

HCC IN FOCUS

Current Developments in the Management of Hepatocellular Carcinoma

Section Editor: Robert G. Gish, MD

Current Status of the GALAD and BALAD Biomarker Models for Hepatocellular Carcinoma



Lewis R. Roberts, MB ChB, PhD
Professor of Medicine
Division of Gastroenterology and Hepatology
Mayo Clinic
Rochester, Minnesota

G&H How has the role of biomarkers evolved in hepatocellular carcinoma?

LR Perhaps the most common way that biomarkers are being used is to aid in the surveillance of patients who have cirrhosis and are at risk for hepatocellular carcinoma (HCC). The biomarker that is most commonly used in this context is α -fetoprotein (AFP), which is expressed in the liver during the development of this organ but then turned off after birth. AFP was probably first used as a biomarker for cancer in the late 1960s and early 1970s when researchers found that the protein was re-expressed in cancers of the liver.

For many years, AFP was used by itself in studies. When doctors started to recognize that patients with cirrhosis were at particular risk for HCC, they began to use ultrasound in combination with AFP to try to maximize their ability to detect cancers. However, AFP began to fall out of favor around the time that many HCCs were being caused by hepatitis C. Hepatitis C is somewhat unique as a cause of cirrhosis in that it has a relapsing nature, which means that patients go through cycles when the viral infection has more activity, followed by periods when there is less activity. During the periods of heightened viral activity, there is increased damage and death to liver cells. Regeneration of these liver cells is associated with an increase in AFP production by the immature liver cells. Therefore, the AFP of many patients with hepatitis C fluctuates, making it difficult to determine whether the

patients are truly developing cancer or simply undergoing a cycle of liver regeneration. Doctors in Europe and North America (where there was a lot of hepatitis C) began to believe that AFP was not helpful, unlike doctors in the rest of the world (where there was a lot of hepatitis B, which did not cause such profound fluctuations in AFP). US and European guidelines deemphasized AFP, recommending that doctors stop using it and use ultrasound alone.

However, over the past 10 to 15 years, there has been an increase in HCC associated with obesity and diabetes, causing fatty liver disease and cirrhosis. There has been a corresponding decrease in the performance of ultrasound in surveillance for HCC because this modality relies on the ability to see through to the back of the liver when the probe is placed on the skin. This can be difficult in patients with central obesity. Also during this period, better treatments were being developed, in particular for hepatitis C but also for hepatitis B. Researchers found that when hepatitis C is treated, AFP tends to normalize and become more consistently low, without significant fluctuation. As previously mentioned, this means that increases in AFP are more likely to reflect true development of cancer rather than episodes of liver regeneration. Thus, there has been a recent resurgence in the use of AFP as a biomarker and an understanding that, particularly in patients with a high body mass index, ultrasound is not as sensitive for the detection of new masses in the cirrhotic liver as previously thought. Recent guidelines reflect the

recognition that ultrasound and AFP are likely best used in combination.

G&H How and why were the GALAD and BALAD biomarker models developed for HCC?

LR The GALAD model (which includes gender, age, Lens culinaris agglutinin-reactive AFP [AFP-L3], total AFP, and des- γ -carboxyprothrombin [DCP]) and the

... neither AFP-L3 nor DCP seem to perform significantly better than AFP alone; it is in combination, especially when adding in gender and age to develop the GALAD score, that the biomarkers perform substantially better than AFP alone.

BALAD model (which includes bilirubin, albumin, AFP-L3, AFP, and DCP) were developed in part because of the recognition that both AFP and ultrasound were imperfect. During the period of biomarker development in the 1960s to 1980s, it was recognized that there was a subfraction of AFP capable of binding to lectin, which became designated as AFP-L3. It was found that this fraction tended to be more specific for HCC. Around this time, investigators studying abnormal prothrombin at the National Institutes of Health identified DCP, a modified prothrombin protein also known as PIVKA-II (protein induced by vitamin K absence II), which was found to be expressed by HCCs and also appears to be increased in patients with vitamin K deficiency.

