

Keywords: hepatocellular carcinoma; surveillance; α -fetoprotein; cirrhosis

A new approach to the use of α -fetoprotein as surveillance test for hepatocellular carcinoma in patients with cirrhosis

M Biselli¹, F Conti^{*1}, A Gramenzi¹, M Frigerio¹, A Cucchetti¹, G Fatti¹, M D'Angelo¹, M Dall'Agata¹, E G Giannini², F Farinati³, F Ciccarese⁴, P Andreone¹, M Bernardi¹ and F Trevisani¹

¹Dipartimento di Scienze Mediche e Chirurgiche, Università di Bologna, Bologna, Italy; ²Dipartimento di Medicina Interna, Unità di Gastroenterologia, Università di Genova, Genova, Italy; ³Dipartimento di Scienze Chirurgiche e Gastroenterologiche, Unità di Gastroenterologia, Università di Padova, Padova, Italy and ⁴Divisione di Chirurgia, Policlinico San Marco, Zingonia, Italy

Background: Surveillance for hepatocellular carcinoma (HCC) is recommended in patients with cirrhosis. As α -fetoprotein (AFP) is considered a poor surveillance test, we tested the performance of its changes over time.

Methods: Eighty patients were diagnosed with HCC (cases) during semiannual surveillance with ultrasonography and AFP measurement were recruited and matched for age, gender, etiology and Child-Pugh class with 160 contemporary cancer-free controls undergoing the same surveillance training group (TG). As a validation group (VG) we considered 36 subsequent patients diagnosed with HCC, matched 1:3 with contemporary cancer-free controls. α -Fetoprotein values at the time of HCC diagnosis (T0) and its changes over the 12 ($\Delta 12$) and 6 months ($\Delta 6$) before cancer detection were considered.

Results: In both TG and VG, >80% of HCCs were found at an early stage. In TG, AFP significantly increased over time only in cases. T0 AFP and a positive $\Delta 6$ were independently associated with HCC diagnosis (odds ratio: 1.031 and 2.402, respectively). The area under the curve of T0 AFP was 0.76 and its best cutoff (BC) was 10 ng ml⁻¹ (sensitivity 66.3%, specificity 80.6%). The combination of AFP >10 ng ml⁻¹ or a positive $\Delta 6$ composite α -fetoprotein index (CAI) increased the sensitivity to 80% with a negative predictive value (NPV) of 86.2%. Negative predictive value rose to 99%, considering a cancer prevalence of 3%. In the VG, the AFP-BC was again 10 ng ml⁻¹ (sensitivity 66.7%, specificity 88.9%), and CAI sensitivity was 80.6% with a NPV value of 90.5%.

Conclusions: CAI achieves adequate sensitivity and NPV as a surveillance test for the early detection of HCC in cirrhosis.

Hepatocellular carcinoma (HCC) is the third most common cause of cancer-related deaths worldwide, and in most cases it develops in a cirrhotic liver (Kamangar *et al*, 2006). Surveillance of patients at risk for developing HCC can detect tumours amenable to curative therapies, with positive impact on survival (Trevisani *et al*, 2002; Zhang *et al*, 2004). The α -fetoprotein (AFP) is a serum marker still used for the diagnosis and surveillance of HCC. A value of 20 ng ml⁻¹ is considered as best cutoff (BC) to suspect the development of HCC in the setting of chronic liver disease (Trevisani *et al*, 2001). However, about 32–59% of patients with

HCC have normal AFP levels and, conversely, non-tumour-related AFP elevations may occur in patients with cirrhosis or chronic hepatitis, making AFP inadequate as a surveillance test (Gupta *et al*, 2003; Colli *et al*, 2006; Lok *et al*, 2010). To detect HCC at an early stage, it has been proposed to lower the AFP cutoff to 10.9 ng ml⁻¹, obtaining a sensitivity of 66% and a specificity of 82% (Marrero *et al*, 2009). However, even with this cutoff, about 30% of HCC escape from an early diagnosis. Therefore, Western guidelines consider AFP as inadequate to survey patients at risk of HCC and recommend the use of ultrasound (US) alone

*Correspondence: Dr F Conti; E-mail: fabio.conti2@studio.unibo.it

Received 8 May 2014; revised 25 July 2014; accepted 15 September 2014; published online 14 October 2014

© 2015 Cancer Research UK. All rights reserved 0007–0920/15

(Bruix and Sherman, 2011; European Association for the Study of the Liver; European Organization for Research and Treatment of Cancer, 2012). Nevertheless, being an operator-dependent imaging technique, US has been shown to suffer from a limited sensitivity (63%) for an early detection of HCC (Singal *et al*, 2009). Thus, some authors as well as Eastern guidelines suggest maintaining the use of AFP in HCC surveillance (Poon *et al*, 2009; Lee *et al*, 2013). Waiting for the new reliable biomarkers to complement US in detecting early HCC, our retrospective case-control study aimed at optimising the efficiency of AFP as a surveillance test in cirrhotic patients, considering both its value at the time of HCC diagnosis and its changes overtime before the diagnosis.

PATIENTS AND METHODS

Between January 2000 and February 2009, we recruited 80 patients newly diagnosed with HCC in the outpatients' clinic of our centers during a regular semiannual surveillance program based on US and AFP measurement. These patients served as a training group (TG). As a validation group (VG), we enrolled 36 patients newly diagnosed with HCC between March 2009 and May 2013 in the cohort of cirrhotic patients on semiannual surveillance at the Bologna center.

HCC patients (HCC cases) were matched at a 1:2 ratio for TG and 1:3 for VG to simultaneously surveyed patients who remained cancer-free for at least 18 months after enrollment. Matching variables were gender, age (within a 5-year interval), etiology of cirrhosis and Child-Pugh class collected at the time of HCC diagnosis (time 0; T0); in controls, T0 was considered the time of the surveillance visit closest to the HCC diagnosis of the corresponding case. At T0, the following data were collected: serum AFP, bilirubin, alanine aminotransferase, aspartate aminotransferase, albumin, creatinine, international normalised ratio, glucose and model for end-stage liver disease (MELD) score. α -Fetoprotein values at 12 and 6 months before T0 (T-12 and T-6, respectively) were also recorded. To avoid interference with AFP levels, patients who began or stopped antiviral therapy during the 18 months preceding the HCC occurrence or the enrollment (controls) were excluded. α -Fetoprotein serum levels were measured using a commercially available Immunoassay (COBAS ROCHE Diagnostics GmbH, Milan, Italy).

