Predictive performance and clinical utility of HCC risk scores in chronic hepatitis C: a comparative study

Gamal Shiha1,2 · Nabiel N. H. Mikhail1,3 · Reham Soliman1,4 · Ayman Hassan1 · Mohammed Eslam5

Received: 9 June 2021 / Accepted: 6 December 2021 / Published online: 16 January 2022
© Asian Pacific Association for the Study of the Liver 2022

Abstract

Background and aim Many HCC risk prediction scores were developed to guide HCC risk stratification and identify CHC patients who either need intensified surveillance or may not require screening. There is a need to compare different scores and their predictive performance in clinical practice. We aim to compare the newest HCC risk scores evaluating their discriminative ability, and clinical utility in a large cohort of CHC patients.

Patients and methods The performance of the scores was evaluated in 3075 CHC patients who achieved SVR following DAAs using Log rank, Harrell’s c statistic, also tested for HCC-risk stratification and negative predictive values.

Results HCC developed in 212 patients within 5 years follow-up. Twelve HCC risk scores were identified and displayed significant Log rank \( p \leq 0.05 \) except Alonso-Lopez TE-HCC, and Chun scores \( p = 0.374, p = 0.053 \), respectively. Analysis of the remaining ten scores revealed that ADRES, GES pre-post treatment, GES algorithm and Watanabe (post-treatment) scores including dynamics of AFP, were clinically applicable and demonstrated good statistical performance; Log rank analysis < 0.001, Harrell’s C statistic (0.66–0.83) and high negative predictive values (94.38–97.65%). In these three scores, the 5 years cumulative IR in low risk groups be very low (0.54–1.6), so screening could be avoided safely in these patients.

Conclusion ADRES, GES (pre- and post-treatment), GES algorithm and Watanabe (post-treatment) scores seem to offer acceptable HCC-risk predictability and clinical utility in CHC patients. The dynamics of AFP as a component of these scores may explain their high performance when compared to other scores.

Keywords CHC · HCC risk scores · AFP

Introduction

Chronic hepatitis C infection is a major public health problem with an estimated 71 million persons chronically infected with hepatitis C worldwide [1]. Hepatitis C virus (HCV) is the most common cause of hepatocellular carcinoma (HCC), with an annual incidence of HCC is approximately 3–8% in cirrhotic patients [2]. The use of highly effective and safe direct acting antivirals (DAAs) had revolutionized the management of chronic HCV patients particularly in patients with liver cirrhosis and advanced hepatic fibrosis. The majority of patients with HCV infection are expected to be treated over the next years. Several studies found that viral clearance after DAAs lowered but did not completely eliminate the occurrence of HCC in post-SVR patients [3–6].

Current guidelines recommend biannual HCC surveillance by ultrasound with or without alpha-fetoprotein (AFP) in patients with cirrhosis [7, 8]. According to
cost-effectiveness analyses, an annual incidence of 1.5% or higher would warrant systematic surveillance of HCC [9]. These recommendations are backed up by data indicating improved survival, a higher rate of early tumor diagnosis and curative treatments among patients undergoing screening for HCC [10]. However, ‘one-size-fits-all’ strategy increases the health care costs particularly in low- to middle-income countries, with a high HCV prevalence, furthermore, it is estimated that a small percentage of patients with cirrhosis are monitored according to guidelines, highlighting the urgent unmet clinical need for a better prediction model to guide HCC surveillance among patients with advanced liver fibrosis who had SVR [11].

In this context, recently, many HCC risk prediction scores (Table 1) were developed to guide HCC risk stratification and identify CHC patients who either need intensified surveillance or may not require screening. In the literature, there is no data comparing these different scores, consequently there is a need for direct comparison of the performance, applicability, and clinical utility of these HCC risk scores in independent patient populations. This comparison is highly needed to help the health authorities to properly focus resources towards patients at high risk of development of HCC and to avoid screening of patients with very low HCC risk, also, the assessment of these scores will provide data which is necessary to support the mandatory modification of the current guidelines of HCC screening.

Our aim is to evaluate the newest HCC risk scores comparing their discriminative ability, applicability and clinical utility in a large cohort of CHC patients who achieved SVR following DAAs.

**Patients and methods**

**Cohort**

3075 consecutive CHC patients, with liver cirrhosis (F4) or advanced liver fibrosis (F3) who had a sustained virologic response (SVR) after receiving DAAs were included in this observational study.

Between January 2014 and July 2019, patients were recruited from the out-patient clinics at the Egyptian Liver Research Institute and Hospital (ELRIAH) and its satellites throughout the Nile Delta.

**Patients’ evaluation**

All patients in the cohort had their initial data (before treatment) recorded, together with the data in the follow-up visits, up to the last follow-up. HCC incidence, together with expected HCC incidence data, were also recorded. For score comparison, we depended on pre-treatment data (immediately before the onset of DAAs) and post-treatment data (24 weeks after end of treatment).

**Diagnosis of fibrosis and HCC**

According to the standard guidelines, patients were diagnosed as having advanced liver fibrosis (F3) by transient elastography (> 10.2 and ≤ 16.3 kPa) [13]. While, they were considered to have cirrhosis (F4) when they fulfilled more than one of the following criteria: (a) definite clinical signs and laboratory parameters of liver cirrhosis (e.g. splenomegaly, ascites, albumin ≤ 3.5 g/dL, platelets count ≤ 100 cmm3); (b) abdominal ultrasonographic signs suggestive of cirrhosis (e.g. mild splenomegaly, minimal ascites, PV dilatation and collaterals); (c) transient elastography (> 16.3 kPa).

The diagnosis of HCC was made in accordance with EASL and AASLD guidelines. Multiphase CT or MRI was done to the patient if there were any focal hepatic lesions diagnosed by abdominal ultrasound and/or AFP value > 20 ng/mL MSCT. Diagnosis of HCC was based on the characteristic arterial enhancement and early washout in delayed phase [2, 14, 15].