Investigators in Japan sought to determine whether AFP-L3 and DCP could be used as biomarkers for HCC. They found that in some patients with hepatitis C–induced liver cirrhosis without cancer, AFP was nonspecifically elevated, but AFP-L3 would typically be low. If both AFP and AFP-L3 were elevated, there was a much higher probability that the patient had cancer. Investigators also found that patients who have HCC do not always make AFP. Even among patients with fairly advanced HCC, only 60% to 70% have a high AFP. On the other hand, some of the patients with HCC but a negative AFP have a high DCP. Thus, to be able to identify as many patients

with cancer as possible, investigators proposed using the biomarkers AFP, AFP-L3, and DCP together.

To develop the GALAD model, Dr Philip Johnson and colleagues in the United Kingdom and other parts of Europe and in Japan and other parts of Asia examined the ability of a combination of patient demographic information and biomarker levels to better predict the presence of cancer. The variables with the greatest ability to predict the presence of HCC were the patient's gender and age and the biomarkers AFP, AFP-L3, and DCP.

The BALAD model was developed by Dr Hidenori Toyoda and colleagues in Japan to predict the survival of patients with HCC. Most of these patients have underlying liver disease (eg, hepatitis C or fatty liver disease), and HCC subsequently develops. Thus, when developing the BALAD model, bilirubin and albumin, which are measures of the underlying liver disease, were combined with AFP, AFP-L3, and DCP, which are measures of HCC, to generate a score that could predict patient survival after HCC develops. A modified version of the model, BALAD-2, was developed by Dr Johnson and colleagues to classify patients into different groups to predict their survival.

G&H What are the advantages of these biomarker models?

LR The main advantages of the models are that they provide additional information beyond what can be obtained from any one biomarker and also provide a way to easily integrate all of that information into a score that is more accessible for physicians. A physician who has all of this information might try to integrate it intuitively, but the models provide a way to easily calculate the probability that the patient has HCC or calculate the patient's probability of survival.

G&H What are the main challenges or limitations of these models?

LR The main challenge involves establishing a sufficient evidence base to achieve more widespread use of the biomarkers and the models. For example, the AFP-L3 and DCP tests are currently approved by the US Food and Drug Administration to help determine a patient's risk of HCC, rather than specifically to use as surveillance biomarkers for screening. However, neither AFP-L3 nor DCP seem to perform significantly better than AFP alone; it is in combination, especially when adding in gender and age to develop the GALAD score, that the biomarkers perform substantially better than AFP alone. Therefore, it has been difficult to obtain approval for the use of the individual biomarkers for screening.

The main limitation of blood-based biomarker assays such as the GALAD model is that if the biomarkers are positive, indicating that a patient has cancer, the doctor still has to find the cancer using additional imaging tests such as ultrasound. Alternatively, if the doctor is

In data from cohorts in Europe, Asia, and North America, the BALAD and BALAD-2 models both appear to be useful for predicting the outcomes of patients with HCC.

convinced that something is abnormal in the liver despite a normal biomarker test, a computed tomography scan or magnetic resonance imaging is needed, both of which are expensive. Thus, it can be difficult to determine the optimal utility of the models in screening and diagnosis algorithms and whether these tools are cost-effective.

G&H Could you discuss more specifically the research that has compared the performance of the GALAD model to that of AFP by itself?

LR Much of the work that Dr Johnson has published on the GALAD model has looked at its performance compared to AFP. He has shown very clearly in studies conducted in Hong Kong, Japan, Germany, and the United Kingdom that the GALAD model performs better than AFP alone. My colleagues and I conducted a study published earlier this year where we examined the performance of the GALAD model in the US population and confirmed that it is better than AFP alone.

Furthermore, we proposed a score called GALADUS, which integrates the GALAD model with ultrasound results. We found that the GALADUS score performs a little better than the GALAD score alone. Thus, if a patient has undergone an ultrasound, the result can be incorporated into the scoring system, regardless of whether the result was positive or negative.

G&H How sensitive and specific are these models for different stages of HCC?

