Letters to the Editor


Tockman and colleagues reported the preliminary findings of two prospective studies evaluating hnRNP\(^1\) A2/B1 overexpression in sputum cells as an early detector/predictor of subsequent lung cancer, predominantly non-small cell lung cancer (1). These studies developed from earlier work involving anti-lung cancer antibodies reported in 1988 (2). Both recent studies involved individuals at high risk of developing lung cancer. The first study evaluated the development of SPLCs in stage I resected, disease-free lung cancer patients whose annual risk of second primary lung cancer was 1–5%. The second study involved Chinese miners exposed to tobacco smoke, radon, and arsenic and whose average annual incidence of primary lung cancer was 1%. The authors summarized the preliminary study findings using the test’s PPV. Tockman and colleagues concluded that “in these advance reports of two prospective studies, where the background cancer risks were 2.2 and 0.9%, respectively, hnRNP A2/B1 overexpression consistently and correctly predicted lung cancer in 67% and 96%, a 35- and 76-fold improvement in positive predictive value.” The authors’ summary contingency tables for these two studies are presented here in Tables 1 and 2. These initially reported PPVs are now reemerging in the secondary literature (3, 4). Unfortunately, the two reported PPVs as presented are biased, misinforming, and artificial.

A test’s PPV \(\frac{A}{A+B}\) in Table 1] measures the probability of a positive test detecting a real positive case. The PPV is not determined only by the sensitivity and specificity of the test but is very strongly influenced by the proportion of cases and controls sampled. A test’s PPV estimated in a sample will not reflect the PPV in the population, unless the proportion of cases (or conversely, controls) in the sample is similar to that occurring in the population. With uncommon or rare diseases, the relatively large numbers of controls (cell N in Table 1) and the consequent large numbers of individuals (cell B of Table 1) tend to deflate PPVs, even when tests have high specificities. In these two studies, most of the cases were sampled, but only a very small fraction of the controls were sampled. As a consequence, the sampling proportions used to estimate the PPVs were unrepresentative; sample and population case proportions were 32.5% \(\text{versus}\) 2.2% and 48% \(\text{versus}\) 0.9% for the SPLC and Chinese miner studies, respectively (Tables 1–4).

A simple approach for estimating the PPV for the population is to apply the sensitivity and specificity obtained by evaluating the test in cases and controls, respectively, back to the population that was sampled. This was done using the data presented in the Tockman et al. report (1) and is presented in Tables 3 and 4. The resultant PPV was 8.5% for the SPLC population and was 2.1% for the Chinese miner population. In both “high risk” populations, less than 1 in 10 individuals with a positive test are expected to develop lung cancer in the time period considered. Applying this test to the general population will yield an even lower PPV.

The PPVs as presented by Tockman and colleagues are arbitrary. They can be manipulated just by altering the number of controls sampled. For example, in the Chinese miners study (Table 2), if instead of sampling 49 controls they had sampled only 17 controls and the sensitivity and specificity remained unchanged, then cell B would contain 11 individuals and the PPV would have improved to 37/(37+11) = 77%. In the samples presented, given the worst case scenario, \(i.e.,\) if the test specificity was zero and all controls without lung cancer tested positive, the PPV in both studies would still have been relatively high: 27% and 43% for the SPLC and Chinese miner studies, respectively. Given the sampling strategies, the PPVs of the preliminary studies as they apply to the samples could not have done poorly. The crux of the error lies in generalizing the PPVs of the study samples directly to the populations that the samples were drawn from.

Received 8/2/99; accepted 8/27/99.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

\(^1\) The abbreviations used are: hnRNP, heterogeneous nuclear ribonucleoprotein; SPLC, second primary lung cancer; PPV, positive predictive value.

---

**Table 1** (Table 3) SPLC study

<table>
<thead>
<tr>
<th>Samples</th>
<th>Lung cancer</th>
<th>No lung cancer</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test +</td>
<td>A 10</td>
<td>B 5</td>
<td>R 15</td>
</tr>
<tr>
<td>Test -</td>
<td>C 3</td>
<td>D 22</td>
<td>S 25</td>
</tr>
<tr>
<td>Totals</td>
<td>M 13</td>
<td>N 27</td>
<td>T 40</td>
</tr>
</tbody>
</table>

Sensitivity \(= 77\%\)
Specificity \(= 81\%\)
Proportion of cases \(= 32.5\%\)
PPV \(= 67\%\)

\(^a\) Table number in original article.

