Primary Lung Cancer vs Metastatic Breast Cancer

A Probabilistic Approach

Robin T. Vollmer, MD

Abstract

In this study, a mathematical and probabilistic model is used to study the probability that a lung tumor is a primary vs a metastasis from cancer of the breast. The model uses information from immunohistochemical stains for thyroid transcription factor (TTF)-1, mammaglobin, p63, and estrogen receptor and epidemiologic data about primary lung and metastatic breast cancers in women. The results demonstrate that these 4 stains can yield nearly certain diagnoses in approximately 80% of tumors falling into the pool of this differential diagnosis. Nevertheless, uncertainty of diagnosis remains for the 19% of tumors in the pool that are negative for TTF-1, mammaglobin, and p63.

One of the most vexing differential diagnoses for pathologists and clinicians is primary lung cancer vs breast cancer metastatic to the lung. When a woman who has been a smoker and who has also had breast cancer has a non–small cell carcinoma in her lung, choosing between lung cancer and breast cancer metastasis is important because treatments for these two tumors differ. The difficulty is due to how lung cancer and breast cancer comprise the 2 most common fatal malignancies in women, to how tumors from both sites comprise similar appearing cells, and to how both tumors commonly stain positively for cytokeratin (CK) 7 and negative for CK20. It is due to how many lung carcinomas do not stain for thyroid transcription factor (TTF)-1, to how some breast cancers do not stain for estrogen receptor (ER), and finally to how these lung tumors are increasingly being sampled via small bronchial biopsies, bronchial cytology specimens, and transthoracic needle biopsies, all of which provide limited tissue for doing anything other than diagnosing them simply as non–small cell carcinomas. The purpose of this study was to examine this differential diagnosis using concepts from probability theory, collected information from previous studies of immunohistochemical analysis, and epidemiologic results from studies of non–small cell lung carcinomas (NSCLCs) and breast carcinomas metastatic to the lung.

Materials and Methods

Bayes Probabilistic Approach

In what follows, I concentrate on women who have been smokers, who have had a prior carcinoma of the breast, and who now have a non–small cell carcinoma in a small biopsy...
sample of the lung. In addition, I assume that the sample of tumor provides sufficient tissue for several immunohistochemical stains that demonstrate that the tumor stains for CK7 but not for CK20. Finally, for the sake of simplicity, I assume that there is no history of other malignancies, such as carcinomas of the colon, ovary, or pancreas, that might present with metastases to the lung. Although this scenario may appear restrictive, it is one we routinely face when dealing with the differential diagnosis of a new primary lung cancer vs a metastasis from breast cancer.

Thus, in what follows, we assume there are 2 possible origins for the tumor—primary lung cancer vs metastatic breast cancer. In this circumstance, knowing 1 of 2 conditional probabilities will be most helpful. The first is the probability of metastatic breast cancer, given a specific staining pattern for the tumor. If we symbolize the staining pattern as SP and the diagnosis of metastatic breast cancer as BR, then the conditional probability is written as \( P(BR | SP) \). The second probability is the probability that the tumor is a primary lung cancer, here symbolized as L, given the same staining pattern. This probability is written as \( P(L | SP) \), and in this binary diagnostic situation, it equals \( 1 - P(BR | SP) \).

By using a previously published approach and Bayes theorem, \( P(BR | SP) \) can be rewritten as\(^{14} \):\(^\text{Equation 1}\)

\[
P(BR | SP) = \frac{P(SP | BR) \times P(BR)}{P(SP | BR) \times P(BR) + P(SP | L) \times P(L)}
\]

Here, \( P(SP | BR) \) is the conditional probability that a metastatic breast cancer has the staining pattern, and \( P(SP | L) \) is the conditional probability that a primary lung cancer has the staining pattern. \( P(BR) \) and \( P(L) \) symbolize the prior probabilities of metastatic breast cancer to the lung vs a primary lung cancer, and, in this study, where we assume that the tumor found in the lung must be metastatic breast cancer or a primary lung cancer, \( P(L) \) will be taken to be equivalent to \( 1 - P(BR) \).

\( \text{Table I} \) provides 2 types of information necessary for estimating \( P(BR | SP) \) from Equation 1: the relative probabilities of subtypes of NSCLC in women and the sensitivities of stains for the tumors in the differential diagnosis. For lung cancers, \( P(SP | L) \) must be obtained as a weighted sum of the probabilities of the staining pattern for each type, ie, \( P(SP | type) \) over the 4 types of tumors after multiplying by the relative incidence of each, which here is symbolized as \( P(type) \). Consequently,

\( \text{Equation 2} \)

\[
P(SP | L) = \Sigma P(type) \times P(SP | type)
\]

Here, the summation is taken over the 4 types of NSCLCs. The probabilities of subtypes \( P(type) \) were obtained from Surveillance, Epidemiology, and End Results (SEER) data involving more than 57,000 women with primary lung cancer (http://seer.cancer.gov/csr/1975_2004/).