For patients in whom US detected a new nodule, we adopted the recall policy proposed by the European Bruix *et al* (2001) and, after April 2011, the American and European guidelines (Bruix and Sherman, 2011; European Association for the Study of the Liver; European Organization for Research and Treatment of Cancer, 2012). Patients with a negative US, but an AFP with a value $>10\text{ ng ml}^{-1}$ and doubled compared with the previous one, underwent computed tomography (CT). Computed tomography or magnetic resonance imaging (MRI) were also performed when the quality of US was deemed to be poor by the operator. The HCC diagnosis was based on histology in 12 out of 80 (15%) TG patients and in 7 out of 36 (19.4%) VG patients. In the others, it was based on recommended non-invasive criteria (Bruix *et al*, 2001; Bruix and Sherman, 2011; European Association for the Study of the Liver; European Organization for Research and Treatment of Cancer, 2012). HCC was staged by CT or MRI. All patients underwent chest X-ray, whereas additional investigations were performed when extra-hepatic involvement was suspected. HCC was classified as: single nodule, paucifocal (≤ 3 nodules), multifocal (> 3 nodules), diffuse and massive type (Stefanini *et al*, 1995). Staging was determined according to the Barcelona Clinic Liver of Cancer (BCLC) system (Bruix and Sherman, 2011; European Association for the Study of the Liver; European Organization for Research and Treatment of Cancer, 2012). Diagnosis of cirrhosis

was supported by biopsy or by clinical and laboratory features including those of portal hypertension at endoscopy and/or US. Child-Pugh classification assessed the severity of cirrhosis (Pugh *et al*, 1973).

Statistical methods. Continuous variables were expressed as mean \pm s.d. or median (range), as appropriate. Comparisons between groups were made by χ^2 -test and Fisher's exact test for qualitative variables, and Mann-Whitney *U*-test for continuous ordinal data. The Friedman test was used to compare AFP values at T-12, T-6 and T0 within each group. The correlations were calculated by linear regression analysis and Spearman's rho test. The AFP change between T-12 and T0 was indicated as $\Delta 12$ -AFP, and between T-6 and T0 as $\Delta 6$ -AFP. The AFP changes were considered 'positive' (and indicated as $\Delta 6+$ or $\Delta 12+$, respectively) when an increase of at least 1 ng ml^{-1} occurred during the monitored period. In the TG, the association between HCC diagnosis and categorical (gender, etiology, ascites, encephalopathy, Child-Pugh class, $\Delta 6+$ and $\Delta 12+$) and continuous variables (age, AFP at different time points, $\Delta 12$ -AFP, $\Delta 6$ -AFP, creatinine, albumin, bilirubin, international normalised ratio, alanine aminotransferase, aspartate aminotransferase, glucose, Child-Pugh and MELD scores) was tested by logistic regression. Variables associated with HCC ($P < 0.10$) were included in multivariate binary logistic regression analysis. The odds ratios with 95% confidence interval (95% CI) were calculated. In both the TG and VG, the discrimination of each variable in predicting the risk of HCC at T0 was assessed by the area under the receiver operating characteristic curve (AUROC). AUROCs were compared using the algorithm described by Hanley and McNeil (1983). Finally, the cutoff values ensuring the lowest false negative and false positive results (best cutoff) were utilised to calculate sensitivity, specificity, positive (PPV) and negative predictive values (NPV). The cost of different surveillance strategies was calculated using a decision algorithm including two models: (1) semiannual US (without AFP) performed at primary health-care institutions; (2) semiannual AFP determination followed by US, performed in a tertiary referral center, if the AFP crossed the thresholds selected to suspect HCC development. In the last model, patients continued to be surveyed with US without other radiological techniques. Sensitivity and specificity of AFP were those obtained in the VG, whereas sensitivity and specificity of US performed in general and expert centers derived from the literature (Teefey *et al*, 2003; Singal *et al*, 2009). Costs were assumed from a health-care system perspective, only including *direct* costs of surveillance. They derived from Italian National Healthcare System reimbursement schedules, as follows: US = €44; AFP = €11. A probabilistic sensitivity analysis was then performed assuming a hypothetical scenario of 1000 patients at risk, and sensitivity and specificity were varied within their 95% CI, whereas costs were varied within 20% of base-case values. All analyses were performed using SPSS 13.0 (SPSS Inc., Chicago, IL, USA) and MedCalc 9.2.1.0 (MedCalc Software, Mariakerke, Belgium). $P < 0.05$ was considered statistical significant.

RESULTS

Clinical and laboratory features of HCC cases and controls in the TG and VG are reported in Table 1. HCC cases and controls were well comparable, except for modest, but statistically significant differences of MELD score in TG, and albumin and international normalised ratio in both the TG and VG. A significant difference between HCC cases and controls was found for AFP at T-12, T-6 and, more evidently, at T0 in both the TG and VG. Most HCC were diagnosed at a very early or early stage (TG: 83.7%; VG: 91.6%). Considering all 116 HCC cases (TG + VG) detected by the