**Scores selection**

We undertook a systematic literature search of PubMed database for studies reporting on HCC prediction scores for HCC over the last 5 years. We used the following search terms: (“Hepatocellular carcinoma” [Mesh] AND “HCV” AND “risk score” [Mesh]. Citations generated by electronic scanning were assessed for relevance based on title, abstracts and key words. Prediction risk scores for which the parameters are available in our cohort were included.

**Statistical methods**

Statistical analyses were performed using version 26, SPSS (Statistical Package for Social Sciences) (IBM Corp., USA). The follow-up duration was calculated as the time between the end of treatment and the last follow-up, or the date of event development (HCC occurrence), whichever occurred first. Times to events and cumulative incidences were calculated with the Kaplan–Meier method.

The performance of the scores was evaluated using:

- Receiver operator curve analysis (ROC) for numeric scores to assess the HCC predictive ability of the score. Accuracy is measured by the area under the ROC curve (AUROC). An area of 1 represents a perfect test; an area of 0.5 represents a worthless test. A rough guide for
| Score                      | Data set                  | All patients | HCC Score components | Categories       | No./group | HCC IR                           | Harrell’s C |
|----------------------------|---------------------------|--------------|----------------------|------------------|-----------|---------------------------------|-------------|
| Sharma et al. [33] THRI   | Derivation (2000–2009)    | 2079, F4     | 226                  | Age, etiology,  | Low       | NA                              | 1.2%/5 years | 0.76          |
|                            |                           |              |                      | sex, PLT         | Intermediate |                                   |             |
|                            |                          |              |                      |                  | High       | NA                              | 4.4%/5 years | 0.77          |
|                            | Validation                | 1144, F4     | 107                  | Age, etiology,  | Low       | NA                              | 1.1%/5 years | 0.77          |
|                            |                           |              |                      | sex, PLT         | Intermediate |                                   |             |
|                            |                           |              |                      |                  | High       | 4.9%/5 years                    | 13.1%/5 years | 0.77          |
| Watanabe et al. [34]       | Pre-treatment             | 1174, 33     | 384 (32.7%)          | FIB4, Alb, sex  | Low       | 0.4% 2 years                    | NA          |
|                            | (2014–2017)               |              |                      |                  | Intermediate| 4.4% 2 years                    |             |
|                            |                           |              |                      |                  | High       | 13.1%/5 years                   |             |
| Watanabe et al. [34]       | Post-treatment            | 553, 33      | 311 (56.2%)          | FIB4,AFP         | Low       | 0.4% 2 years                    | NA          |
|                            | (2014–2017)               |              |                      |                  | Intermediate| 3.2% 2 years                    |             |
|                            |                           |              |                      |                  | High       | 14.4% 2 years                   |             |
| Hu et al. [35]             | Derivation (2002–2016)    | 665, 65      | 453 (68.1%)          | Age, bilirubin, | Low       | 1.1% 5 years                    | 4.2% 10 years| 0.82          |
|                            | Validation                | 78, 13       | 51 (65.4%)           | AFP, SVR, cirrhosis | Intermediate | 8.3% 10 years              | 30.4% 10 years|             |
|                            |                          |              |                      |                  | High       | 36.2% 10 years                  |             |
| Tani et al. [36]           | Post-treatment            | 1088, 26     | 630 (57.9%)          | Age, follow up  | Low       | 0.3% 2 years                    | NA          |
|                            | (2014–2018)               |              |                      | AFP              | Intermediate| 6.27% 2 years                   |             |
|                            |                           |              |                      |                  | High       | 18.37% 2 years                  |             |
| Fan et al. aMAP score [37] | Training (2014–2018)      | 3688, 95     | 2158 (58.9%)         | Age, sex, Alb,  | Low       | 0.3% 2 years                    | 0.82        |
|                            | Validation                | 13,686, 536  | 5348 (40.1%)         | Plt, bilirubin  | Intermediate| 1.5–4.8% 5 years              |             |
|                            |                          |              |                      |                  | High       | 8.1–17.8% 5 years              |             |
| Hiraoka et al. ADRES score | Training (2014–2017)      | 484, 22      | 134 (27.7%)          | Sex, FIB4,     | Very low | 0.0% 2 years                    | 0.835       |
|                            | Validation                | 585, 14      | 225 (46.5%)          | Follow up AFP   | Low       | 0% 2 years                      |             |
|                            |                          |              |                      |                  | Intermediate| 2.1% 2 years                    |             |
|                            |                           |              |                      |                  | High       | 15.9% 2 years                   |             |
|                            |                           |              |                      |                  | Very low   | 18.37% 2 years                  |             |
|                            |                           |              |                      |                  | Low        | 0.0% 2 years                    | 0.899       |
|                            |                           |              |                      |                  | Intermediate| 7.9% 2 years                    |             |
|                            |                           |              |                      |                  | High       | 19.5% 1 year                     |             |
|                            |                           |              |                      |                  | Very low   | 3.2% 1 year                     |             |
|                            |                           |              |                      |                  | Low        | 3.2% 1 year                     |             |
|                            |                           |              |                      |                  | Intermediate| 33.9% 4 years                   |             |
| Abe et al. [39]            | Post-treatment            | 181, NA      | 141 (77.9%)          | ALBI, Plt, DM   | Low       | 0.0% 2 years                    | NA          |
|                            | (2014–2019)               |              |                      |                  | High       | 33.9% 4 years                   |             |
| Shiha et al. GES score [40]| Training (2015–2018)      | 2372, 109    | 1368                 | Age, sex, Alb,  | Low       | 1.9% 3 years                    | 0.801       |
|                            | Internal validation       | 687, 14      | 590                  | AFP, fibrosis   | Intermediate| 5.8% 3 years                   |             |
|                            | (2017–2018)               |              |                      | stage           | High       | 9.5% 3 years                    |             |
|                            | External evaluation      | 1314, 46     | 478                  | Age, sex, Alb,  | Low       | 0.21% years                     | 0.812       |
|                            | (2015–2018)               |              |                      | AFP, fibrosis   | Intermediate| 2.11% years                     |             |
|                            |                           |              |                      | stage           | High       | 6.14% years                     |             |
|                            |                           |              |                      |                  | Low        | 0.22% years                     | 0.816       |
|                            |                           |              |                      |                  | Intermediate| 2.46% years                     |             |
|                            |                           |              |                      |                  | High       | 6.11% years                     |             |
|                            |                           |              |                      |                  | Low        | 1.23% 5 years                   | 0.832       |
|                            |                           |              |                      |                  | Intermediate| 2.93% 5 years                   |             |
|                            |                           |              |                      |                  | High       | 7.15% 5 years                   |             |
classifying the accuracy of a diagnostic test is using 0.7 to indicate fair, 0.8 to indicate good and 0.9 to indicate excellent. Acceptable discrimination is indicated when AUROC is > 0.70 [16].

