Use of Probability Theory to Evaluate Cost Effectiveness of Laboratory Tests in Insurance Medicine

John R. Iacovino, M.D.
Vice President-Medical Director
New York Life Insurance Company
New York, NY

As technology advances, new laboratory tests are spawned to improve diagnostic efficiency. These tests are marketed to insurance companies by their laboratories which, on occasion, do not consider the problems unique to insurance testing. In order to respond to marketing pressures, the medical director must evaluate two factors. First, whether the new test, which is almost always developed for clinical use, is useful for screening populations where disease prevalence is low. An example is fructosamine. It has clinical utility, but what is its role in screening for glucose intolerance in a low prevalence population? Second, is the test cost effective? The evaluation of the latter factor will be the basis of this paper.

Our emphasis should not be on cost per test but cost per new individual correctly diagnosed and, most importantly in insurance medicine, accurately underwritten.

In addition to correctly underwritten cases, the medical director requires tools to evaluate the percentage of overrated and underrated cases as both will affect profitability of their company. Those overrated will not likely take their policies; those underrated will take their toll at claim time.

Figure 1 is a 2x2 table illustrating the possible underwriting results on testing a population with and without a disease.

Figure 1.
2x2 table illustrating underwriting possibilities.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Test</td>
<td>Number of individuals diseased and who have a positive test (a)</td>
<td>Number of individuals without disease who have a positive test (c)</td>
</tr>
<tr>
<td>Negative Test</td>
<td>Number of individuals diseased and who have a negative test (b)</td>
<td>Number of individuals without disease who have a negative test (d)</td>
</tr>
</tbody>
</table>

The sensitivity (Sn) and specificity (Sp) of a test will yield the proportion of individuals in each of the four preceding underwriting groups.

Sensitivity (Sn) — the proportion of those diseased, who have a positive test.

\[
\frac{a}{a + b}
\]

Specificity (Sp) — the proportion of those disease-free, who have a negative test.

\[
\frac{d}{c + d}
\]

Sn and Sp are independent of the prevalence of disease in a population; they are a measure of the test itself. However, when one wishes to calculate the number of individuals in a tested group, disease prevalence must be considered.

Another way to evaluate a test is the use of predictive values. The proportion and, if prevalence is considered, the number of correctly underwritten individuals can be predicted through their use.

Positive Predictive Value (PPV) — the proportion of those with a positive test who are diseased.

\[
\frac{a}{a + c}
\]

Negative Predictive Value (NPV) — the proportion of those with a negative test who are disease-free.

\[
\frac{d}{b + d}
\]

Several laboratories offer to reflex various apolipoproteins from predetermined levels of HDL and cholesterol or their ratio. The medical director must decide: a) Is this cost effective? and b) What will be the underwriting implication? To answer these, facts must be blended with assumptions. The facts are Sn, Sp and cost of the reflex tests; all are available from the laboratory. The assumptions are the responsibility of the medical director and include: a) At what levels do I want to reflex the HDL and cholesterol? and b) What is the prevalence of the disease, for which I am testing, in the population?

The assumptions are difficult! The medical director, not the laboratory, should determine the appropriate reflex levels predicated on their company's marketing, agency and financial goals and constraints. If the reflex level is too low, costs
Use of Probability Theory to Evaluate Cost Effectiveness

Volume 21, No. 3 1989

will increase since more reflex testing will be generated. This diminishes cost effectiveness. Also if the reflex level is too low, you may classify too many individuals as abnormal, resulting in a marketing and agency revolt. The medical director must keep the number of rated and declined cases within a given range if a product is to be successfully marketed. Various degrees of mortality must be accepted; this is considered in the actuarial pricing. Conversely, if you set the reflex too high, you will not diagnose enough individuals to be cost effective. Additionally, you will underrate too many individuals; however, marketing and agency will love you.

Let us proceed to determine the cost effectiveness of a reflex test.

The facts are the Sn and Sp. For the HDL/cholesterol ratio which is superior to the cholesterol alone in predicting mortality, the Sn is .71 and the Sp is .87. The calculations use a cohort of 1000 proposed insureds. The reflex level is established to identify only the worst 10% of the population at risk. Remember, if we go lower, we risk rating too many individuals with the potential loss of business. Figure 2 shows the underwriting results.

![Figure 2. HDL/Cholesterol underwriting result](image)

<table>
<thead>
<tr>
<th>Disease</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>71</td>
</tr>
<tr>
<td>-</td>
<td>29</td>
</tr>
<tr>
<td>100</td>
<td>783</td>
</tr>
</tbody>
</table>

From the preceding figure, the following underwriting results are obtained:
- Correctly underwritten 85%
- Overrated 12%
- Underrated 3%

The HDL/cholesterol ratio with its relatively low sensitivity (.71) coupled with a low risk population (.10) will yield more false positive than true positive results. The PPV would be only 38%; the NPV is 96%. Combining these, we would correctly underwrite 85% of those tested by the HDL/cholesterol ratio alone.

The laboratory will automatically reflex all our “positive” HDL/cholesterol ratios above a preset level, despite false positives comprise 62% of the total positives. There is no way to separate the true from false positives.