Table 1. Baseline characteristics of HCC cases and matched controls

| Variable | Training group | | | Validation group | | |
|---------------------------------------|--------------------|----------------------------|--------|--------------------|----------------------------|--------|
| | HCC cases (n = 80) | Matched controls (n = 160) | P | HCC cases (n = 36) | Matched controls (n = 108) | P |
| Males | 56 (70) | 112 (70) | 1 | 22 (61) | 66 (61) | 1 |
| Age, years | 66 (51–81) | 65 (48–81) | 0.728 | 65.5 (53–77) | 62 (33–90) | 0.89 |
| Etiology (%) | | | 1 | | | 0.231 |
| Hepatitis B | 14 (17.5) | 28 (17.5) | | 3 (8.3) | 11 (10.2) | |
| Hepatitis C | 55 (68.8) | 110 (68.8) | | 26 (72.2) | 58 (53.7) | |
| Alcohol | 9 (11.2) | 18 (11.2) | | 5 (13.9) | 23 (21.3) | |
| Others | 2 (2.5) | 4 (2.5) | | 2 (5.6) | 16 (14.8) | |
| Child-Pugh score | 6 (5–10) | 5 (5–11) | 0.156 | 6 (5–10) | 5 (5–12) | 0.405 |
| MELD score | 10 (6–18) | 9 (6–20) | 0.034 | 9 (6–23) | 9 (6–23) | 0.299 |
| T0 AFP, ng ml ⁻¹ | 17.5 (0.6–1238) | 5 (1–75) | <0.001 | 15.5 (1–267) | 4.5 (1–52) | <0.001 |
| T-6 AFP, ng ml ⁻¹ | 9.75 (1–129.2) | 5 (1–145) | <0.001 | 11 (1–114) | 4 (1–70) | <0.001 |
| T-12 AFP, ng ml ⁻¹ | 9.5 (0.4–69) | 5 (1–359) | <0.001 | 9 (1–354) | 4 (1–83) | <0.001 |
| ALT, IU l ⁻¹ | 42 (10–626) | 38 (8–235) | 0.506 | 53.5 (15–301) | 32.5 (10–322) | 0.067 |
| Glucose, mg ml ⁻¹ | 95 (60–342) | 99 (63–296) | 0.612 | 105 (77–261) | 100 (58–269) | 0.143 |
| Serum creatinine, mg ml ⁻¹ | 0.9 (0.4–1.7) | 0.9 (0.5–1.9) | 0.784 | 0.8 (0.5–2.7) | 0.8 (0.5–2.3) | 0.427 |
| Albumin, g dl ⁻¹ | 3.6 (2.1–4.6) | 3.8 (2.0–6.1) | 0.031 | 3.8 (2.2–5.1) | 3.8 (2.6–5.1) | 0.026 |
| Total bilirubin, mg dl ⁻¹ | 1.1 (0.4–5.9) | 1.1 (0.3–4.4) | 0.543 | 1 (0.4–3.9) | 1.1 (0.2–7.2) | 0.546 |
| INR | 1.2 (0.9–1.6) | 1.1 (0.9–2.4) | 0.048 | 1.2 (1–1.6) | 1.2 (0.9–1.7) | 0.039 |
| Ascites | 26 (32.6) | 35 (21.9) | 0.053 | 8 (22.2) | 20 (18.5) | 0.632 |
| Porto-systemic encephalopathy | 1 (1.3) | 6 (3.8) | 0.43 | 3 (8.3) | 2 (1.9) | 0.1 |

Abbreviations: AFP = α-fetoprotein; ALT = alanine aminotransferase; HC = hepatocellular carcinoma; INR = international normalised ratio; MELD = model for end-stage liver disease. Values are expressed as number (%) or median with range.

surveillance, 95 (81.9%) were suspected by US, 105 (90.5%) by US combined with AFP as described in the Patients and Methods section of the paper. Eleven (9.5%) HCC cases were detected by CT or MRI following poor US quality. Interestingly, adding AFP > 10 ng ml⁻¹ to US allow to suspect up to 111 (96.5%) HCC cases. Therefore, the added value of AFP to US was 8.6% using AFP as described in the Patients and Methods and 14.6% with AFP > 10 ng ml⁻¹.

Training group. Median AFP increased from 9.5 ng ml⁻¹ (0.4–69) at T-12 to 17.5 ng ml⁻¹ (0.6–1238) at T0 (*P* < 0.001) in HCC cases, whereas it did not change in controls (5 ng ml⁻¹ (1–359) and 5 ng ml⁻¹ (1.0–75), *P* = 0.126) (Figure 1A). In HCC cases, median AFP also increased between T-6 and T0 (from 9.75 ng ml⁻¹ (1–129.2) to 17.5 ng ml⁻¹ (0.6–1238), *P* < 0.001). Consequently, the median Δ6 AFP (2 vs 0 ng ml⁻¹) and Δ12 AFP (4.6 vs 0 ng ml⁻¹) were greater in HCC cases than in controls (*P* < 0.001). Finally, at each time point, AFP was higher in HCC cases than in controls (*P* < 0.001).

At univariate analysis the following variables were associated with HCC: MELD, albumin, T0 AFP, T-6 AFP, Δ6 AFP, Δ12 AFP, Δ6 + AFP and Δ12 + AFP. At multivariate analysis, only T0 AFP (odds ratios: 1.031, 95% CI: 1.008–1.055, *P* = 0.008) and Δ6 + AFP (odds ratios: 2.402, 95% CI: 1.246–4.631, *P* = 0.009) were associated with HCC. No correlations were found between alanine aminotransferase and T0 AFP in either HCC cases or controls. The AUROC of T0 AFP showed an useful diagnostic accuracy (0.76, 95% CI: 0.701–0.813) (Figure 2A). Its BC was 10 ng ml⁻¹ (AFP-BC), with a sensitivity of 66.3% and a specificity of 80.6%. A similar sensitivity (67.5%) but a lower specificity (70.6%) was observed using the Δ6 + AFP. Assuming HCC prevalences expected in clinical setting (3 and 5%), PPV and NPV of Δ6 + AFP and AFP-BC were similar (Table 2). The use of T0 AFP and

Δ6 + AFP in a *combined-sequential way*, first a T0 AFP > 10 ng ml⁻¹ and, in patients with a value below this threshold, the Δ6 + AFP (Combined α-fetoprotein Index: CAI), improved the sensitivity up to 80% (95% CI: 74.3–84.8%). Notably, the NPV of CAI was extremely high at HCC prevalence of both 3% (99%, 95% CI: 96.5–99.8) and 5% (98.3%, 95% CI: 95.5–99.5%). As expected, the CAI specificity dropped to 62.5 (95% CI: 56–68.6). Finally, the comparison between patients with a positive or negative CAI did not show any difference in number and size of nodules, and BCLC stage (Table 3). Importantly, about 80% of HCC were identified by CAI in an early stage.