\(^b\) In the Tockman et al. report, the specificity 5/27 = 0.8148 had be rounded up to 82%.

**Table 2** (Table 5) Chinese miners study

<table>
<thead>
<tr>
<th>Samples</th>
<th>Lung cancer</th>
<th>No lung cancer</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test +</td>
<td>37</td>
<td>17</td>
<td>54</td>
</tr>
<tr>
<td>Test -</td>
<td>8</td>
<td>32</td>
<td>40</td>
</tr>
<tr>
<td>Totals</td>
<td>45</td>
<td>49</td>
<td>94</td>
</tr>
</tbody>
</table>

Sensitivity \(= 82\%\)
Specificity \(= 65\%\)
Proportion of cases \(= 48\%\)
PPV \(= 69\%\)

\(^a\) Table number in original article.
On the basis of the report by Tockman et al., Smith concluded that hnRNP A2/B1 overexpression was “a powerful predictor of early subclinical cancer in high risk groups” (3). Given that the data indicate that in two high-risk populations the PPVs were <10%, one needs to reconsider the applicability of such a screening test, at least in isolation, to the general population or even “high-risk” populations, as they are presently defined.

The discussion presented here highlights the innate dilemma that confronts screening and early detection of lung cancer. Although sophisticated molecular tests may be highly sensitive and specific in detecting biomarkers in the laboratory, because there are no known biomarkers that are necessarily and sufficiently (hence, exclusively) associated with lung cancer, biomarker screening for lung cancer will always be less sensitive and specific for predicting invasive cancer than estimated in laboratory validation studies.

The PPV is a relevant parameter that should be estimated for the target population. However, for uncommon diseases, obtaining a test with a high PPV so as to make screening economical and minimize the worry and risk associated with false positives is problematic. Given disease frequencies of 2%, a screening test with 90% sensitivity and specificity will result in a PPV of 16%. At this phase of our understanding, for early detection to succeed, we need to identify and characterize groups with still higher risk and develop screening tests with higher sensitivities and specificity and consider applying them jointly or applying different sets of tests to identify unique lung cancer subtypes.

C. Martin Tammemagi
Josephine Ford Cancer Center
Henry Ford Health System
1 Ford Place, 5c
Detroit, Michigan 48202-3450

References

Reply
In the wake of the recent report on the promise of spiral CT1 for lung cancer screening, the whole field of early lung cancer detection is receiving a new level of scrutiny (1). From that perspective, we welcome the opportunity to address the important issue raised by Dr. Tammemagi. Two years ago, we published a preliminary report of early lung cancer detection using an immunocytochemical assay that recognized hnRNP A2/B1 overexpression in shed bronchial epithelial cells recovered in the sputum (2). The letter of Tammemagi raises an issue long understood by the screening community. Specifically, the prevalence of a disease in the population being screened strongly influences the predictive value of a positive or negative test. This point, which was elegantly described by Vecchio (3) in 1966, is correct but leads only to a hypothetical “straw-man” generic to how one conducts screening for an uncommon condition. We believe the central points are as follows.

First, three laboratories using clinical material from over 6000 subjects have independently validated the utility of hnRNP A2/B1 overexpression as an early detection marker for lung cancer (2, 4, 5). Two independent laboratories have published their results confirming the detection of the hnRNP A2/B1 epitope in sputum from individuals in advance of the clinical diagnosis of lung cancer (2, 4). This achievement, along with the report that helical CT can detect complementary preclinical lung lesions (1), raises the possibility that new screening techniques have progressed sufficiently to detect lung cancer at a point in its natural history when current treatments might allow for a lower mortality from this disease.

In his letter, Tammemagi assumes that hnRNP A2/B1 detection will be successful. He raises concern that its application in clinical populations will not result in a positive predictive value as we reported. The positive predictive value we report is based on sampling all of the cases and a subset of controls. As we clearly described, these preliminary results represent observations on individuals who had a definite clinical outcome after a follow-up of 1 year. The trial is being completed so that we

Table 3  SPLC study

<table>
<thead>
<tr>
<th>Populations</th>
<th>Lung cancer</th>
<th>No lung cancer</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test +</td>
<td>10</td>
<td>108</td>
<td>118</td>
</tr>
<tr>
<td>Test −</td>
<td>3</td>
<td>474</td>
<td>477</td>
</tr>
<tr>
<td>Totals</td>
<td>13</td>
<td>582</td>
<td>595</td>
</tr>
</tbody>
</table>

| Sensitivity       | 77%         |
| Specificity       | 81%         |
| Proportion of cases | 2.2%       |
| PPV               | 8.5%        |

* In the Tockman et al. report, the specificity 5/27 = 0.8148 had been rounded up to 82%.