• Evaluating the performance of the risk stratification as a screening procedure against HCC development as the gold standard. Using the risk stratification results, patients are classified into risky group (intermediate and high risk score) and less-risky group (low risk score) and then performance statistics (sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy) are calculated. NPV (the probability that subjects with a negative screening test truly do not have the disease) is of special importance in score comparisons [17].

• Log rank (Mantel–Cox) analysis for comparison of incidence curves. p value ≤ 0.05 was considered significant.

• Harrell’s C statistic [18]. The C-statistic only gives a general idea about a model (goodness of fit measure), especially its discrimination ability. A value below 0.5 indicates a very poor model. A value of 0.5 means that the model is no better than predicting an outcome than random chance. Values over 0.7 indicate a good model. Values over 0.8 indicate a strong model. A value of 1 means that the model perfectly predicts those group members who will experience a certain outcome and those who will not [19, 20].

• Overall performance by Brier score. The lower the Brier score is for a set of predictions, the better the predictions are calibrated [21].

• Calibration using Hosmer–Lemeshow test. The output returns a chi-square value (a Hosmer–Lemeshow chi-squared) and a p-value. Small p-values mean that the model is a poor fit. Small p-values (usually under 5%) mean that model is not a good fit [19].

### Comparison of the scores

The different scores will be evaluated in the same cohort in a stepwise manner:

• Statistical performance of the scores using the above mentioned methods.

• Scores that are statistically valid will be tested for its applicability and its clinical utility by: (a) studying its ability to stratify patients into the different risk groups either two groups only (high and low), three groups (high, intermediate and low) or four groups where the low is subdivided into low and very low; (b) calculating the percentage of patients in each category together with its 5-year cumulative incidence of HCC; and (c) reporting the AUROC and NPV of the score.

### Results

Using the previously mentioned search criteria, 95 papers resulted.

### Score exclusion criteria

- Prediction scores that depend on molecular and genetic risk factors which are expensive, not done routinely and were not available for patients in our cohort, like fat-genetic risk score [22], TLL1 [23, 24], IFNL3 [25, 26], MICA [27] and DEPDC5 [28].

- Scores that need data not routinely available in our dataset, as GGT (FIB4-HCC score by Alonso López et al. [29] and Ganne-Carrié et al. [30]).

- Scores that are depending on complicated mathematical methods like El-Serag et al. (HES) score [31] or the methods used for stratification of patients into risk groups was

| Score | Data set | All patients | HCC | Score components | Categories | No./group | HCC IR | Harrell’s C |
|-------|----------|--------------|-----|------------------|------------|-----------|-------|-------------|
| Chun et al. | Training (2003–2016) | 669 | 19 | Sex, AFP, FIB4 | Low | 315 (47.1%) | 0.01% 6 years | AUROC at 6 years = 0.911 |
| Chun et al. | Validation (2003–2016) | 524 | 11 | Sex, AFP, FIB4 | Intermediate | 309 (46.2%) | 0.03% 6 years |
| Chun et al. | | | | | High | 45 (6.7%) | 0.31% 6 years |
| Alonso López et al. [29] | TE-HCC (2015–2017) | 993 | 35 | Alb, LSM, delta LSM | Low | 218 (41.6%) | 0.01% 6 years | AUROC at 6 years = 0.809 |
| Alonso López et al. | Validation (2015–2017) | 524 | 11 | Alb, LSM, delta LSM | Intermediate | 274 (52.3%) | 0.07% 6 years |
| Alonso López et al. | | | | | High | 32 (6.1%) | 0.17% 6 years |
| Alonso López et al. | | | | | Very low | 194 (38.4%) | 0.0% 3 years |
| Alonso López et al. | | | | | Low | 330 (39.8%) | 2.1% 3 years |
| Alonso López et al. | | | | | Intermediate | 257 (31.0%) | 5.8% 3 years |
| Alonso López et al. | | | | | High | 49 (5.9%) | 16.3% 3 years |
not fully reported in the published papers as Ioannou et al. score [32].

Accordingly, eleven scores and an algorithm were included in this study (Table 1).

All these scores were developed and/or validated using data from HCV patients. Only Sharma et al. [33] THRI score was developed and validated using multiple etiologies and the etiology was included as a part of the score whereas Fan et al. aMAP score [37] was developed using HBV patients and it was validated in HBV, HCV and non-viral hepatitis patients. Patients were followed up in all score for more than a year, up to 7.3 years in Sharma et al. [33] THRI score [33] (validation group).

Cohort characteristics

The study included 3075 chronic HCV patients (1037 patients with F3 and 2038 with F4 stage) with SVR who met the inclusion criteria in our study between January 2014 and July 2019. Characteristics of the patients are shown in Supplementary Table 1. The mean observation period was 24.32 ± 12.14 months after the end of DAAs treatment (range 6–72 months). HCC developed in 212 cases during the study period. Characteristics of the patients according to the development of HCC are shown in Supplementary Table 2. Out of the 212 HCC cases identified during follow-up, 38 occurred during the first year of follow-up, 67 during the second year, 62 during the third year, and 26 during the fourth year, and 19 after the fourth year of follow-up.

