Generalized Likelihood Ratio Concept and Logistic Regression Analysis for Multiple Diagnostic Categories

Gilbert Reinbnegger, Dietmar Fuchs, Arno Hausen, Ernst R. Werner, Gabriele Werner-Felmayer, and Helmut Wachter

Albert (Clin Chem 1982;28:1113–9) has proposed estimation of likelihood ratios by logistic regression analysis. The usual likelihood-ratio approach for estimation of post-test probability of disease from sensitivity and specificity data of a diagnostic test has been extended by Birkett (J Clin Epidemiol 1988;41:491–4) for situations with more than two diagnostic categories. We suggest here a combination of these ideas, demonstrating this by a re-evaluation of previously published data on the validity of neopterin as a tool for differential diagnosis between chronic non-A, non-B hepatitis and fatty liver. Analysis of neopterin data in combination with the ratio between serum concentrations of aspartate aminotransferase and of alanine aminotransferase yielded a good discrimination between three mutually exclusive diagnostic categories, namely, fatty liver and chronic persistent and chronic aggressive non-A, non-B hepatitis. The approach is flexibly applicable to situations with different pre-test probabilities. The sum of estimated post-test probabilities deviates slightly from the sum of pre-test probabilities. This deviation is a function of the coefficients obtained in logistic regression, and an analytical expression for the deviation is given. The generalized likelihood-ratio approach appears promising in complex diagnostic situations when multiple diagnostic tests are available.

Additional Keyphrases: clinical decision-making · test evaluation · neopterin · hepatitis · fatty liver

The likelihood ratio (LR) has been suggested as a single numerical value to quantify the diagnostic ability of a test (1). The relationship between the LR and the diagnostic sensitivity (Se) and specificity (Sp) of a binary diagnostic test (e.g., "normal" vs "abnormal") is simple in the situation of two mutually exclusive diagnostic classes (e.g., "healthy" vs "diseased"):

\[
LR = \frac{Se}{1 - Sp}
\]

By Bayes' theorem, use of the LR permits computation of post-test probabilities of disease if estimates of pre-test probabilities are available. Importantly, LR provides a straightforward means to convert to pre-test odds (for the presence of disease vs absence of disease) into post-test odds, by simple multiplication. A generalization of the LR concept for evaluating dichotomous diagnostic tests for use in the situation of multiple (i.e., more than two) mutually exclusive diagnostic categories has been presented recently by Birkett (2). On the other hand, for dichotomous clinical outcome, Albert (3) has advocated estimation of the LR by logistic regression analysis, which allows better exploitation of the information content provided by a test with multiple possible results, e.g., a quantitative clinical chemical test, or even a combination of multiple diagnostic tests with both quantitative and qualitative results.

We have recently shown that, in a situation with mutually exclusive diagnostic classes, not all binary LRs can be freely chosen (4). Rather, if LR_{AB} and LR_{AC} are given (where A, B, and C denote different diagnostic classes), then LR_{BC} must obey the following relationship:

\[
LR_{BC} = LR_{AC}/LR_{AB}
\]

Otherwise, the sum of computed post-test probabilities will not match the sum of pre-test probabilities. (This condition is easily understandable if one remembers the simple relationship between LR, Se, and Sp.) If LR_{BC} differs from the above-given condition by a factor k, the deviation of the sum of post-test probabilities from the sum of pre-test probabilities, which always can be chosen as unity by appropriate definition of the diagnostic categories, is equal to

\[
[(r + s)/(r + s + rs)] - [(r)/(r + rs + ks)] - [(ks)/(r + ks + kr)k]
\]

where

\[
r = LR_{AB} (\Pi_A/\Pi_B),
\]

\[
s = LR_{AC} (\Pi_A/\Pi_C),
\]

and \Pi_A, \Pi_B, and \Pi_C are the pre-test probabilities of the respective diagnostic classes.

Here we explore a combination of these more general concepts (2, 3). Data are re-analyzed from a previously published study on patients with fatty liver, with chronic persistent non-A, non-B hepatitis, and with chronic aggressive non-A, non-B hepatitis (5). In that study, urinary concentrations of neopterin, which is a marker for the activation of cell-mediated immunity (6), were compared with a panel of several routine laboratory tests.

Thus, we describe a situation with three different diagnostic categories that are regarded to be mutually exclusive. Our aim is to discriminate among these diagnostic categories on the basis of laboratory data, using a combination of Birkett's approach (2) and of logistic regression analysis for estimation of the required binary likelihood ratios as proposed by Albert (3).