Validation group. Median AFP increased from 9 ng ml⁻¹ (1–354) at T-12 to 15.5 ng ml⁻¹ (1–267) at T0 in HCC cases (*P* < 0.001), whereas it remained stable in controls (4 ng ml⁻¹ (1–83) and 4.5 ng ml⁻¹ (1–52) (*P* = 0.884) (Figure 1B). At each time point, AFP was significantly higher in HCC cases than in controls (*P* < 0.001). The AUROC of T0 AFP was significant (0.783, 95% CI: 0.706–0.847) (Figure 2B) and, as in the TG, the AFP-BC value was 10 ng ml⁻¹, with a comparable sensitivity (66.7%), but a higher specificity (88.9%). The PPV and NPV of AFP-BC and Δ6 + AFP at 3% and 5% of HCC prevalence were similar (Table 2). The VG confirmed that CAI improves the sensitivity up to 80.6% (95% CI: 73%–86.5%), maintaining a high NPV at the HCC prevalences expected in clinical practice. As in the TG, the results of CAI did not segregate patients for different number or size of nodules and BCLC stage (Table 3). Lastly, about 90% of HCC were identified at an early stage.

Cost analysis of the potential use of CAI compared with US as surveillance test. (Figure 3) We assumed a hypothetical scenario of 1000 patients with an annual HCC risk of 3% to be surveyed for 1 year. Accordingly, standard US surveillance performed outside

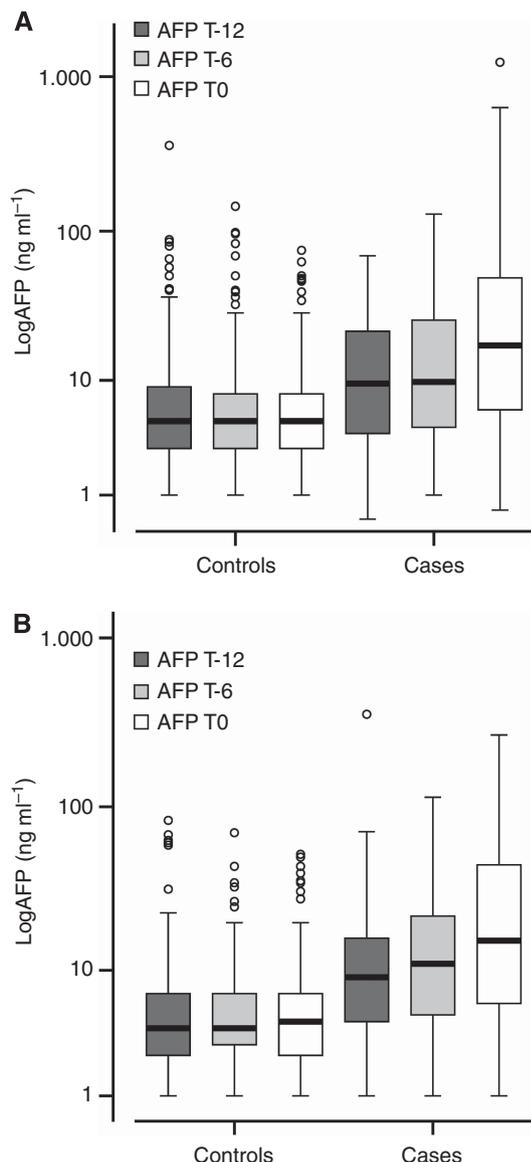


Figure 1. Box plot of AFP shown as log10 in HCC cases and controls in training (A) and validation (B) group. The box shows the 25th and 75th percentile with a line indicating the median. The interquartile range spreads outside the box. Points outside the interquartile range indicate outliers.

referral centers resulted in the detection of 21 tumours, with a total cost of € 88 270 ± 7380. Instead, the use of CAI as a first line surveillance tool, followed by US performed in expert centers in the presence of a positive CAI (Figure 4), led to the detection of 20 tumours, with a total cost of €50 030 ± 4300. Consequently, the average cost per each HCC diagnosed was € 4203 for standard US surveillance and €2501 for CAI→US strategy.

DISCUSSION

Ultrasound is considered the basic tool for HCC surveillance, whereas the use of AFP in this context is controversial (Bruix *et al*, 2001; Poon *et al*, 2009; Bruix and Sherman, 2011; European Association for the Study of the Liver; European Organization for Research and Treatment of Cancer, 2012).

This is the first case-control study evaluating the performance of AFP in a setting of semiannual surveillance of patients

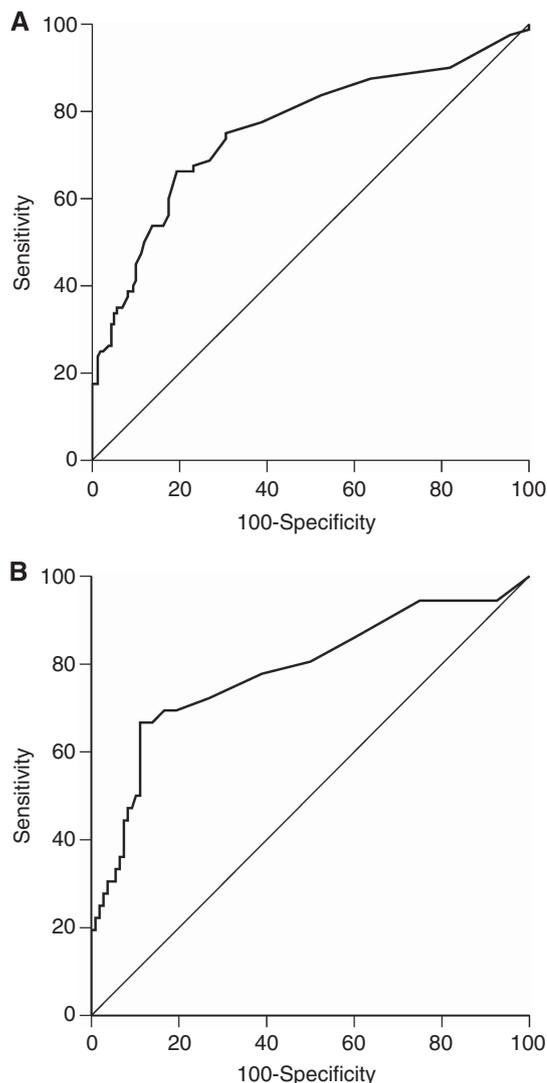


Figure 2. AUROC of T0 AFP evaluating the discrimination accuracy between HCC cases and controls in training (A) and validation group (B).