Performance of eleven scores and an algorithm was compared and the results were presented in Tables 2 and 3, and Fig. 1. Most scores stratified risk of HCC into three strata (low, intermediate and high), except two scores (Alonso López et al. TE-HCC score and Hiraoka et al. ADRES score) that stratified risk into four strata (very low, low, intermediate, and high), and two scores (Hu et al. score and Abe et al. score) that stratified risk into two strata (low and high).

The performance of each score is listed in Tables 2 and 3

Except for the TE-HCC score of Alonso López et al. [29] and Chun et al. [42], all of the HCC risk scores investigated had adequate statistical performance with significant Log rank (Mantel–Cox) analysis for comparison of incidence curves (p value ≤ 0.05) except TE-HCC score of Alonso López et al. [29], and Chun et al. [42] (p = 0.374, p = 0.053, respectively). The scores of Sharma et al. (THRI score) [33], Watanabe et al. (pre-treatment) [34] and Fan et al. (aMAP score) [37] revealed strong statistical performance with highly significant Log rank (p < 0.001), Harrell’s c statistic ≥ 0.64, area under the ROC curve (AUROC) values ≥ 0.73 and very high negative predictive values (> 98%). However, these scores stratified less than 25% of our cohort into the low risk group suggesting the need for HCC surveillance in the remaining 75–85% of patients (Fig. 2). Including the majority of patients in the screening will not only lead to a diminished cost-effectiveness of surveillance program but also may impose substantial physical harms on patients including multiple CT/MRI. Recently Fan et al. agreed that stratifying most of the patients into the high and intermediate risk groups based on the aMAP score, would reduce the cost-effectiveness of the surveillance [44]. This finding showed that good statistical performance of any HCC risk score may not be directly translated to clinical usefulness. It is interesting that the same observation was emphasized during comparison of HCC risk scores in CHB [43].

Discussion

Direct comparison of the predictive performance and clinical utility of HCC risk scores in the same patient population have been recently reported in chronic hepatitis B [43]. To the best of our knowledge, this work is the first comparative study assessing the predictive performance of many HCC risk scores in the same large cohort of CHC patients who achieved SVR following DAAs with follow-up period of more than 5 years. Most of these scores showed acceptable performance for HCC prediction in our cohort and were able to stratify CHC patients into low and risky groups.

As regards discrimination ability as measured by Harrell’s c statistic, all results were in the range 0.5678–0.832, the lowest Harrell’s c statistic was that of Alonso López et al. TE-HCC score [29] and the best was that of Shiha et al. GES algorithm [41].

As regards calibration using Hosmer–Lemeshow test, significant p-value (poor fit) was noted for Sharma et al. THRI score [33] and Hu et al. score [35].

HCC-SVR score [42] and Alonso López et al. TE-HCC score [29]).

As regards discrimination ability as measured by Harrell’s c statistic, all results were in the range 0.5678–0.832, the lowest Harrell’s c statistic was that of Alonso López et al. TE-HCC score [29] and the best was that of Shiha et al. GES algorithm [41].

As regards calibration using Hosmer–Lemeshow test, significant p-value (poor fit) was noted for Sharma et al. THRI score [33] and Hu et al. score [35].
Conversely, the score of Tani et al. \[36\] and Abe et al. \[39\], stratified most of the patients 70.8–84% into the low risk group with 5 years cumulative IR (95% CI) of 2.13 (1.73–2.59) and 2.51 (2.00–3.10), respectively which is higher than all studied scores. A large number stratified into the low risk group means that a lot of risky patients will not be screened, consequently many HCC cases will be missed, so HCC surveillance cannot be safely avoided.