Materials and Methods

Subjects

As described in detail in a previous publication (5), 42 patients with liver disease—16 with fatty liver (FL), 16
with chronic persistent non-A, non-B hepatitis (CPH), and 10 with chronic aggressive non-A, non-B hepatitis (CAH)—were tested for a panel of laboratory variables: erythrocyte sedimentation rate, aspartate aminotransferase (AST), alanine aminotransferase (ALT), \( \gamma \)-glutamyltransferase, alkaline phosphatase, bilirubin, cholesterol, triglycerides, gamma globulins, Quetelet index, and urinary neopterin concentration (measured as micromoles of neopterin per mole of creatinine). Of all variables studied, we found neopterin to discriminate best between FL on the one hand and CPH and CAH on the other. However, no difference was found for neopterin concentrations among CPH and CAH patients. The ratio between AST and ALT was shown to discriminate well between CPH patients on the one hand and FL and CAH patients on the other. However, this variable was not able to discriminate between FL and CAH patients. Figure 1 shows the cumulative frequencies of these two variables, neopterin and AST/ALT, among the three diagnostic categories. It seemed reasonable for the present study to explore a combination of these two laboratory tests for the problem of discrimination between all three diagnostic classes, because these variables apparently provide independent and complementary information.

Methods of Computation

Two variants of the analysis are shown: First, a combination of neopterin data, coded as (quasi)continuous variable (i.e., the measurement results are used without any transformation), and AST/ALT ratios, transformed into a dichotomous variable (below or above the limit AST/ALT = 0.7), is investigated. Second, both variables are used without transformation.

With either approach, logistic regression analyses are performed with subsequent use of each of the three possible binary combinations of diagnostic categories (i.e., FL vs CPH, FL vs CAH, and CPH vs CAH). Program BMDPLR (BMDP software) is used for this step, which also includes an analysis of a possible significant interaction between both variables, neopterin and AST/ALT ratio. As detailed by Albert (3), LR's are then computed from the computed regression coefficients. Here it is necessary to correct for unequal pre-test probabilities of the two diagnostic classes used in the logistic regression analysis (3). The LR's are finally used with Birkett's formulas (2) for estimating post-test probabilities of each diagnostic class, in dependence on neopterin concentration and AST/ALT ratio. Because pre-test probabilities are required for this step, we use equal relative frequencies of each diagnostic category. However, post-test probabilities for any other combination of pre-test probability are easily computed with this approach.

For comparison of the classification results obtained by the method presented here, we perform a linear discriminant analysis on the same data set, using program BMDP7M.

Finally, we investigate the deviation of the sum of computed post-test probabilities from unity by using the formula (4) repeated in the Introduction.

**Results**

One Continuous Variable, One Dichotomous Variable

Table 1 shows the results of logistic regression analyses that are performed for each of the three possible binary combinations of diagnostic categories. For these analyses, neopterin data are used without transformation, and AST/ALT ratios are transformed into a binary variable. A previous exploratory analysis did not reveal a significant interaction between the two variables (not shown). According to the procedure given by Albert (3), LR's for binary comparisons of diagnostic classes are then estimated, whereby one has to adjust for unequal pre-test probabilities of the diagnostic categories used in the logistic regression analyses (Table 1). Making use of Birkett's formula to transform these LR's into post-test probabilities for a given set of pre-test probabilities, we obtained Figure 2, showing post-test probabilities in dependence on neopterin and the two levels of AST/ALT ratios. For low neopterin concentrations the post-test probability for diagnosis FL is seen to be highest. For higher neopterin concentrations, the results depend strongly on AST/ALT ratios: an AST/ALT ratio below 0.7 yields a high post-test probability for CPH over a

<table>
<thead>
<tr>
<th>Diagnoses compared</th>
<th>Regression coefficients (and P-value) *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neopterin</td>
<td>AST/ALT</td>
</tr>
<tr>
<td>FL vs CPH</td>
<td>-0.034798</td>
</tr>
<tr>
<td></td>
<td>(&lt;0.0001)</td>
</tr>
<tr>
<td>FL vs CAH</td>
<td>-0.037957</td>
</tr>
<tr>
<td></td>
<td>(&lt;0.0001)</td>
</tr>
<tr>
<td>CPH vs CAH</td>
<td>-0.004387</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
</tr>
</tbody>
</table>

**Table 1. Estimation of Likelihood Ratios for Binary Comparisons of Diagnostic Categories by Using One Continuous Variable (Urinary Neopterin Concentration, µmol/mol Creatinine) and One Categorical Variable (AST/ALT Ratio)**

I) Logistic regression analysis

<table>
<thead>
<tr>
<th>Diagnoses compared</th>
<th>Likelihood ratio function *</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL vs CPH</td>
<td>exp[-0.034798x1 + 2.4667x2 + 7.9644 - log(16/16)]</td>
</tr>
<tr>
<td>FL vs CAH</td>
<td>exp[-0.037957x1 + 1.4733x2 + 8.1324 - log(16/10)]</td>
</tr>
<tr>
<td>CPH vs CAH</td>
<td>exp[-0.004387x1 - 1.8066x2 + 1.2641 - log(16/10)]</td>
</tr>
</tbody>
</table>

* Coded +1, if AST/ALT >0.7; -1, otherwise.