with cirrhosis due to different etiologies, as occurs in Southern Europe. AUROC analysis of T0 AFP indicated 10 ng ml⁻¹ as the best cutoff. This threshold perfectly matches with that (10.9 ng ml⁻¹) proposed by Marrero *et al* (2009), for the diagnosis of early stage HCC in U.S. patients with established HCC and not in surveillance. In addition, the sensitivity (66% vs 66%) and specificity (81% vs 82%) of these cutoffs were equivalent in our and in the U.S. series. Importantly, we paid great attention to matching HCC cases and controls. Therefore, our results support the use of an AFP cutoff of 10 ng ml⁻¹ (and not 20 ng ml⁻¹) to suspect HCC development in cirrhotic patients undergoing semiannual surveillance. However, the sensitivity of AFP remained poor, allowing one third of HCCs to escape a subclinical diagnosis. Indeed, in considering the results of a prospective investigation on HCV patients, Lok *et al* (2010) concluded that biomarkers such as AFP or des- γ -carboxy prothrombin are needed to complement US in the detection of early HCC, but neither des- γ -carboxy prothrombin nor AFP were optimal. Being convinced of the inadequacy of the conventional use of available oncomarkers in HCC surveillance, we attempted to improve the performance of AFP by testing the ability of AFP changes over time in differentiating HCC cases from those who remained cancer-free.

Table 2. PPV and NPV for the diagnosis of HCC of the AFP-BC (10 ng ml⁻¹) and Δ6 + AFP calculated for the training group (HCC prevalence 33%) and the validation group (HCC prevalence 25%), and for two tumour prevalences encountered in clinical practice

| HCC prevalence (%) | Training group | | | | Validation group | | | |
|--------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | PPV (95% CI) | NPV (95% CI) | Sens (95% CI) | Spec (95% CI) | PPV (95% CI) | NPV (95% CI) | Sens (95% CI) | Spec (95% CI) |
| AFP-BC | | | | | | | | |
| Study | 63.1 (56.6–69.1) | 82.7 (77.2–87.1) | 66.3 (59.8–72.1) | 80.6 (74.9–85.3) | 66.7 (58.3–74.2) | 88.9 (82.3–93.3) | 66.7 (58.3–74.2) | 88.9 (82.3–93.3) |
| 5 | 15.3 (11.1–20.6) | 97.8 (94.8–99.2) | | | 24 (16.3–33.8) | 98.1 (92.4–99.7) | | |
| 3 | 9.6 (6.3–14.2) | 98.7 (96–99.7) | | | 15.7 (9.4–24.6) | 98.9 (93.5–99.9) | | |
| Δ6 + AFP | | | | | | | | |
| Study | 53.5 (46.9–59.9) | 81.3 (75.4–85.9) | 67.5 (61.1–73.3) | 70.6 (64.4–76.2) | 43.5 (35.4–52.1) | 89 (82.5–93.4) | 75 (67–81.7) | 67.6 (59.2–75) |
| 5 | 10.8 (7.3–15.6) | 97.6 (94.6–99.1) | | | 10.9 (5.8–19.1) | 98.1 (92.4–99.7) | | |
| 3 | 6.6 (4–10.8) | 98.6 (95.9–99.6) | | | 6.5 (2.8–13.8) | 98.9 (93.5–99.9) | | |
| CAI | | | | | | | | |
| Study | 51.6 (45.1–58.1) | 86.2 (81–90.2) | 80 (74.3–84.8) | 62.5 (56–68.6) | 41.4 (33.4–49.9) | 90.5 (84.2–94.6) | 80.6 (73–86.5) | 62 (53.5–69.9) |
| 5 | 10.1 (6.7–14.8) | 98.3 (95.5–99.5) | | | 10 (5.2–18.1) | 98.4 (92.8–99.8) | | |
| 3 | 6.2 (3.6–10.2) | 99 (96.5–99.8) | | | 6.2 (2.6–13.3) | 99 (93.8–100) | | |

AFP-BC = alpha-fetoprotein best-cut-off at the time of HCC detection; Δ6 + AFP = increasing alpha-fetoprotein in the semester prior to HCC detection; CAI = combined Alpha-fetoprotein Index; HCC = hepatocellular carcinoma; PPV = positive predictive value; NPV = negative predictive value; Sens = sensitivity; Spec = specificity; CI = confidence interval.

Table 3. BCLC stage and cancer burden in training and validation groups subdivided for the result of the combined-sequential AFP test (AFP-BC/Δ6 + AFP)

| Variable | Training group | | | Validation group | | |
|---------------------------------|-----------------------------------|-----------------------------------|-------|----------------------------------|-----------------------------------|-------|
| | AFP-BC/Δ6 + AFP negative (n = 16) | AFP-BC/Δ6 + AFP positive (n = 64) | P | AFP-BC/Δ6 + AFP negative (n = 7) | AFP-BC/Δ6 + AFP positive (n = 29) | P |
| Number of lesions | | | 0.772 | | | 0.813 |
| Monofocal | 12 (75) | 47 (73.5) | | 5 (71.4) | 22 (75.9) | |
| Paucifocal | 4 (25) | 15 (23.4) | | 2 (28.6) | 6 (20.7) | |
| Multifocal | 0 (0) | 2 (3.1) | | 0 (0) | 1 (3.4) | |
| Size of the largest lesion (mm) | 2.1 (1–3.3) | 2.3 (0.8–4.8) | 0.201 | 2.7 (1.2–4.5) | 2.1 (1–5.3) | 0.841 |
| BCLC | | | 0.391 | | | 0.618 |
| Very-early | 5 (31.2) | 13 (20.3) | | 3 (42.9) | 9 (31.1) | |
| Early | 11 (68.8) | 38 (59.4) | | 4 (57.1) | 17 (58.6) | |
| Intermediate | 0 (0) | 9 (14.1) | | 0 (0) | 3 (10.3) | |
| Advanced | 0 (0) | 2 (3.1) | | 0 (0) | 0 (0) | |
| End-stage | 0 (0) | 2 (3.1) | | 0 (0) | 0 (0) | |

Values are expressed as number (%) or median with range. AFP-BC = alpha-fetoprotein best-cut-off at the time of HCC detection; Δ6 + AFP = increasing alpha-fetoprotein in the semester prior to HCC detection; BCLC = Barcelona Clinic Liver Cancer.