LR Sensitivity and specificity for biomarkers are often determined by the area under the receiver operating

characteristic (ROC) curve. In the aforementioned study that compared the GALAD and GALADUS models, the area under the ROC curve for the GALAD model was 0.95 for the detection of HCC overall and 0.92 for the detection of early-stage HCC. In addition, we found that the GALAD score had a sensitivity of 89% and a specificity of 86% for the detection of HCC overall. For the detection of early-stage disease, these numbers were 82% and 86%, respectively. As for the GALADUS model, the area under the ROC curve was 0.98 for HCC overall and 0.97 specifically for early-stage disease. In contrast, the best sensitivities for AFP are typically approximately 60%.

In general, the GALAD and GALADUS models perform acceptably well for patients with early-stage HCC, which is important because the major goal of surveillance is to detect disease before it becomes advanced. In addition, adding ultrasound to the GALAD model produced excellent performance in our study.

G&H Has there been any research on the use of the GALAD model in different underlying liver diseases?

LR Yes. For example, my colleagues and I conducted a cohort study with the National Cancer Institute–funded Early Detection Research Network to compare use of the GALAD model in patients with different etiologies of HCC. The GALAD model had a sensitivity of 71% and specificity of 92% among patients with alcoholic liver disease, a sensitivity of 83% and specificity of 100% among patients with hepatitis B, and a sensitivity of 79% and specificity of 85% among patients with hepatitis C. For patients who did not have a viral or alcohol etiology, which includes patients with fatty liver disease, the sensitivity was 78% and specificity was 93%.

G&H What recent research has been conducted on the BALAD model?

LR In data from cohorts in Europe, Asia, and North America, the BALAD and BALAD-2 models both appear to be useful for predicting the outcomes of patients with HCC. My colleagues and I found that a modified score incorporating the size of the largest tumor also performed slightly better than the BALAD or BALAD-2 scores in predicting recurrence or death of patients after liver transplantation for HCC.

G&H Are the GALAD and BALAD models being used in routine clinical practice yet?

LR There has been slow uptake by doctors, and the models have not been formally approved by US or European

liver societies. However, the scores can be calculated easily online, and more and more doctors are at least trying the models to see whether they are useful in their own practices.

G&H Do you think that eventually the use of these models will become more widespread?

LR That is my suspicion, although there are also other biomarkers in development that perform quite well. Many of the new biomarker assays are based on genomic markers such as differentially methylated regions. It will be interesting to see how some of the new biomarkers that are being developed compare with the GALAD and GALADUS models and whether the addition of more biomarkers would substantially improve the performance of the models. I foresee quite a bit of activity in this field over the next several years.

G&H What are the next steps in research regarding the GALAD and BALAD biomarker models?

LR It is important to continue to build the evidence base and to have more studies on the use of these models in comparison to the use of AFP by itself. In addition, when

only AFP is used, experienced clinicians follow the trends in biomarker levels, rather than just isolated individual measurements of the biomarker. The same may need to be done with the GALAD and BALAD models, and although we know that trends are important, doctors need to gain experience with the biomarkers to understand what the trends mean. Some investigators, including Dr Johnson, are beginning to develop models that include multiple measurements of the biomarkers.

Dr Roberts has received grant funding from Wako Diagnostics, Exact Sciences, GRAIL, and Glycotest, Inc.

Suggested Reading

Ahmed Mohammed HF, Roberts LR. Should AFP (or any biomarkers) be used for HCC surveillance? *Curr Hepatol Rep.* 2017;16(2):137-145.

Marrero JA, Su GL, Wei W, et al. Des-gamma carboxyprothrombin can differentiate hepatocellular carcinoma from nonmalignant chronic liver disease in American patients. *Hepatology.* 2003;37(5):1114-1121.

Wongjarupong N, Negron-Ocasio GM, Chaiteerakij R, et al. Model combining pre-transplant tumor biomarkers and tumor size shows more utility in predicting hepatocellular carcinoma recurrence and survival than the BALAD models. *World J Gastroenterol.* 2018;24(12):1321-1331.

Yang JD, Addissie BD, Mara KC, et al. GALAD score for hepatocellular carcinoma detection in comparison with liver ultrasound and proposal of GALADUS score. *Cancer Epidemiol Biomarkers Prev.* 2019;28(3):531-538.