Staining for 5 markers was considered: TTF-1, mammmaglobin, gross cystic disease fluid protein 15 (GCDFP), estrogen receptor protein (ER), and p63. Their sensitivities are given in Table 1 as probabilities of a positive stain, given a specific tumor type, and they were obtained mostly from the PathIQ ImmunoQuery Web site (https://immunoquery.pathiq.com/PathIQ/) and from Sasaki et al,\(^8\) Bhargava et al,\(^15\) and Nadji et al.\(^16\) Altogether, the probabilities for immunohistochemical results in Table 1 were derived from more than 11,000 tumor-stain encounters. For the sake of simplicity and

<table>
<thead>
<tr>
<th>Tumor Type</th>
<th>Relative Incidence</th>
<th>TTF-1</th>
<th>MAM</th>
<th>GCDFP</th>
<th>ER</th>
<th>p63</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenocarcinoma</td>
<td>.533</td>
<td>.77</td>
<td>.02</td>
<td>.06</td>
<td>.07</td>
<td>.31</td>
</tr>
<tr>
<td>Squamous carcinoma</td>
<td>.194</td>
<td>.07</td>
<td>.01</td>
<td>0</td>
<td>.64</td>
<td>.97</td>
</tr>
<tr>
<td>Large cell carcinoma</td>
<td>.0589</td>
<td>.44</td>
<td>0</td>
<td>0</td>
<td>.94</td>
<td>.47</td>
</tr>
<tr>
<td>Non-small cell, not otherwise specified</td>
<td>.214</td>
<td>.58</td>
<td>.06</td>
<td>.56</td>
<td>.75</td>
<td>.70</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>—</td>
<td>.63</td>
<td>.07</td>
<td>.75†</td>
<td>.06</td>
<td></td>
</tr>
</tbody>
</table>

ER, estrogen receptor; GCDFP, gross cystic disease fluid protein; MAM, mammaglobin; TTF-1, thyroid transcription factor 1.

\(^*\) Tumor types apply to invasive lung cancers, and the relative incidences were obtained from SEER data (http://seer.cancer.gov/csr/1975_2004/). (For breast cancer, this relative incidence is not needed.) The probabilities of tumor staining positive were obtained from the PathIQ ImmunoQuery Web site (https://immunoquery.pathiq.com/PathIQ/) and from Sasaki et al,\(^8\) Bhargava et al,\(^15\) and Nadji et al.\(^16\)

\(^†\) For tumors that were negative for MAM, the probability for positive ER was adjusted to .71, based on statistical dependence between results for MAM and ER as demonstrated by Sasaki et al and Bhargava et al.
because there are few available data regarding combinations of stains, I will, for the most part, assume that the staining results for combinations of 2 or 3 of these stains are approximately and statistically independent. Thus, if the staining pattern is TTF-1−, ER+, and p63+, then the probability of this staining pattern in breast cancer, P(SP | BR), will be taken to be:

\[ P(SP | BR) = (1 - P(TTF)) \times P(ER) \times P(p63) \]

The exception will be for staining for ER, given that staining for mammaglobin is negative, because staining for these 2 markers has been demonstrated to be statistically dependent.8,15 Thus, using the combined results of these 2 studies, the probability that ER is positive in a mammaglobin– case was reduced from .75 to .71.

Estimating P(BR), the Prior Probability of Breast Cancer Metastatic to the Lungs

The approach taken here will consider not only immuno histochemical results, but also epidemiologic data derived from women with breast or lung cancer. This is because the decision about the likelihood of a breast cancer metastasis to the lungs vs a primary lung cancer must consider the underlying frequencies of both diagnoses. For example, SEER data give us reasonably accurate estimates of the number of women expected to have lung cancer each year, which, in data give us reasonably accurate estimates of the number of lying frequencies of both diagnoses. For example, SEER the lungs vs a primary lung cancer must consider the under-

The results are summarized in Table 2. The first column gives the staining pattern. The second column provides the estimated P(BR | SP) for the staining pattern and is obtained from Equation 1. The third column provides the information content, ie, 1 − S, for the staining pattern’s ability to discriminate a metastatic breast cancer from a primary lung cancer.21 The information content for a binary outcome ranges from 0 (no relevant information) to 1 (maximum information). Consider next the total pool of women with metastatic breast cancer to the lungs or primary lung cancer. The fourth column provides the estimated fraction in this pool of tumors expected to have the staining pattern.