| Score                          | Risk group      | No./group (%) | HCC/group | HCC 5 years cumulative IR (95% CI) |
|-------------------------------|-----------------|---------------|-----------|-----------------------------------|
| Sharma et al. THRI score      | Low             | 766 (24.9%)   | 9         | 0.51 (0.23–1.02)                  |
|                               | Intermediate    | 2072 (67.4%)  | 165       | 3.25 (2.73–3.84)                  |
|                               | High            | 237 (7.7%)    | 38        | 6.47 (4.45–9.12)                  |
| Watanabe et al. pre-treatment score | Low           | 676 (22.0%)   | 6         | 0.40 (0.10–1.09)                  |
|                               | Intermediate    | 2077 (67.5%)  | 126       | 2.42 (1.86–3.09)                  |
|                               | High            | 322 (10.5%)   | 80        | 8.29 (5.95–11.25)                 |
| Watanabe et al. post-treatment score | Low          | 1462 (47.5%)  | 38        | 1.10 (0.68–1.68)                  |
|                               | Intermediate    | 1227 (39.9%)  | 97        | 3.11 (2.33–4.08)                  |
|                               | High            | 386 (12.6%)   | 77        | 7.58 (5.44–10.30)                 |
| Hu et al. score               | Low             | 1814 (59.0%)  | 70        | 1.70 (1.29–2.19)                  |
|                               | High            | 1261 (41.0%)  | 142       | 4.60 (3.80–5.51)                  |
| Tani et al. score             | Low             | 2178 (70.8%)  | 94        | 2.13 (1.73–2.59)                  |
|                               | Intermediate    | 1227 (39.9%)  | 97        | 3.11 (2.33–4.08)                  |
|                               | High            | 386 (12.6%)   | 77        | 7.58 (5.44–10.30)                 |
| Fan et al. aMAP score         | Low             | 521 (16.9%)   | 4         | 0.35 (0.09–0.95)                  |
|                               | Intermediate    | 1315 (42.8%)  | 38        | 1.17 (0.79–1.68)                  |
|                               | High            | 1239 (40.3%)  | 170       | 5.40 (4.50–6.44)                  |
| Hiraoka et al. ADRES score    | Very low        | 504 (16.4%)   | 6         | 0.54 (0.14–1.46)                  |
|                               | Low             | 1244 (40.5%)  | 36        | 1.18 (0.72–1.83)                  |
|                               | Intermediate    | 1032 (33.6%)  | 96        | 3.64 (2.72–4.79)                  |
|                               | High            | 295 (9.6%)    | 74        | 9.08 (6.49–12.38)                 |
| Abe et al. score              | Low             | 2604 (84.7%)  | 136       | 2.51 (2.00–3.10)                  |
|                               | High            | 471 (15.3%)   | 76        | 6.43 (4.76–8.50)                  |
| Shiha et al. GES pre-treatment score | Low        | 1857 (60.4%)  | 59        | 1.66 (1.27–2.12)                  |
|                               | Intermediate    | 719 (23.4%)   | 71        | 4.45 (3.50–5.57)                  |
|                               | High            | 499 (16.2%)   | 82        | 7.64 (6.11–9.43)                  |
| Shiha et al. GES post-treatment score | Low       | 1764 (57.4%)  | 46        | 1.35 (1.00–1.79)                  |
|                               | Intermediate    | 944 (30.7%)   | 68        | 3.49 (2.73–4.40)                  |
|                               | High            | 367 (11.9%)   | 98        | 11.09 (9.05–13.45)                |
| Shiha et al. GES algorithm    | Low             | 1574 (51.2%)  | 37        | 1.23 (0.88–1.67)                  |
|                               | Intermediate    | 609 (19.8%)   | 38        | 2.93 (2.11–3.99)                  |
|                               | High            | 892 (29.0%)   | 137       | 7.15 (6.03–8.43)                  |
| Chun et al. HCC-SVR score     | Low             | 957 (31.1%)   | 31        | 1.75 (1.21–2.45)                  |
|                               | Intermediate    | 1659 (54.0%)  | 120       | 3.52 (2.92–4.21)                  |
|                               | High            | 459 (14.9%)   | 61        | 5.79 (4.47–7.39)                  |
| Alonso López et al. TE-HCC score | Very low      | 490 (15.9%)   | 15        | 1.09 (0.44–2.27)                  |
|                               | Low             | 848 (27.6%)   | 42        | 1.82 (1.10–2.85)                  |
|                               | Intermediate    | 1236 (40.2%)  | 98        | 2.90 (2.10–3.90)                  |
|                               | High            | 501 (16.3%)   | 57        | 3.65 (2.37–5.39)                  |

Five scores are remaining, Watanabe et al. (post-treatment) \[34\], Hu et al. score \[35\], GES score \[40\], GES algorithm \[41\] and ADRES \[38\]; the score of Hu et al. stratified the patients into two categories only low and high-risk groups, consequently, a large number (41%) of our patient cohort should undergo more intense screening which may lead to reduced cost-effectiveness and increased physical harms. Screening of this large number of patients using HU
| Score                          | Brier score | Harrell’s C | HL test | Log rank p | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) | Accuracy (95% CI) | AUROC (95% CI) |
|-------------------------------|-------------|-------------|---------|------------|----------------------|----------------------|--------------|-------------|-------------------|----------------|
| Sharma et al. THRI score      | 0.6981      | 0.6505      | 0.006   | <0.001     | 95.75 (92.13–97.75)  | 8.78 (7.70–10.02)    | 98.83 (97.78–99.38) | 98.32 (97.78–99.38) | 31.22 (29.61–32.88) | 0.739 (0.704–0.775) |
| Watanabe et al. pre-treatment | 0.7268      | 0.6406      | 0.615   | <0.001     | 97.17 (93.96–98.70)  | 8.59 (7.53–9.78)     | 99.11 (98.08–99.59)  | 99.11 (98.08–99.59)  | 28.49 (26.92–30.11) | NA              |
| Watanabe et al. post-treatment| 0.4874      | 0.6693      | 0.708   | <0.001     | 82.08 (73.36–86.65)  | 7.25 (6.28–8.36)     | 94.38 (92.38–95.88)  | 94.38 (92.38–95.88)  | 26.41 (24.88–27.99) | NA              |
| Hu et al. score               | 0.3902      | 0.6158      | 0.001   | <0.001     | 66.98 (60.40–72.96)  | 11.26 (9.63–13.12)   | 96.14 (95.15–96.93)  | 96.14 (95.15–96.93)  | 61.33 (59.60–63.04) | 0.691 (0.651–0.731) |
| Tani et al. score             | 0.2839      | 0.6388      | 0.922   | <0.001     | 55.66 (48.93–62.19)  | 7.29 (7.13–74.39)    | 95.68 (94.75–96.46)  | 95.68 (94.75–96.46)  | 71.61 (69.99–73.18) | NA              |
| Fan et al. aMAP score         | 0.7778      | 0.6984      | 0.613   | <0.001     | 98.11 (95.25–99.26)  | 8.14 (7.15–9.27)     | 99.23 (98.04–99.70)  | 99.23 (98.04–99.70)  | 23.58 (22.11–25.11) | 0.765 (0.728–0.801) |
| Hiraoka et al. ADRES score    | 0.3964      | 0.7183      | 0.915   | <0.001     | 80.19 (71.31–85.00)  | 13.73 (11.72–15.63)  | 97.60 (96.77–98.22)  | 97.60 (96.77–98.22)  | 61.01 (59.27–62.72) | NA              |
| Abe et al. score              | 0.1721      | 0.5678      | 0.654   | 0.013      | 35.85 (29.70–42.50)  | 32.53 (29.70–35.35)  | 94.78 (93.85–95.57)  | 94.78 (93.85–95.57)  | 82.73 (81.36–84.03) | NA              |
| Shiha et al. GES pre-treatment| 0.3655      | 0.801       | 0.143   | <0.001     | 72.17 (65.78–77.77)  | 12.56 (10.82–14.54)  | 96.82 (95.92–97.53)  | 96.82 (95.92–97.53)  | 63.45 (61.73–65.13) | NA              |
| Shiha et al. GES post-treatment| 0.3873     | 0.828       | 0.126   | <0.001     | 78.30 (72.28–83.32)  | 12.66 (10.97–14.57)  | 97.39 (96.54–98.04)  | 97.39 (96.54–98.04)  | 61.27 (59.53–62.98) | NA              |
| Shiha et al. GES algorithm    | 0.4433      | 0.832       | 0.971   | <0.001     | 82.6 (76.9–87.1)     | 11.7 (10.1–13.4)     | 97.7 (96.8–98.3)     | 97.7 (96.8–98.3)     | 55.8 (53.9–57.4)   | NA              |
| Chun et al. HCC-SVR score     | 0.6400      | 0.6038      | 0.568   | 0.053      | 85.38 (79.99–89.50)  | 32.34 (30.65–34.08)  | 96.76 (95.44–97.71)  | 96.76 (95.44–97.71)  | 36.00 (34.32–37.71) | 0.660 (0.623–0.697) |
| Alonso López et al. TE-HCC score| 0.5428     | 0.5975      | 0.943   | 0.374      | 73.11 (66.77–78.63)  | 9.82 (7.67–10.36)    | 95.74 (94.52–96.70)  | 95.74 (94.52–96.70)  | 46.70 (44.94–48.47) | NA              |

