Test Characteristics of α-Fetoprotein for Detecting Hepatocellular Carcinoma in Patients with Hepatitis C
A Systematic Review and Critical Analysis

Samir Gupta, MD; Stephen Bent, MD; and Jeffrey Kohlwes, MD, MPH

Background: Patients with hepatitis C virus (HCV) are at increased risk for hepatocellular carcinoma. Although serum α-fetoprotein (AFP) is often used to detect hepatocellular carcinoma in HCV-infected individuals, its utility is unclear.

Purpose: To define the test characteristics of AFP for the diagnosis of hepatocellular carcinoma in patients with HCV.

Data Sources: MEDLINE search from 1966 to December 2002 for English- and non–English-language articles examining the test characteristics of AFP for identifying hepatocellular carcinoma.

Study Selection: Articles were included if they reported the sensitivity and specificity of AFP for detecting hepatocellular carcinoma in patients with HCV. Articles were excluded if the cause of hepatitis was ambiguous or if 50% or more of the study patients did not have HCV.

Data Extraction: Relevant articles were evaluated for quality of evidence; test characteristics were abstracted and calculated.

People with hepatitis C virus (HCV) have a 2% annual risk and a 7% to 14% five-year risk for hepatocellular carcinoma (1–3), a tumor with an estimated median survival duration of 4.3 to 20 months after diagnosis (4–7). Some studies suggest a possible survival advantage when small tumors are detected (8, 9), but no randomized, controlled trials of screening for hepatocellular carcinoma in patients with HCV have been conducted.

Although the National Cancer Institute currently recommends against screening for hepatocellular carcinoma (10), many physicians currently screen high-risk populations with various strategies, including serum α-fetoprotein (AFP), ultrasonography, and computed tomography (11). The use of AFP, a tumor marker variably secreted by hepatocellular carcinomas, to detect these tumors has been widely debated (12–14). Many conclude that AFP is not a useful diagnostic test (12, 15), but AFP continues to be commonly used (11). To determine a summary estimate of the test characteristics of AFP for detecting hepatocellular carcinoma in patients with HCV, we conducted a systematic review.

Methods
Study Search Protocol
We performed a MEDLINE search from 1966 through December 2002 for English- and non–English-language articles using the following search terms: hepatitis C, hepatocellular carcinoma, screening, diagnosis, alpha-fetoprotein, sensitivity, and specificity. Bibliographies of all reviewed articles were searched to identify additional relevant titles. Titles that mentioned hepatocellular carcinoma or HCV and screening were identified for abstract review. Abstracts that described the use of AFP as a diagnostic or screening test for hepatocellular carcinoma were marked for full article review.

Inclusion and Exclusion Criteria
Study designs accepted for analysis included randomized, controlled trials, cohort studies, or case–control studies that used AFP to detect hepatocellular carcinoma in HCV-infected patients with or without cirrhosis. We required that the authors report sensitivity and specificity for the use of serum AFP (or data sufficient to calculate these test characteristics) and that they identify some gold standard for diagnosis.

Computed tomography, magnetic resonance imaging, histopathology, and disease-free time greater than 2 years were considered adequate gold standards. Ultrasonography was not considered an adequate gold standard because its sensitivity for hepatocellular carcinoma is controversial (12, 14, 16, 17).

Studies were excluded from analysis if the cause of viral hepatitis was unclear, if at least 50% of the study patients did not have HCV, and if the same data were presented in a separate article by the same investigators.

Data abstracted were study design, cause of hepatitis, whether the AFP test was used for diagnosis or screening, type of gold standards used, percentage of the study sample...
with cirrhosis, and reported sensitivity and specificity of AFP for detecting hepatocellular carcinoma.

### Analysis

To grade the quality of evidence for use of serum AFP as a screening test for hepatocellular carcinoma in patients with HCV, we independently determined study design, whether application of the gold standard for each study was blinded to AFP result, whether the patient selection was independent, the type of gold standard implemented, and presence of partial verification bias. Disparity in grade was resolved by discussion and consensus among all three authors.