For example, row 1 deals with tumors that stain positive for TTF-1. The value of 0 for P(BR | SP) implies that metastatic breast cancer is very unlikely and that a primary lung cancer is nearly certain (ie, P(L | SP) = 1). The information content of 1 also implies near certainty of outcome. The expected frac-

| Staining Pattern (SP) | P(BR | SP) | 1 − S | Fraction of Tumors |
|-----------------------|----------|-------|-------------------|
| (1) TTF-1+            | .00      | 1.0   | 0.42              |
| (2) TTF-1+, MAM+      | .98      | .02   | 0.17              |
| (3) TTF-1+, GCDFP+    | .96      | .04   | 0.15              |
| (4) TTF-1+, ER+       | .63      | .37   | 0.22              |
| (5) TTF-1+, MAM−, p63+| .03      | .97   | 0.22              |
| (6) TTF-1+, MAM−, p63−| .48      | .52   | 0.19              |
| (7) TTF-1+, MAM−, p63−, ER+ | .76      | .24   | 0.09              |
| (8) TTF-1+, MAM−, p63−, ER− | .26      | .74   | 0.10              |

ER, estrogen receptor; GCDFP, gross cystic disease fluid protein; MAM, mammaglobin; TTF-1, thyroid transcription factor 1.

Table 2

| Staining Pattern (SP) | P(BR | SP) | 1 − S | Fraction of Tumors |
|-----------------------|----------|-------|-------------------|
| (1) TTF-1+            | .00      | 1.0   | 0.42              |
| (2) TTF-1+, MAM+      | .98      | .02   | 0.17              |
| (3) TTF-1+, GCDFP+    | .96      | .04   | 0.15              |
| (4) TTF-1+, ER+       | .63      | .37   | 0.22              |
| (5) TTF-1+, MAM−, p63+| .03      | .97   | 0.22              |
| (6) TTF-1+, MAM−, p63−| .48      | .52   | 0.19              |
| (7) TTF-1+, MAM−, p63−, ER+ | .76      | .24   | 0.09              |
| (8) TTF-1+, MAM−, p63−, ER− | .26      | .74   | 0.10              |

ER, estrogen receptor; GCDFP, gross cystic disease fluid protein; MAM, mammaglobin; TTF-1, thyroid transcription factor 1.

The value of 1 − S range from no information (0.0) to maximal information at a value of 1.0. Fraction of tumors refers to the fraction of lung tumors expected to have the particular staining pattern, when the total pool comprises women with metastatic breast cancer to the lungs or primary lung cancer.

Results

In the results are 8 staining patterns: (1) positive staining of tumor for TTF-1; (2) negative staining for TTF-1 and positive staining for mammaglobin; (3) negative staining for TTF-1 and positive staining for GCDFP; (4) negative staining for TTF-1 and positive staining for ER; (5) negative staining for TTF-1, negative staining for mammaglobin, and positive staining for p63; (6) negative staining for TTF-1, mammaglobin, and p63; (7) negative staining for TTF-1, mammaglobin, and p63 with positive staining for ER; and, finally, (8) negative staining for TTF-1, mammaglobin, p63, and ER.
together that nearly all such tumors will be metastatic breast cancers. The fraction of 0.17 implies, however, that just 17% of the pool will have this pattern. The results of rows 3 and 4 in Table 2 demonstrate that, when combined with negative TTF-1, staining for GCDFP and ER provide less information about tumor origin than the combination of TTF-1 and mammaglobin. In fact, the information content of the combination of negative staining for TTF-1 and positive staining for ER is very nearly zero.

Altogether, the results in rows 2 to 4 demonstrate that in a pairwise combination with TTF-1, mammaglobin provided more information than combinations of TTF-1 with either GCDFP or ER. For this reason and because mammaglobin, GCDFP, and ER are statistically dependent, I turned next to an analysis of p63 in combination with negative stains for TTF-1 and mammaglobin, and the results appear in rows 5 and 6. Row 5 demonstrates that a tumor staining negative for TTF-1 and mammaglobin and staining positive for p63 should most probably be a primary lung tumor (ie, P(L | SP) = 1 – 0.03 = 0.97). The information content for this staining combination is high, and approximately 22% of tumors in the diagnostic pool should have this staining pattern. By contrast, row 6 demonstrates that a staining pattern that is negative for TTF-1, mammaglobin, and p63 provides no information about the origin of the tumor (ie, 1 – S = 0), and approximately 19% of the tumors are expected to fall into this nondiagnostic category.