In addition to T0 AFP, an increasing AFP level over the 6 months before HCC detection (Δ6+) was the sole variable independently associated with tumour. Unfortunately, even the sensitivity of Δ6+ was not superior to that of T0 AFP, suggesting that both tests, if used alone, are inefficient. These results could be considered as another bullet for killing the biomarker surveillance for HCC, but this would conflict with a recent study that proposes the AFP fluctuation over time as a tool to improve the accuracy of surveillance (Lee *et al*, 2013). So, we ideated a combined-sequential use of static (T0) and dynamic (Δ6+) AFP values, named CAI. This index achieved a sensitivity of 80% (confirmed in the VG), with a NPV of about 99% at the cancer prevalences (3% or 5%) observed in clinical practice. These figures enable CAI to be considered as adequate to survey patients at risk of HCC and are competitive with US, the sensitivity of which has been estimated to be 60–70% for the detection of early stage HCC (Colli *et al*, 2006; Singal *et al*, 2009). Indeed, US detection of early HCC in cirrhotic patients may be challenged by a coarse liver echo-texture. Moreover, a complete evaluation of liver parenchyma may be difficult or impossible because of the body habitus (for example

obesity, intestinal gas, colonic interposition, ascites) or poor compliance to the breath hold command. Finally, the effectiveness of US is highly dependent on operator expertise, as testified by important differences in the results of US surveillance performed at community hospitals or referral centers (Lee *et al*, 2011; Giannini *et al*, 2013). On the other hand, it is unrealistic to expect that expert centers can fulfill the number of US scans required by a systematic application of semiannual surveillance in all patients at risk of HCC, and that all these patients agree to be semiannually checked outside local primary care centers.

It should be pointed out that in our study about three-quarters of the HCC cases in which US was not diagnostic showed a positive CAI. Monitoring of serum AFP has several advantages, being an easy and cheap test for which well-standardised methods with an intra-assay variability <4% are available, as that utilised in our study. Therefore, in patients with ‘normal’ AFP levels (≤10 ng ml⁻¹), even small increases during monitoring deserve attention, because they may reflect a biological rather than a methodological variation and this new concept could change the general opinion that many HCCs ‘do not produce’ AFP. In fact,

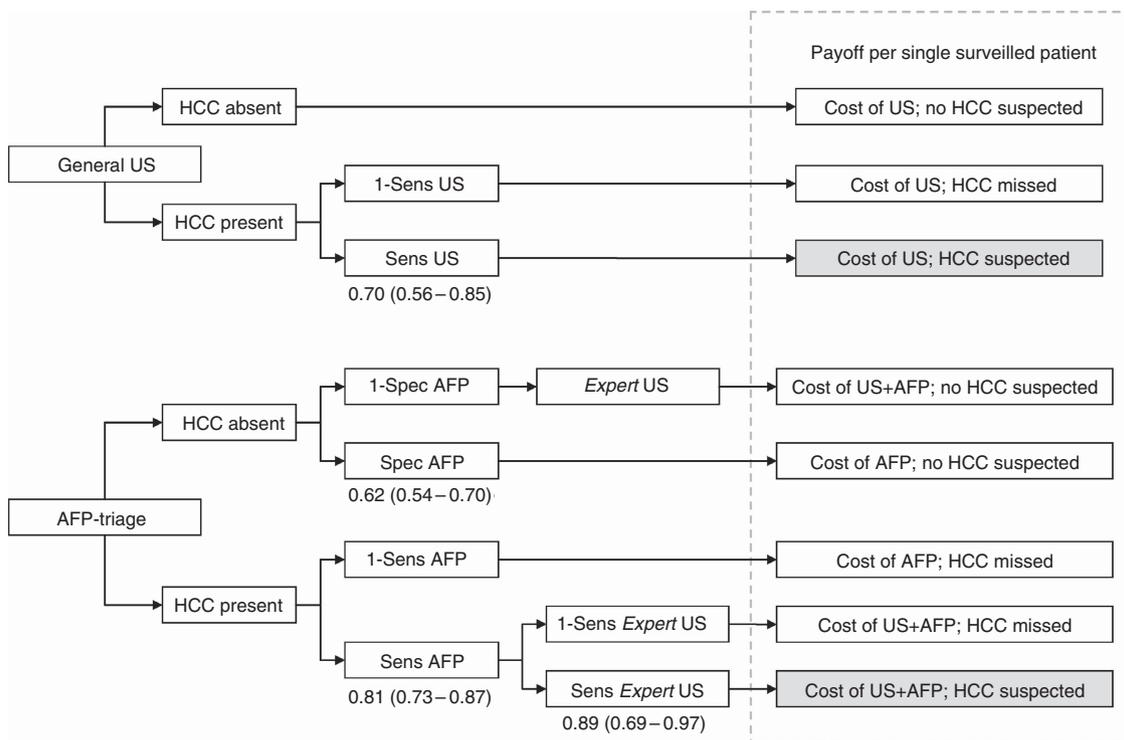


Figure 3. Decision algorithm considering two different surveillance strategies: conventional US and CAI → US strategy. Sensitivity and specificity derived from the validation group and from literature. Costs derived from the NHS reimbursement as follows: US = €44 (35–53); AFP = €11 (9–13). Costs of diagnosis confirmation were not included.