### Results

A total of 1239 titles were identified, 55 relevant abstracts were reviewed, and 18 articles were identified as potentially relevant. Five studies met all inclusion criteria and were included in the analysis (15, 18–21). Of the 18 potentially relevant articles, 5 were excluded because they were uncontrolled case series (22–26), 6 were excluded because study patients did not have HCV (27–32), 1 was excluded because it did not identify the cause of hepatitis in all study patients (33), and 6 were excluded because they did not provide both sensitivity and specificity data for AFP in their study sample (8, 34–38). Characteristics of the 5 studies meeting all inclusion criteria and no exclusion criteria are shown in Table 1 (15, 18–21).

Two studies were prospective cohort studies (15, 18) and 3 were case–control studies (19–21). One study (15) universally applied an acceptable gold standard test to both case-patients and controls, and each study used a different gold standard.

Table 2 presents abstracted sensitivity and specificity data and abstracted or calculated positive and negative likelihood ratios for the diagnosis of hepatocellular carcinoma at an AFP cutoff value of 20 μg/L. This cutoff value was chosen because each included article provided data for a cutoff value of 20 μg/L, and an AFP level of 20 μg/L is considered a level that prompts further testing (19). Other cutoff values were not reported in every article. Exclusive data for patients with HCV were available for all studies but one (18); for the latter study, only combined data for patients with HCV and hepatitis B virus were available.

Sensitivity of AFP levels higher than 20 μg/L ranged from 41% to 65%, while specificity ranged from 80% to 94%. Positive likelihood ratios for AFP levels higher than

### Table 1. Characteristics of Included Studies*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Peng et al., 1999 (20)</td>
<td>Case–control</td>
<td>131 patients with HCV†</td>
<td>Angiography or biopsy</td>
<td>Ultrasonography</td>
<td>NR</td>
<td>Yes</td>
<td>Unknown</td>
</tr>
<tr>
<td>Cedrone et al., 2000 (18)</td>
<td>Prospective cohort</td>
<td>350 patients: 78% with HCV, 42% with cirrhosis</td>
<td>Ultrasonography</td>
<td>Ultrasonography</td>
<td>Yes</td>
<td>No gold standard test used‡</td>
<td>Yes</td>
</tr>
<tr>
<td>Tong et al., 2001 (15)</td>
<td>Prospective cohort</td>
<td>601 patients: 73% with HCV, 29% with cirrhosis</td>
<td>CT and biopsy</td>
<td>Ultrasonography and follow-up time</td>
<td>NR</td>
<td>Yes</td>
<td>Unknown</td>
</tr>
<tr>
<td>Trevisani et al., 2001 (21)</td>
<td>Case–control</td>
<td>340 patients; most patients with HCV‡</td>
<td>Pathology, “imaging,” or autopsy</td>
<td>Ultrasonography and follow-up time ≥ 6 mo</td>
<td>NR</td>
<td>Yes</td>
<td>Unknown</td>
</tr>
<tr>
<td>Nguyen et al., 2002 (19)</td>
<td>Case–control</td>
<td>312 patients with HCV and cirrhosis</td>
<td>Biopsy</td>
<td>Ultrasonography, CT, or MRI</td>
<td>NR</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* AFP = α-fetoprotein; CT = computed tomography; HCV = hepatitis C virus; MRI = magnetic resonance imaging; NR = not reported.
† Percentage of patients with cirrhosis was not reported.
‡ Ultrasonography applied to both case-patients and controls as gold standard test; therefore, partial verification bias does not apply.

### Table 2. Abstracted Test Characteristics of α-Fetoprotein Levels Higher than 20 μg/L for Detecting Hepatocellular Carcinoma*