*Comparing risky patients (high + intermediate risk groups) with less risky patients (low risk group)
Fig. 1 Cumulative risk of HCC according to different prediction scores

(b) Sharma et al. (THRI) score

(b) Watanabe et al. (pre-treatment) score

(c) Watanabe et al. (post-treatment) score

(d) Hu et al. score

(e) Tani et al. score

(f) Fan et al. (aMAP) score

(g) Hiraoka et al. (ADRES) score

(h) Abe et al. score

(i) Shiha et al. (GES pre-treatment)

(j) Shiha et al. (GES post-treatment)

(k) Shiha et al. (GES algorithm)

(l) Chun et al. (HCC-SVR) score

(m) Alonso López et al. (TE-HCC) score
Fig. 2 Flowchart of the evaluation and analysis of the eleven scores

**HCC risk scores In CHC patients**

- 12 Scores (table 1)
  - 10 Scores statistically valid:
    - Sharma et al.
    - Tani et al.
    - Watanabe et al. (post)
    - Hu et al.
    - Watanabe et al. (pre)
    - Fan et al.
    - Hiraoka et al.
    - Shiha et al. score
    - Shiha et al. algorithm
    - Abe et al.

**Analysis and Evaluation in 3075 CHC patients**

- 2 Scores not statistically valid:
  - Chun et al.
  - Alonso-López et al.
  - Log rank p-value > 0.05

- 10 Scores statistically valid:
  - Sharma et al.
  - Tani et al.
  - Watanabe et al. (post)
  - Hu et al.
  - Watanabe et al. (pre)
  - Fan et al.
  - Hiraoka et al.
  - Shiha et al. score
  - Shiha et al. algorithm
  - Abe et al.

**Potential Clinical Application**

- Not applicable for screening in our cohort

**Flowchart**

- 3 Scores with small low risk groups:
  - Sharma et al.
  - Watanabe et al. (pre-treatment)
  - Fan et al.
  - Low risk groups < 25% of patients
  - (16.9%-24.9% of patients)
  - Screening of most of the cohort (75-85% of patients)
  - Cost-effectiveness???

- 7 Scores:
  - Watanabe et al. (post)
  - Hu et al.
  - Tani et al.
  - Abe et al.
  - Shiha et al. score

- 2 Scores (Tani et al. and Abe et al.):
  - High risk group (0.4%-15.3% of patients)
  - Low risk group (70.8%-84.7% of patients)
  - Most patients stratified into the low risk group but 5y cum. IR is > 2.1-2.5 which is higher than all other scores
  - A lot of risky patients will not be screened. So, many HCC cases will be missed

- 5 Scores:
  - Watanabe et al. (post)
  - Hu et al.
  - Tani et al.
  - Shiha et al. GES score
  - Shiha et al. GES algorithm
  - Hiraoka et al. ADRES score

- 4 Scores:
  - Watanabe et al. (post)
  - Shiha et al. GES (pre and post) score
  - Shiha et al. GES algorithm
  - Hiraoka et al. ADRES score

- Hu et al. Score:
  - No intermediate risk group
  - The high risk group (41% of patients) will require more intense screening
  - Cost-effectiveness???

- Low risk group about 50% of patients with 5Y cum. IR < 1.7
- Intermediate risk group about 30% of patients
- High risk group (9.6%-29% of patients) 5Y cum. IR 7.15-9.08

- Screening can be avoided safely in 50% of the patients
- Screening in less than one third of patients to be continued as current guidelines
- Intense screening diagnosis of many patients with early HCC

**Statistically valid, applicable, can be utilized clinically for HCC risk-based individualized surveillance and potentially cost-effective**
et al. score could be explained by the lack of an intermediate risk group, highlighting the importance of this group during application of HCC risk scores in clinical practice.

Finally, after evaluation and comparison of these eleven HCC risk scores we ended with four scores namely Watanabe et al. (post-treatment), GES score, GES algorithm and ADRES score. These scores were clinically applicable being simple, easy to calculate and based on readily available clinical and laboratory parameters. In addition, these scores demonstrated good statistical performance; Log rank analysis < 0.001, Harrell’s C statistic (0.66–0.83) and high negative predictive value (94.38–97.65%). Also, these scores stratified our patients successfully into low, intermediate, and high groups with very low 5 years cumulative IR (0.54–1.6) in the low risk group which is about 50% of the cohort, so surveillance could be avoided safely in approximately half of the patients. On the other hand, the high-risk groups had high 5 years cumulative IR in about 20% of the patients only, for whom more intense screening may be required. These scores had intermediate risk group with relatively high 5 years cumulative IR in about one third of patients who may need to continue screening according the current guidelines. Also, these scores had relatively good Brier score (as an indication of overall performance) and non-significant Hosmer–Lemeshow test, indicating good calibration and that they had a good fit.