**Table 2. Likelihood ratios as functions of neopterin concentration (x1) and AST/ALT ratio (x2)**

- Estimated by a chi-square-to-enter statistic.
- Including an adjustment term for unequal pre-test probabilities.
wide range of neopterin concentrations (only very high neopterin concentrations make CAH more likely); an AST/ALT ratio exceeding 0.7 indicates a very high post-test probability for CAH, whereas the post-test probability for CPH is very low for all neopterin concentrations found.

Two Continuous Variables

We made an analogous calculation, using neopterin data as well as AST/ALT ratios as (quasi)continuous variables, i.e., making full use of the information content of the quantitative measurement results. Table 2 shows the results of the logistic regression analyses and the estimation of the LRs for binary comparison of diagnostic classes. Again, no significant interaction between both variables is found (not shown). Figure 3 is obtained by applying Birkett's formula for estimating post-test probabilities of each diagnostic category. Figure 3 shows, in accordance with the preceding results, that low neopterin values are a strong indication for FL, whereas high neopterin and low AST/ALT ratio indicate CPH and high neopterin combined with higher values of AST/ALT ratio indicate CAH. Table 3 shows in detail how post-test probabilities for each of the diagnostic categories are estimated for a particular pair of values for neopterin concentration and AST/ALT ratio. A relatively high degree of numerical accuracy is deliberately chosen to point out the deviation from unity of the sum of estimated post-test probabilities, which stems from the fact that the LRs estimated by logistic regression analyses are only approximative. As Table 3 shows, it is also possible to calculate this deviation by using the estimated LRs directly. Figure 4 demonstrates, for two different sets of pre-test probabilities, the deviation as a function of neopterin and AST/ALT measurements. Obviously, for most situations the deviation is negligible. There are combinations of the test results, however, in which the deviation becomes marked, depending also strongly on the pre-test probabilities.

For this model, the classification matrix is estimated (using the observed frequencies of diagnostic classes as pre-test probabilities) and compared with the results obtained by classical linear discriminant analysis. Overall, 35 of the 42 cases (83.3%) are correctly classified by the model presented here (category FL: 14 cases classified as FL, two cases as CPH; category CPH: 13 cases classified as CPH, two cases as FL, and one as CAH; category CAH: eight cases classified as CAH, one as FL, and one as CPH). The linear discriminant model yields an overall rate of 32/42 (76.2%) correct classifications (category FL: 12 cases classified as FL, two cases as CPH, and two as CAH; category CPH: 13 cases classified as CPH, two cases as FL, and one as CAH; category CAH: seven cases classified as CAH, two as FL, and one as CPH).

Discussion

The concepts of diagnostic sensitivity and specificity of a dichotomous diagnostic test in a dichotomous clinical decision are well known but suffer from several limitations: (a) Clinical problems cannot always be reduced to a "no-yes" decision. (b) Clinical chemists try to establish quantitative diagnostic assay procedures providing a test result on a (quasi)continuous scale, but very often in the step of test interpretation one just looks at whether the result lies below or above a "cutoff" value, thereby seriously reducing the information inherent in the quantitative test value. Various concepts have been developed to overcome this situation.

We are investigating here a combination of two promising concepts, namely, generalizing the scope of clinical decision analyses to situations with multiple diagnostic levels by applying a generalized LR technique (2) and generalizing the estimation of LRs by logistic regression analysis (3). The combination of both ideas appears to be very attractive.

Perhaps most importantly, the technique is very flexibly applicable to clinical situations with varying pre-test probabilities. One can easily adjust to a different set of pre-test probabilities by making use of Birkett's formula (2). The importance of adjusting prediction rules to different sets of pre-test probabilities has been recognized by various authors, and the LR has been used to transport prediction rules among different clinical settings (7).