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Sensitivity of AFP Level &gt; 20 μg/L (95% CI), %</th>
<th>Specificity of AFP Level &gt; 20 μg/L (95% CI), %</th>
<th>Positive Likelihood Ratio (95% CI)</th>
<th>Negative Likelihood Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peng et al., 1999 (20)</td>
<td>65 (58–71)</td>
<td>87 (79–93)</td>
<td>4.9 (3.0–8.0)</td>
<td>0.5 (0.3–0.5)</td>
</tr>
<tr>
<td>Cedrone et al., 2000 (18)</td>
<td>55</td>
<td>88</td>
<td>4.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Tong et al., 2001 (15)</td>
<td>41</td>
<td>94</td>
<td>6.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Trevisani et al., 2001 (21)</td>
<td>60</td>
<td>91</td>
<td>6.7</td>
<td>0.4</td>
</tr>
<tr>
<td>Nguyen et al., 2002 (19)</td>
<td>63 (56–70)</td>
<td>80 (73–86)</td>
<td>3.1†</td>
<td>0.5†</td>
</tr>
</tbody>
</table>

* AFP = α-fetoprotein.
† Data for patients with hepatitis C virus and hepatitis B virus analyzed together.
‡ Data for CIs are not available or calculable.
20 μg/L ranged from 3.1 to 6.8 and negative likelihood ratios ranged from 0.4 to 0.6.

Table 3 shows the sensitivity and specificity data for an AFP cutoff value higher than 200 μg/L, a value that is frequently reported to be specific for the diagnosis of hepatocellular carcinoma (19, 21). Four of the 5 studies reported sensitivity and specificity for this cutoff value. The range of specificities was very high at this cutoff value (99% to 100%), but the sensitivity was very low (20% to 45%).

### Table 3. Abstracted Test Characteristics for α-Fetoprotein Levels Higher than 200 μg/L for Detecting Hepatocellular Carcinoma*

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Sensitivity of AFP Level &gt; 200 μg/L (95% CI), %</th>
<th>Specificity of AFP Level &gt; 200 μg/L (95% CI), %</th>
<th>Positive Likelihood Ratio (95% CI)</th>
<th>Negative Likelihood Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peng et al., 1999 (20)</td>
<td>45 (38–52)</td>
<td>100 (97–100)</td>
<td>0.6 (0.5–0.6)</td>
<td>0.8</td>
</tr>
<tr>
<td>Cedrone et al., 2000 (18)†</td>
<td>20</td>
<td>99</td>
<td>0.7†</td>
<td>0.7†</td>
</tr>
<tr>
<td>Tong et al., 2001 (15)‡</td>
<td>NR</td>
<td>99</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Trevisani et al., 2001 (21)†</td>
<td>32 (25–39)</td>
<td>100 (100–100)</td>
<td>0.6 (0.5–0.6)</td>
<td>0.8</td>
</tr>
<tr>
<td>Nguyen et al., 2002 (19)</td>
<td>NR</td>
<td>NR</td>
<td>NR†</td>
<td>NR†</td>
</tr>
</tbody>
</table>

* AFP = α-fetoprotein; NR = not reported.  † When the reported specificity is 100%, the likelihood ratio is theoretically infinite.  ‡ Data for CIs are not available or calculable.

### Discussion

Our systematic review of the literature shows that the quality of evidence describing the characteristics of AFP as a diagnostic test for hepatocellular carcinoma in patients with HCV is limited. Three of the reviewed studies were case–control studies (19–21), which potentially overestimate the sensitivity and specificity of the test in question (39, 40). In contrast, cohort studies are less susceptible to bias because they are more likely to include patients with a varying spectrum of disease, particularly those patients who present more subly, and therefore more closely reflect the manner in which a test will be implemented in clinical practice (39).

Two studies (15, 20) may have partial verification bias, which occurs when the result of the test being evaluated (in this case, AFP or ultrasonography) influences the decision to administer the gold standard test (39, 40). This may falsely elevate sensitivity and specificity (39). Four of five studies (18–21) applied a gold standard of uncertain validity to both case-patients and controls, resulting in a possible underestimate of disease prevalence and an unknown ultimate effect on sensitivity and specificity. Blinding was not reported in four studies (15, 19–21) and may have affected interpretation of gold standard test results. Without systematic blinding, investigators may be more vigilant in applying gold standards to those patients with positive test results and thereby falsely elevate specificity (39, 40). Finally, four studies (15, 18, 20, 21) included patients with and without cirrhosis. Patients with cirrhosis have a higher risk for cancer (41) but commonly have elevated levels of AFP thought to be unrelated to hepatocellular carcinoma (42), leading to an unknown effect on sensitivity and specificity. Notably, one excluded study reported sensitivity of 80% and specificity of 95% for AFP levels higher than 10 μg/L applied to a subgroup of patients with histologically severe liver injury (35).

