strategy 2	0.04 (−0.02–0.11)	0.04 (−0.02–0.11)	
Strategy 1	0.09 (0.01–0.17)	0.09 (0.01–0.17)	0.05 (−0.03–0.13)

*Numbers in parentheses represent the 95% confidence interval of difference. Strategy 1 used the results of all seven clinical examination findings. Strategy 2 used the results of the single best clinical examination finding. Strategy 3 used the three clinical examination findings chosen by logistic regression. Strategy 4 was the same as strategy 3 except that the likelihood ratios were adjusted by the logistic model.

Table 6. Mean Predicted Versus Observed Probabilities for the Four Strategies*

Tertile	Strategy 1		Strategy 2		Strategy 3		Strategy 4	
	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed
1	0.04	0.33	0.17	0.09	0.09	0.05	0.13	0.05
2	0.15	0.23	0.35	0.46	0.30	0.46	0.35	0.46
3	0.69	0.62	0.73	0.65	0.68	0.67	0.65	0.67

*Strategy 1 used the results of all seven clinical examination findings. Strategy 2 used the results of the single best clinical examination finding. Strategy 3 used the three clinical examination findings chosen by logistic regression. Strategy 4 was the same as strategy 3 except that the likelihood ratios were adjusted by the logistic model.

ments. Because the clinical examination finding “drives” subsequent investigations, inaccurate diagnoses may cause harm to patients in the form of inappropriate additional testing or treatment. When researchers study several tests that are likely to be used together, they should investigate how violations in the independence assumption affect the accuracy of clinicians’ posterior probability assessments, using strategies like those outlined in this manuscript. Likewise, clinicians should be cautious when using multiple clinical findings for the same condition, such as might be found in review articles or textbooks, when the independence assumption has not been tested. We urge clinicians to search for combinations of findings that have been chosen by variable reduction methods that eliminate the more egregious violations of the independence assumption (our strategy 3 or 4). When the independence assumption has not been substantiated, relying on the finding with the single best LR (strategy 2) would be a valid alternative for clinicians—one free of any conditional independence assumptions.

REFERENCES

- Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd ed. Boston, Mass: Little, Brown and Co; 1991.
- Sox HC, Blatt MA, Higgins MC, Marton KI. Medical Decision Making. Boston, Mass: Butterworths; 1988.
- Weinstein MC, Fineberg HV, Elstein AS, et al. Clinical Decision Analysis. Philadelphia, Pa: WB Saunders; 1980.
- Detrano R. Optimal use of literature knowledge to improve the Bayesian diagnosis of coronary artery disease. *J Clin Epidemiol*. 1989;42:1041-7.
- Diamond GA. Off Bayes: effect of verification bias on posterior probabilities calculated using Bayes’ theorem. *Med Decis Making*. 1992;12:22-31.
- Patterson RE, Eng C, Horowitz SF. Practical diagnosis of coronary artery disease: a Bayes’ theorem nomogram to correlate clinical data with noninvasive exercise tests. *Am J Cardiol*. 1984;53:252-6.
- Morise AP, Duval RD. Comparison of three Bayesian methods to estimate posttest probability in patients undergoing exercise stress testing. *Am J Cardiol*. 1989;64:1117-22.
- Fryback DG. Bayes’ theorem and conditional nonindependence of data in medical diagnosis. *Comput Biomed Res*. 1978;11:423-34.
- Russek E, Kronmal RA, Fisher LD. The effect of assuming independence in applying Bayes’ theorem to risk estimation and classification in diagnosis. *Comput Biomed Res*. 1983;16:537-52.
- Detrano R, Leatherman J, Salcedo EE, Yiannikas J, Williams G. Bayesian analysis versus discriminant function analysis: their relative utility in the diagnosis of coronary disease. *Circulation*. 1986;73:970-7.
- Sackett DL. A primer on the precision and accuracy of the clinical examination. *JAMA*. 1992;267:2638-44.
- Weintraub WS, Madeira SW, Bodenheimer MM, et al. Critical analysis of the application of Bayes’ theorem to sequential testing in the noninvasive diagnosis of coronary artery disease. *Am J Cardiol*. 1984;54:43-9.
- Crichton NJ, Fryer JG, Spicer CC. Some points on the use of “independent Bayes” to diagnose acute abdominal pain. *Stat Med*. 1987;6:945-59.
- Ohmann C, Kunneke M, Zaczek R, Thon K, Lorenz W. Selection of variables using “independence Bayes” in computer-aided diagnosis of upper gastrointestinal bleeding. *Stat Med*. 1986;5:503-15.
- Chard T. The effect of dependence on the performance of Bayes’ theorem: an evaluation using a computer simulation. *Comput Methods Programs Biomed*. 1989;29:15-9.
- Hilden J. Statistical diagnosis based on conditional independence does not require it. *Comput Biol Med*. 1984;14:429-35.
- Holleman DR, Simel DL, Goldberg JS. Diagnosis of obstructive airways disease from the clinical examination. *J Gen Intern Med*. 1993;8:63-8.
- Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS recommendations. *Am Rev Respir Dis*. 1981;123:659-64.
- American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Respir Dis*. 1991;144:1202-18.
- Simel DL, Feussner JR, DeLong ER, Matchar DB. Intermediate, indeterminate, and uninterpretable diagnostic test results. *Med Decis Making*. 1987;7:107-14.
- Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol*. 1991;44:763-70.
- Hosmer DW, Lemeshow S. Applied Logistic Regression. New York, NY: John Wiley and Sons; 1989.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
- Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148:839-43.