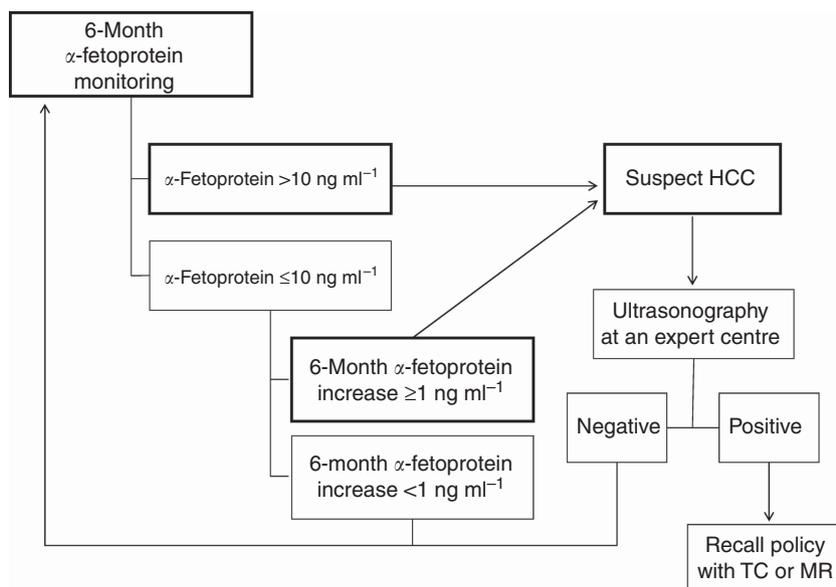


Figure 4. Surveillance algorithm of patients at risk of hepatocellular carcinoma based on the Combined α -fetoprotein index (CAI). The recall policy is entrained by the detection of a new nodule, according to the recommendations of practice guidelines for HCC management (Bruix and Sherman, 2011). CT = computed tomography; MRI = magnetic resonance imaging.

despite a ‘normal’ AFP value, increasing levels over the previous 6 months should raise the suspicion of HCC occurrence. Thus, CAI overcame the drawback of a poor sensitivity of AFP measurement alone, reaching a figure of 80%. Moreover, its NPV approached 100% at the 3–5% cancer prevalence observable in clinical practice.

Another disadvantage of AFP monitoring relies on the fact that several extra-tumoral factors can affect its levels (Trevisani *et al*, 2001; Nguyen *et al*, 2002; Di Bisceglie *et al*, 2005; Chen *et al*, 2007), reducing the specificity and remarkably increasing the cost of surveillance (European Association for the Study of the Liver;

European Organization for Research and Treatment of Cancer, 2012; Giannini *et al*, 2013). In this respect, we did not observe any significant correlation between alanine aminotransferase and AFP values either in HCC cases or controls. This finding is not surprising in HCC patients in whom the AFP levels primarily depend on the tumour, but conflicts with what was reported by a case registry analysis regarding HCV patients (Richardson *et al*, 2012). A possible explanation lies in the differences in stage of liver disease, degree of necroinflammation and etiology of the two cohorts. Nonetheless, in our study, the specificity of AFP was only

62%, and its use as an initial test could increase the cost of surveillance due to false positive results. Having this in mind, using a simulation model we compared the cost of standard US surveillance (with US performed at primary health-care institutions) vs CAI followed by US performed in specialised liver centers. The comparison indicated that the second strategy led to a 43% reduction of total direct costs for each HCC detected, paying the clinical price of one HCC per year lost for every 1000 surveyed patients. The sparing effect of the CAI strategy suggests that the over-cost produced by false positive results of AFP was over-compensated by the reduced use of US.

Our study has several limitations, the first because it was a retrospective case-control study. However, it was nested in a prospective cohort, and its results were validated in an independent cohort. Second, our findings were obtained in stable cirrhotic patients (most of them with HCV infection) and avoiding the confounding effect of starting or ending an antiviral treatment (Chen *et al*, 2007). Thus, they cannot be extrapolated to different categories of patients. A third limitation relies on the applicability of CAI only to patients with an AFP baseline $\leq 10 \text{ ng ml}^{-1}$. However, based on the prevalence of patients with an AFP $> 10 \text{ ng ml}^{-1}$ at T-12 observed both in controls and HCC cases, and considering HCC prevalences up to 5%, more than 80% of cirrhotic patients could be amenable to CAI surveillance (data not showed). Fourth, because the reference standard procedure (MRI with hepato-specific contrast medium) cannot be systematically used to verify the result of US and AFP as surveillance tools, our investigation – as almost all on this topic – suffers from the ‘verification bias’ that can artificially increase the sensitivity of both tests (Kobayashi *et al*, 1985). Finally, we provided a concise measurement of direct costs and effectiveness of the two surveillance strategies without considering indirect costs and costs of HCC diagnosis and treatment (Cucchetti *et al*, 2013). It should also be taken into account that the proposed CAI strategy consider the resumption of AFP monitoring when second-line US proved negative, and does not include CT or MRI as confirmatory tests of a negative US result. Therefore, specific cost-effectiveness studies, including all management costs of the disease, survival and quality of life and presenting different applicative scenario of imaging techniques in the case of a positive AFP test, are needed to refine the comparison of CAI \rightarrow US vs standard US surveillance.

In conclusion, HCC surveillance with US performed by properly trained operators remains the ideal solution, having a high sensitivity and an excellent specificity. Moreover, US has an unavoidable advantage over AFP: as the cancer grows, US sensitivity improves so that a false negative result may be corrected by the subsequent examination, a paradigm not so pertinent to AFP. However, for many national health-care systems it is impossible to realise the ideal solution because of the saturation of resources that it would require.