The good performance of GES score and algorithm may be explained not only by the fact that it was derived from a similar population with the same HCV genotype but also that it included both F4 and F3 in its components [40, 41]. It should be noted that the study cohort included both patients with F3 and F4 as well as the cohorts from which Watanabe et al. [34] and ADRES [38] scores were derived as FIB-4 > 3.25 i.e. F3 and F4.

There is an argument about HCC surveillance after SVR in individuals with cirrhosis (F4) vs. pre-cirrhotic advanced fibrosis (F3). EASL supports ongoing surveillance in patients with advanced fibrosis (F3) whereas AASLD does not [7]. Ioannou in his recent review of HCC surveillance after SVR in patients with F3 and F4 attempted to explain this important issue by reporting that this disagreement is due to the difficulty in precise determination of patients with F3 fibrosis and the fact that they are a heterogeneous group with some patients having F3–F4 fibrosis and higher HCC risk and others having F2–F3 fibrosis with lower risk. Furthermore, there is possibility for misclassification of cirrhosis; certain patients are under staged by biopsy or non-invasive markers of fibrosis and hence their risk of HCC is underestimated [11]. HCC that occurred in non-cirrhotic patients as advanced fibrosis (F3) often diagnosed at late stages due to low index of suspicious and lack of screening which is a significant gap in the clinical care. As a result so there is unmet need for scores that can help to identify patients with F3 who have a high enough HCC risk that warrant HCC surveillance.

It is not clear why Watanabe et al. (post-treatment) and ADRES scores had good statistical performance and clinical utility in our cohort although they were derived from totally different population and genotypes. However, careful analysis of the components of these scores showed that they are very similar including age, fibrosis stage and AFP; this similarity of the components of these HCC risk scores and its relation to the clinical utility was recently highlighted by Voulgaris et al. [43] who stated that in addition to the predictive performance, the components of each score and its formulas are crucial factors for the clinical utility of any risk score. Interestingly it is not AFP cut-off but its post-treatment changes and dynamics during the follow-up. Although AFP is the most widely used biomarker in HCC surveillance, it is not included in the international guidelines based on its suboptimal sensitivity and specificity [43–46]. However, recently several reports confirmed longitudinal AFP measurement rather than an absolute cut-off value [47, 48] may further increase the sensitivity of HCC detection. Based on these reports and our findings, we suggest that a good score for HCC risk prediction should include the dynamics of AFP during the follow-up. This suggestion may be reinforced by the observation that the scores of Watanabe et al., Fan et al. (aMAP score), and Sharma et al. (THRI score) which did not include AFP in their components were not clinically useful in our patient cohort although they displayed the highest statistical performance among all studied scores.

The strength of this study is that it is the first direct comparison of recently published HCC risk prediction scores in CHC. This work was done in the same large cohort of CHC patients with follow-up period more than 5 years allowing assessment of the statistical performance, applicability, clinical utility, and potential cost-effectiveness of these scores. The results of this work may pave the way for related methodology together with objective and validated clinical interpretation of HCC risk scores.

The study has some limitations; it reflects the judgment of one single center and most patients were predominantly genotype 4 so validation in cohorts of different ethnicities and other HCV-genotypes may be required before any further recommendations as HCV genotype showed increased HCC risk in different populations; Patients with HCV genotype 3 is known to be associated with a higher incidence of HCC [49]. Also, HCV genotype 6 increased the Risk for Hepatocellular Carcinoma among CHC Asian patients with liver cirrhosis [50]. Second; the good performance of GES score and algorithm may be explained by the fact that it was derived from similar patient population.

In conclusion, many scores including parameters which were not readily available or had complicated formulas
restricted their applicability in clinical practice. Statistical performance of HCC risk scores in CHC may not be directly transferable to clinical utility. ADRES, GES score, GES algorithm, Watanabe et al. (post-treatment) scores seem to offer high predictability and clinical utility for HCC in CHC patients. The dynamics of AFP as a component of these three scores may explain the good clinical usefulness when compared to other scores.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12072-021-10284-6.

Author contributions GS designed the study. GS and RS supervised clinical work. NM performed the statistical analyses. GS, NM, RS, AH and ME interpreted the data. AH supervised the laboratory work. All authors drafted the paper provided input into the manuscript and approved the final version.

Funding NA.

Declarations

Conflict of interest Gamal Shiha, Nabei N. M. Mikhail, Rehman Soliman, Ayman Hassan and Mohammed Eslam declare no conflicts of interest.

Ethical approval This study was conducted in accordance with the protocol and the principles of the Declaration of Helsinki [12] and its amendments in 2008. The protocol was approved by the Institutional Research Board of ELRIAH (IORG0008819, IRB00010534) in accordance with the local regulations. The need to obtain informed consent from the participants was waived by the IRB due to the retrospective nature of the study.