I have elected to use the ApoA1/ApoB ratio for the reflex test. Its Sn and Sp are considerably better than for ApoA1 alone. Assume the laboratory charges $8.00 per reflex test. Up front, this will cost $1500 per 1000 blood samples tested (188 positives x $8.00). Is this cost justified? For the ApoA1/ApoB ratio the Sn is .87 and Sp .80. We will assume (guess) the disease prevalence is now 85% in those reflexed (recall we took the top 10% of abnormals yet due to the low specificity had more false positives than true positives). The total number of reflexed tests is 188. Eighty five percent of those we assume are diseased, so we have 160 diseased and 28 nondiseased. This underwriting result is illustrated in Figure 3.

![Figure 3. ApoA1/ApoB ratio reflexed underwriting result](image)

<table>
<thead>
<tr>
<th>Disease</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>140</td>
</tr>
<tr>
<td>-</td>
<td>20</td>
</tr>
<tr>
<td>160</td>
<td>22</td>
</tr>
</tbody>
</table>

From the above 2x2 table, we have the following results on these 188 tests.
- Correctly underwritten 86%
- Overrated 3%
- Underrated 11%

For those original positives reflexed by the ApoA1/ApoB ratio, the PPV is 96% and the NPV is 52%. We have nearly as many false negatives as true negatives due to the relatively low specificity (.80) of the test.

Finally, we must assimilate all the results from the original and the reflex test groups (Figure 4).

![Figure 4. HDL/Cholesterol and ApoA1/ApoB ratio combined underwriting result](image)

<table>
<thead>
<tr>
<th>Disease</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>140</td>
</tr>
<tr>
<td>-</td>
<td>49</td>
</tr>
<tr>
<td>160</td>
<td>805</td>
</tr>
</tbody>
</table>

- Correctly underwritten 94%
- Overrated 1%
- Underrated 5%

The PPV for our entire initially tested and reflexed group is 96%; the NPV is 94%.

The following (Table 1) summarizes the results of testing.

**Discussion:**

Laboratory costs are dramatically escalating as testing limits, necessitated by the AIDS epidemic, continue to drop.
Use of Probability Theory to Evaluate Cost Effectiveness

Table 1.

Summary of Test Results (1000 tested individuals)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Original group tested (1000)</td>
<td>71</td>
<td>783</td>
<td>117</td>
<td>29</td>
</tr>
<tr>
<td>Result of positive tests reflex</td>
<td>(188)</td>
<td>140</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>Final outcome of original group tested (1000)</td>
<td>140</td>
<td>805</td>
<td>6</td>
<td>49</td>
</tr>
<tr>
<td>Net change from reflex testing</td>
<td>+69</td>
<td>+22</td>
<td>-111</td>
<td>+20</td>
</tr>
</tbody>
</table>

may come a time, in the not distant future, when we test all applicants for insurance. Facing this expense, medical directors must become involved in the financial planning of their company to help maintain increasingly thin profit margins. By the use of probability theory, we can make informed financial decisions to strengthen our importance in our companies.

Sn and Sp are determined by testing diseased and nondiseased reference or gold-standard populations. Their calculation is independent of the prevalence of disease. However, when used to determine cost effectiveness, disease prevalence must be considered. Prevalence is estimated both by knowledge of disease and the company’s agency and marketing needs. Each medical director must derive their own prevalence assumptions.

Prevalence can profoundly affect the underwriting outcome of a particular test depending on its Sn and Sp. A test with a low sensitivity in a low prevalence population carries the risk of excessive false positives; cases which will be overrated or declined. Conversely, a low specificity carries the risk of excessive false negatives; cases which will be underrated with subsequent excessive mortality. Given the same Sn and Sp, these disparities narrow as the prevalence increases.

The purpose of reflexing the HDL/cholesterol ratio is to use a test of higher sensitivity to identify highest risks in the abnormal or high risk group; a secondary gain is the removal, via a higher specificity, from the reflex group of those not at highest risk. However, those removed are not necessarily low risk; we cannot rate every abnormal risk.

Initially, we identified 71 true positives (those presumed at highest risk) through our basic testing profile. After reflexing all 188 positives, we had 140 true positives for a net gain of 69 true positives out of our original 1000 tested individuals (Table 1).

Per 1000 tests, we spent $1500 to identify an additional 69 true positives and 22 true negatives. In actuality, the 22 true negatives are not clinical true negatives. They were derived from our original 188 positives (71 true, 117 false); those initially identified as high risk. As discussed, for marketing reasons, we are unable to rate all abnormals in a group. With the new cholesterol normal limits we could theoretically rate up to 25% of our insurance buying population — this is clearly not acceptable. We must cull out of this large group those individuals at highest risk, in this case the reflexed true positives. Therefore, the number of new correctly underwritten tests is not the 91 total true test results but only the 69 true positives. The cost per 1000 per new correctly underwritten test with the preceding assumption would be $21.75. Additionally, we generated 20 more false negatives per 1000 that would be underrated.

Mitigating this $21.75 cost per new correctly underwritten test per 1000 is our false positives (overrated cases) are reduced by 111.

The difficult question to be answered in this reflex scenario is whether the net gain of 69 additional correctly underwritten tests per 1000 at a cost of $21.75 each, is cost effective in view of its potential mortality savings.

In summary, we have examined a methodology to evaluate the cost effectiveness of laboratory tests. Each medical director must use both objectivity and subjectivity to arrive at a decision which is consistent with their company’s goals.

References

1. Riegelman RK. Diagnostic discrimination of test. In, Studying a Study and Testing a Test — How to read the medical literature. 1981. Little, Brown and Company, Boston:

2. Home Office Reference Laboratory. Laboratory Bulletin 88-06.