We therefore proposed a method to optimise the sensitivity of AFP in stable cirrhotic patients and limiting the access to US performed in specialised institutions, without paying an unacceptable price in terms of both missed tumour and cancer stage at diagnosis. In fact, CAI was obtained and validated in a population where more than 80% of patients had a very early/early stage HCC. Future studies should prospectively evaluate the use of AFP monitoring in combination with US performed by experts to make HCC surveillance more cost-effective and sustainable by many national health systems, until better tools become available.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Bruix J, Sherman M, Llovet JM, Beaugrand M, Lencioni R, Burroughs AK, Christensen E, Pagliaro L, Colombo M, Rodés J (2001) Clinical management of hepatocellular carcinoma: conclusions of the Barcelona-2000 EASL Conference. *J Hepatol* **35**: 421–430.
- Bruix J, Sherman M (2011) Management of hepatocellular carcinoma: an update. *Hepatology* **53**: 1020–1022.
- Chen TM, Huang PT, Tsai MH, Lin LF, Liu CC, Ho KS, Siau CP, Chao PL, Tung JN (2007) Predictors of alpha-fetoprotein elevation in patients with chronic hepatitis C, but not hepatocellular carcinoma, and its normalization after pegylated interferon alfa 2a-ribavirin combination therapy. *J Gastroenterol Hepatol* **22**: 669–675.
- Colli A, Fraquelli M, Casazza G, Massironi S, Colucci A, Conte D, Duca P (2006) Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review. *Am J Gastroenterol* **101**: 513–523.
- Cucchetti A, Cescon M, Erroi V, Pinna AD (2013) Cost-effectiveness of liver cancer screening. *Best Pract Res Clin Gastroenterol* **27**: 961–972.
- Di Bisceglie AM, Sterling RK, Chung RT, Everhart JE, Dienstag JL, Bonkovsky HL, Wright EC, Everson GT, Lindsay KL, Lok AS, Lee WM, Morgan TR, Ghany MG, Gretch DR (2005) Serum alpha-fetoprotein levels in patients with advanced hepatitis C: results from the HALT-C Trial. *J Hepatol* **43**: 434–441.
- European Association for the Study of the Liver; European Organization for Research and Treatment of Cancer (2012) EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* **56**: 908–943.
- Giannini EG, Cucchetti A, Erroi V, Garuti F, Odaldi F, Trevisani F (2013) Surveillance for early diagnosis of hepatocellular carcinoma: How best to do it? *World J Gastroenterol* **19**: 8808–8821.
- Gupta S, Bent S, Kohlwe J (2003) Test characteristics of alpha-fetoprotein for detecting hepatocellular carcinoma in patients with hepatitis C. A systematic review and critical analysis. *Ann Intern Med* **139**: 46–50.
- Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**: 839–843.
- Kamangar F, Dores GM, Anderson WF (2006) Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol* **24**: 2137–2150.
- Kobayashi K, Sugimoto T, Makino H, Kumagai M, Unoura M, Tanaka N, Kato Y, Hattori N (1985) Screening methods for early detection of hepatocellular carcinoma. *Hepatology* **5**: 1100–1105.
- Lee JM, Trevisani F, Vilgrain V, Wald C (2011) Imaging diagnosis and staging of hepatocellular carcinoma. *Liver Transpl* **17**: S34–S43.
- Lee E, Edward S, Singal AG, Lavieri MS, Volk M (2013) Improving screening for hepatocellular carcinoma by incorporating data on levels of α -fetoprotein, over time. *Clin Gastroenterol Hepatol* **11**: 437–440.
- Lok AS, Sterling RK, Everhart JE, Wright EC, Hoefs JC, Di Bisceglie AM, Morgan TR, Kim HY, Lee WM, Bonkovsky HL, Dienstag JL (2010) Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. *Gastroenterology* **138**: 493–502.
- Marrero JA, Feng Z, Wang Y, Nguyen MH, Befeler AS, Roberts LR, Reddy KR, Harnois D, Llovet JM, Normolle D, Dalhgren J, Chia D, Lok AS, Wagner PD, Srivastava S, Schwartz M (2009) Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* **137**: 110–118.
- Nguyen MH, Garcia RT, Simpson PW, Wright TL, Keeffe EB (2002) Racial differences in effectiveness of alpha-fetoprotein for diagnosis of hepatocellular carcinoma in hepatitis C virus cirrhosis. *Hepatology* **36**: 410–417.
- Poon D, Anderson BO, Chen LT, Tanaka K, Lau WY, Van Cutsem E, Singh H, Chow WC, Ooi LL, Chow P, Khin MW, Koo WH (2009) Asian Oncology Summit. Management of hepatocellular carcinoma in Asia: consensus statement from the Asian Oncology Summit 2009. *Lancet Oncol* **10**: 1111–1118.
- Pugh RN, Murray-Lyon IM, Dawson JL, Pietroni MC, Williams R (1973) Transection of the oesophagus for bleeding oesophageal varices. *Br J Surg* **60**: 646–649.

- Richardson P, Duan Z, Kramer J, Davila JA, Tyson GL, El-Serag HB (2012) Determinants of serum alpha-fetoprotein levels in hepatitis C-infected patients. *Clin Gastroenterol Hepatol* **10**: 428–433.
- Singal A, Volk ML, Waljee A, Salgia R, Higgins P, Rogers MA, Marrero JA (2009) Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. *Aliment Pharmacol Ther* **30**: 37–47.
- Stefanini GF, Amorati P, Biselli M, Mucci F, Celi A, Arienti V, Roversi R, Rossi C, Re G, Gasbarrini G (1995) Efficacy of transarterial targeted treatments on survival of patients with hepatocellular carcinoma. An Italian experience. *Cancer* **75**: 2427–2434.
- Teefey SA, Hildeboldt CC, Dehdashti F, Siegel BA, Peters MG, Heiken JP, Brown JJ, McFarland EG, Middleton WD, Balfe DM, Ritter JH (2003) Detection of primary hepatic malignancy in liver transplant candidates: prospective comparison of CT, MR imaging, US and PET. *Radiology* **226**: 533–542.
- Trevisani F, D'Intino PE, Morselli-Labate AM, Mazzella G, Accogli E, Caraceni P, Domenicali M, De Notariis S, Roda E, Bernardi M (2001) Serum alpha-fetoprotein for diagnosis of hepatocellular carcinoma in patients with chronic liver disease: influence of HBsAg and anti-HCV status. *J Hepatol* **34**: 570–575.
- Trevisani F, De Notaris S, Rapaccini G, Farinati F, Benvegnù L, Zoli M, Grazi GL, Del PP, Di N, Bernardi M (2002) Semiannual and annual surveillance of cirrhotic patients for hepatocellular carcinoma: effects on cancer stage and patient survival (Italian experience). *Am J Gastroenterol* **97**: 734–744.
- Zhang BH, Yang BH, Tang ZY (2004) Randomized controlled trial of screening for hepatocellular carcinoma. *J Cancer Res Clin Oncol* **130**: 417–422.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.