To resolve the limited information that negative staining for TTF-1, mammaglobin, and p63 provided, I next turned to ER. Because mammaglobin and ER are statistically dependent, I had to adjust the probability of a mammaglobin–breast tumor staining for ER from 0.75 to 0.71, and this adjustment was based on data published in 2 prior studies. The results of the analysis including ER appear in rows 7 and 8. They demonstrate that the 19% of tumors in row 6 can be further subdivided into approximately 9% that should be ER+ and 10% that should be ER–. The probability that the ER+ tumors are breast metastases should be approximately 0.76, and the probability that the ER– tumors are primary lung cancers should be approximately 0.74 (ie, P(L | SP) = 1 – 0.26 = 0.74). Nevertheless, results demonstrate that the information content for these 2 staining patterns with added ER is limited. In other words, for this expected 19% of tumors, there remains significant uncertainty about a metastasis vs a primary lung cancer when one considers just immunohistochemical and epidemiologic results.

Finally, Figure 1 demonstrates a stepwise approach to the differential diagnosis of metastatic breast cancer vs a primary lung tumor, given the presence of a non–small cell carcinoma in the lung of a woman who has had breast cancer and has been a smoker. The suggested stains, branch points, and probabilities derive from the Bayes approach that uses published results about how breast cancers and lung cancers stain for these markers and epidemiologic data about expected breast cancer deaths and new primary cancers of the lung. In the boxes of Figure 1, Pb stands for the Bayes probabilities derived from Equation 1. In the boxes marked as Breast Metastasis, Pb is the Bayes probability P(BR | SP). In the boxes marked as Lung Ca, Pb is the Bayes probability P(L | SP) [ie, 1 – P(BR | SP)].

**Discussion**

The process of reaching a diagnosis in surgical pathology can be viewed as deterministic or probabilistic. If deterministic, then the histologic and clinical features lead to a conclusive diagnosis. If probabilistic, then those features lead to a list of possible diagnoses, perhaps ordered according to their probabilities of being observed. Whereas many biopsies of tumors in their primary sites are deterministic, biopsies of metastatic tumors are often probabilistic. The foregoing analyses demonstrate that even in the difficult differential diagnosis of a metastatic breast cancer to the lung vs a primary lung cancer, 3 stains combined with epidemiologically derived data

---

**Figure 1** A stepwise scheme for evaluating the differential diagnosis of metastatic breast cancer to the lung vs a primary lung cancer (Lung Ca) based on the combination of epidemiologic data and stains for thyroid transcription factor 1 (TTF-1), mammaglobin (MAM), p63, and estrogen receptor (ER). Entries for Pb in the boxes give the Bayes-derived probability of a lung primary in boxes marked Lung Ca or probability of a breast metastasis in boxes marked as such. Ca, carcinoma.
can yield very nearly deterministic diagnoses. For example, the Bayes probability approach demonstrated that nearly 80% of tumors in this differential diagnosis should be diagnosable by using TTF-1, mammaglobin, and p63. The residual 19% can be partly resolved by using ER, but the diagnosis for this group should be expected to remain probabilistic.

Although the preceding results rely on a long-standing mathematical model whose logic appears to be solid, the results in Table 2 and Figure 1 equally depend on a large amount of previously collected data, and these data may harbor errors and biases. For example, the autopsy studies used comprised a total of 389 cases of fatal breast cancers—numbers difficult to reproduce in current times. The immunohistochemical results relied on more than 11,000 tumor stain encounters. Nevertheless, the autopsy data and immunohistochemical results came from published reports, so that they could involve errors and biases involved in investigator-designed studies and in the overall publication process. The relative incidence of subtypes of NSCLC and the relative frequencies of lung cancer in women and the mortality of breast cancer came from SEER data comprising more than 100,000 cases, but even this large amount of data could include errors due to sampling biases. Furthermore, other factors not considered here are undoubtedly important to the differential diagnosis, and these include the presence of a single lung mass vs multiple and bilateral lung masses; known foci of metastatic breast cancer to the liver, brain, or bone marrow; and the time lapse since the diagnosis of breast cancer. For these reasons, the results of Table 2 and Figure 1 should be viewed as providing rough guidelines for the probabilistic diagnosis of metastatic breast cancer vs primary carcinoma of the lung. The relative values in Table 2 and Figure 1 may be more important than their specific numbers.

References