Given the significant concerns about the validity of the data generated by the studies reviewed, we could not calculate conclusive summary estimates of the sensitivity and specificity of AFP as a diagnostic test for hepatocellular carcinoma. The biases previously mentioned that affect the reported sensitivities and specificities tend to overestimate the utility of AFP as a diagnostic test, but to guide current interpretation of AFP in practice, we can consider the use of this test if these “best-case” estimates are true.

The most common use of AFP is to screen for hepatocellular carcinoma in asymptomatic patients with HCV. In this scenario, reported prevalence data indicate a pretest probability of hepatocellular carcinoma in patients with HCV is 5% to 12% (41, 43). Using a prevalence of 5% with the range of positive likelihood ratios for an AFP level higher than 20 μg/L (3.1–6.8), results in a post-test probability of 14% to 25%, while an AFP level lower than 20 μg/L results in a post-test probability of 2% to 3%. Although a post-test probability of 25% would prompt further work-up with imaging, a post-test probability of 2% is unlikely to be reassuring enough to preclude the use of other screening strategies, including ultrasonography or computed tomography.

The other common use of AFP involves the evaluation of patients presenting with one or more high-risk features, including a hepatic nodule (found incidentally or with a screening test) or decompensated liver failure. Data for AFP at higher cutoff values, such as an AFP level higher than 200 μg/L (Table 3), suggest that AFP, although not sensitive, can be highly specific for hepatocellular carcinoma. A low AFP level (<200 μg/L) would not be informative enough to stop further search for hepatocellular carcinoma, but an AFP level higher than 200 μg/L would strongly suggest that cancer is present, allowing for earlier counseling of a patient.

In addition to the quality of articles reviewed, our review is limited in two aspects. One of the articles reviewed did not exclusively analyze sensitivity and specificity data for patients with HCV independent of those patients with hepatitis B virus (18). Some studies have shown that AFP has different test characteristics in patients with hep-
attenuates B virus than in those with HCV (44, 45), but the sensitivities and specificities in this study (18) were within the range of values reported by the other studies analyzed, which included only patients with HCV. Finally, we accepted computed tomography as a gold standard for diagnosing hepatocellular carcinoma. Two studies of pretransplant triphasic helical computed tomography in detecting hepatocellular carcinoma in the explanted liver reported sensitivities of only 59% and 80% for computed tomography (29, 46). Two studies in our review may have underestimated the rate of hepatocellular carcinoma in their study sample because computed tomography scan played a key role in their diagnostic algorithm (15, 20). This potential bias has an unknown effect on the reported sensitivity and specificity.

**CONCLUSION**

Current studies examining the test characteristics of AFP for diagnosing hepatocellular carcinoma in patients with HCV have substantial methodologic limitations, making it difficult to define clear estimates of sensitivity and specificity for this test. With use of best-case estimates of sensitivity and specificity, AFP seems to have limited utility in identifying hepatocellular carcinoma in patients with HCV. A prospective study done with careful attention to limitation of bias is clearly needed to define whether any screening strategy can provide clinically important benefits.

From University of California, San Francisco, and Veterans Affairs PRIME program, Veterans Affairs Medical Center, San Francisco, California.

Grant Support: By PRIME Residency Program, University of California, San Francisco, San Francisco, California.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Jeffrey Kohlwees, MD, MPH, Department of Medicine #111, Veterans Affairs Medical Center, 4150 Clement Street, San Francisco, CA 94121; e-mail, samirguptaresearch@yahoo.com.

Current author addresses are available at www.annals.org.

**References**


www.annals.org

Review
Test Characteristics of α-Fetoprotein


Current Author Addresses: Drs. Gupta, Bent, and Kohlweg: Department of Medicine #111, Veterans Affairs Medical Center, San Francisco, 4150 Clement Street, San Francisco, CA 94121.