The combination of two or more diagnostic tests, both quantitative or qualitative, is readily possible in the gen-

Table 2. Estimation of Likelihood Ratios for Binary Comparisons of Diagnostic Categories by Using Two Continuous Variables (Urinary Neopterin Concentration, μmol/mol Creatinine, and AST/ALT Ratio)

<table>
<thead>
<tr>
<th>Diagnoses compared</th>
<th>Regression coefficients (and P-value)*</th>
<th>Likelihood ratio functionb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neopterin</td>
<td>AST/ALT</td>
</tr>
<tr>
<td>FL vs CPH</td>
<td>-0.030526</td>
<td>5.1306</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0054)</td>
</tr>
<tr>
<td>FL vs CAH</td>
<td>-0.034267</td>
<td>2.8498</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>CPH vs CAH</td>
<td>-0.005927</td>
<td>-7.8935</td>
</tr>
</tbody>
</table>

* Estimated by a chi-square-to-enter statistic.

b Including an adjustment term for unequal pre-test probabilities.
The LRs used are given in Table 2. Here, urinary neopterin concentrations (μmol per mol creatinine) and serum AST/ALT ratios are used without transformation. Equal pre-test probabilities for each diagnostic category are assumed.

Table 3. Estimation of Post-Test Probabilities by Using the Model Given in Table 2: Pre-Test Probabilities \( \Pi \), for All Diagnostic Categories Are Assumed to Be Equal; Neopterin = 300 μmol/mol Creatinine, AST/ALT = 0.30

<table>
<thead>
<tr>
<th></th>
<th>FL vs CPH:</th>
<th>FL vs CAH:</th>
<th>CPH vs CAH:</th>
</tr>
</thead>
<tbody>
<tr>
<td>I) Likelihood ratios (compare Table 2):</td>
<td>LR(_{AB}) = 0.009279</td>
<td>LR(_{AC}) = 0.010145</td>
<td>LR(_{BC}) = 15.196215</td>
</tr>
<tr>
<td>II) Post-test probabilities (note that LR(<em>B = LR</em>{B^{-1}})):</td>
<td>FL: (1 + LR_{AB} (\Pi_{A}/\Pi_{AB})^{-1} + LR_{AC} (\Pi_{A}/\Pi_{AC})^{-1}) = 0.004823</td>
<td>CPH: (1 + LR_{AB} (\Pi_{A}/\Pi_{AB}) + LR_{AC} (\Pi_{A}/\Pi_{AC})^{-1}) = 0.930169</td>
<td>CAH: (1 + LR_{AC} (\Pi_{A}/\Pi_{AC}) + LR_{BC} (\Pi_{A}/\Pi_{BC})^{-1}) = 0.061704</td>
</tr>
<tr>
<td>III) Sum of post-test probabilities = 0.996668, deviation from unity = 0.003314</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV) Direct estimation of the deviation from unity from the likelihood ratio functions:</td>
<td>LR(<em>{BC}) differs from the ratio LR(</em>{AC}/LR_{AB}) by a factor (see Table 2) given by (k = \exp(-0.002188\text{ neopterin} - 5.6127\text{ AST/ALT} + 4.9720) = 13.898626). Use of the formula given in the Introduction yields the deviation = 0.003314.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4. Deviation from unity of the sum of post-test probabilities estimated by Birkett's formula (2). The LRs used are given in Table 2. The analytic expression for the deviation as function of urinary neopterin concentration (μmol/mol creatinine) and serum AST/ALT data (and implicitly of the LRs) is given in Table 3. (a) Equal pre-test probabilities for all diagnostic categories are assumed. (b) Pre-test probabilities: FL: 0.10, CPH: 0.80, CAH: 0.10

Generalized LR approach by using logistic regression analysis.

Logistic regression techniques can only provide more-or-less imprecise estimates of the LRs (3). A condition may be violated by these estimates, which we would expect to hold between binary LRs in the case of multiple mutually diagnostic categories (4). Thus, the computed post-test probabilities suffer from the fact that their sum does not exactly match the sum of the assumed pre-test probabilities. However, this defect can be formulated analytically in the generalized LR approach by using logistic regression analysis. As Figure 4 demonstrates, the deviation depends on the test results (by virtue of the regression coefficient) as well as on the pre-test probability estimates. In the examples we studied, the deviation is not very significant in most situations. Here, renormalization of computed post-test probabilities might be sufficient. The deviation may become large, however, particularly in situations with extreme pre-test probabilities of one of the diagnostic classes.

The classification power of the method presented here compares favorably with that of linear discriminant analysis (each case was classified into the category with the highest calculated post-test probability).

Some limitations of the present study must be mentioned: The frequency of patients in the diagnostic classes...
is low. Thus, the regression coefficients obtained by logistic regression analysis, and hence the LRs derived here, might require reassessment in a larger study group. Nevertheless, the primary aim of this paper is to demonstrate the applicability of the evaluation technique, rather than to derive a final prediction rule. Furthermore, it must be stressed that the method presented is applicable only in the case of mutually exclusive diagnostic categories, as has been pointed out and discussed in detail by Birkett (2).

We gratefully acknowledge the financial support by the "Jubiläumsfonds" of the Austrian National Bank, project 3127.

References