Table 4  Chinese miners study

<table>
<thead>
<tr>
<th></th>
<th>Lung cancer</th>
<th>No lung cancer</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test +</td>
<td>47</td>
<td>2161</td>
<td>2208</td>
</tr>
<tr>
<td>Test −</td>
<td>10</td>
<td>4067</td>
<td>4077</td>
</tr>
<tr>
<td>Totals</td>
<td>57</td>
<td>6228</td>
<td>6285</td>
</tr>
</tbody>
</table>

| Sensitivity | 82%          |
| Specificity | 65%          |
| Proportion of cases | 0.9%       |
| PPV         | 2.1%         |

Received 1/13/00; accepted 2/22/00.

1 The abbreviation used is: CT, computed tomography.
can determine the true performance characteristics of the assay. Furthermore, Tammemagi chooses to focus his criticism on the difficulty of relying on positive predictive values to evaluate screening tests for uncommon conditions such as cancer. He ignores our report of hnRNP A2/B1 sensitivity (77–82%) and specificity (65–82%), which exceed those of prostate cancer screening with the routine prostate-specific antigen test (6, 7).

Second, the detection of hnRNP A2/B1 is the first molecular lung cancer screening technique to be prospectively tested in a population trial. The last techniques to be tested for lung cancer screening, routine sputum cytology and chest radiography, failed due to their lack of sensitivity (8). Under population-screening conditions, we reported an 8-fold improvement in sensitivity compared to routine sputum cytology blindly evaluated at the same time (2).

We are deeply committed to determining which positive tests are falsely positive and to avoid detecting lung cancers that do not grow enough to threaten the life of the patient (so-called “overdiagnosis”). For that reason, in our trials with sputum immunodetection, we are correlating marker expression with clinically validated end points to ensure that we identify only clinically significant lung cancer. Overdiagnosis did not appear to be a significant issue with sputum-diagnosed lung cancer cases in the previous National Cancer Institute-sponsored lung cancer trial, but these issues have important implications for the strategy of screening with molecular airway markers.

A single lung cancer screening test with near perfect sensitivity and specificity remains hypothetical. Just as diagnoses are made for other clinical conditions, we propose that sensitive lung cancer screening tests be followed by more specific tests until a diagnosis is reached. Whether this triage will take the form of initial hnRNP A2/B1 screening followed by helical CT examination, as suggested by Smith’s editorial (9), remains to be determined. Comparative testing of these two techniques is being conducted at Weill Medical College of Cornell University, the Mayo Clinic, and the H. Lee Moffitt Cancer Center to address these issues. Furthermore, innovative approaches to the management of the early, screen-identified lesions may be associated with low enough cost and morbidity that through time, concerns about overdiagnosis may be assuaged.

Considering that lung cancer is the world’s greatest cancer killer of both men and women, that the high mortality of lung cancer has persisted for more than three decades, and that lung cancer is the only major cancer for which screening remains unavailable, we consider it a mistake to wait for the ideal screening test. Rather, we propose clinical trials of rational management strategies to triage individuals with positive molecular airway markers for further testing with progressively more specific evaluations. We greatly appreciate Dr. Tammemagi’s concern with regard to the proper conduct of a trial to validate a screening approach. For the reasons outlined above, we do not feel that his criticism was relevant to our previous report. Nevertheless, because lung cancer early detection represents a distinct and important screening challenge, we thank the editors for this opportunity to review these topical issues.

Melvyn S. Tockman
James L. Mulshine
Steven Piantadosi
H. Lee Moffitt Cancer Center Research Institute
12902 Magnolia Drive
Tampa, Florida 33612

References

Editor’s Note
Although hnRNP A2/B2 shows promise for detecting preclinical lung cancer, it is important not to exaggerate its potential benefit. Positive predictive value is a relevant concept, and all parties agree that Dr. Tammemagi’s calculations are correct.

C. Martin Tammemagi


Updated version  Access the most recent version of this article at: http://clincancerres.aacrjournals.org/content/6/4/1601

Cited articles  This article cites 10 articles, 4 of which you can access for free at: http://clincancerres.aacrjournals.org/content/6/4/1601.full#ref-list-1

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.