References

1. WHO. Global Hepatitis Report, 2017. Geneva: World Health Organization; 2017. http://www.who.int/hepatitis/publications/global-hepatitis-report2017/en/. Accessed 20 Mar 2021.
2. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of hepatocellular carcinoma. J Hepatol 2018;69(1):182–236.
3. Li DK, Ren Y, Fierer DS, Rutledge S, Shaikh OS, Lo Re V III, et al. The short-term incidence of hepatocellular carcinoma is not increased after hepatitis C treatment with direct-acting antivirals: an ERCHIVES study. Hepatology 2018;67(6):2244–2253.
4. Carrat F, Fontaine H, Dorival C, Simony M, Diallo A, Herode C, et al. Clinical outcomes in patients with chronic hepatitis C after direct-acting antiviral treatment: a prospective cohort study. Lancet 2019;393(10179):1453–1464.
5. Ide T, Koga H, Nakano M, Hashimoto S, Yatsuhashi H, Higuchi N, et al. Direct-acting antiviral agents do not increase the incidence of hepatocellular carcinoma development: a prospective, multicenter study. Hepatol Int 2019;13:293–301.
6. Shiha G, Mousa N, Soliman R, Mikhail NNH, Adel Elbashony M, Khattab M. Incidence of HCC in chronic hepatitis C patients with advanced hepatic fibrosis who achieved SVR following DAAs: a prospective study. J Viral Hepat 2020;27(7):671–679.
7. European Association for the Study of the Liver. EASL recommendations on treatment of hepatitis C 2018. J Hepatol 2018;69:461–511.
8. Ghany MG, Morgan TR, AASLD-IDSA Hepatitis C Guidance Panel. Hepatitis C guidance 2019 update: American Association for the Study of Liver Diseases-Infectious Diseases Society of America recommendations for testing, managing, and treating hepatitis C virus infection. Hepatology 2020;71:686–721.
9. Sarasin FP, Giostra E, Hadengue A. Cost-effectiveness of screening for detection of small hepatocellular carcinoma in western patients with Child-Pugh class A cirrhosis. Am J Med 1996;101(4):422–434.
10. Mittal S, Kanwal F, Ying J, Chung R, Sada YH, Temple S, et al. Effectiveness of surveillance for hepatocellular carcinoma in clinical practice: a United States cohort. J Hepatol 2016;65:1148–1154.
11. Ioannou G. HCC surveillance after SVR in patients with F3/F4 fibrosis. J Hepatol 2021;74:458–465.
12. CIOMS/WHO. International Ethical Guidelines for Biomedical Research Involving Human Subjects. Geneva: CIOMS; 1993.
13. Shiha G, Seif S, Maher M, Etreby S, Samir W, Zalata K. Comparison between transient elastography (Fibroscan) and liver biopsy for the diagnosis of hepatic fibrosis in chronic hepatitis genotype 4. Egypt Liver J 2014;4(4):106–111.
14. Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecasis MM, Roberts LR, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology 2019;67:358–380.
15. Omata M, Cheng AL, Kokudo N, Kudo M, Lee JM, Jia J, et al. Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. Hepatol Int 2017;11:317–370.
16. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5(9):1315–1316.
17. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017;5:307.
18. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247:2543–2546.
19. Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 2016. https://www.statisticshowto.com/c-statistic/. Accessed 20 Mar 2021.
20. Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev 1950;78:1–3.
21. Degasperi E, Galmozzi E, Pelusi S, D’Ambrosio R, Soffredini R, Borghi M, et al. Hepatic fat—genetic risk score predicts hepatocellular carcinoma in HCV cirrhotic patients treated with DAAs. Hepatology 2020;72(6):1912–1923.
22. Matsuura K, Sawai H, Ikeo K, Ogawa S, Iio E, Isogawa M, et al. Genome-wide association study identifies TLL1 variant associated with development of hepatocellular carcinoma after eradication of hepatitis C virus infection. Gastroenterology 2017;152(6):1383–1394.
23. Iio E, Matsuura K, Shimada N, Atsukawa M, Itokawa N, Abe H, et al. AASLD guidelines for the treatment of hepatocellular carcinoma after eradication of hepatitis C virus by interferon-free therapy. J Gastroenterol 2019;54(4):339–346.
24. Simili A, Mazzella G, Ravaoli F, Festi D, Bacchi-Reggiani ML, Porro A, et al. Interleukin 28B polymorphisms and hepatocellular carcinoma development after direct acting antiviral therapy for chronic hepatitis C. J Gastrointest Liver Dis 2019;28(4):449–456.
25. Qin S, Wang J, Zhou C, Xu Y, Zhang Y, Wang X, et al. The influence of interleukin 28B polymorphisms on the risk of hepatocellular carcinoma among patients with HBV or HCV.
infection: an updated meta-analysis. Medicine (Baltimore) 2019;98(38):e17275.

27. Luo X, Wang Y, Shen A, Deng H, Ye M. Relationship between the rs2596542 polymorphism in the MICA gene promoter and HBV/HCV infection-induced hepatocellular carcinoma: a meta-analysis. BMC Med Genet 2019;20(1):142.

28. Liu W, Ma N, Zhao D, Gao X, Zhang X, Yang L, et al. Correlation between the DEPDC5 rs1012068 polymorphism and the risk of HBV-related hepatocellular carcinoma. Clin Res Hepatol Gastroenterol 2019;43(4):446–450.

29. Alonso López S, Manzano ML, Gea F, Gutiérrez ML, Ahumada AM, Devesa MJ, et al. A model based on non-invasive markers predicts very low hepatocellular carcinoma risk after viral response in Hepatitis C virus-advanced fibrosis. Hepatology. 2020;72(6):1924–1934.

30. Ganne-Carriè N, Layese R, Bourcier V, Cagnot C, Marcellin P, Guyader D, et al. Nomogram for individualized prediction of hepatocellular carcinoma occurrence in hepatitis C virus cirrhosis (ANRS CO12 CirVir). Hepatology 2016;64(4):1136–1147.

31. El-Serag HB, Kanwal F, Davila JA, Kramer J, Richardson P. A new laboratory-based algorithm to predict development of hepatocellular carcinoma in patients with hepatitis C and cirrhosis. Gastroenterology 2014;146(5):1249–1255.e1.

32. Ioannou GN, Green PK, Beste LA, Mun EJ, Kerr KF, Berry K. Development of models estimating the risk of hepatocellular carcinoma after antiviral treatment for hepatitis C. J Hepatol 2018;69(5):1088–1098.

33. Sharma SA, Kowgier M, Hansen BE, Brouwer WP, Maan R, Wong D, et al. Toronto HCC risk index: a validated scoring system to predict 10-year risk of HCC in patients with cirrhosis. J Hepatol 2017;S0168–8278(17):32248–32251.

34. Watanabe T, Tokumoto Y, Joko K, Michitaka K, Horiike N, Tanaka Y, et al. Predictors of hepatocellular carcinoma occurrence after direct-acting antiviral therapy in patients with hepatitis C virus infection. Hepatol Res 2019;49(2):136–146.

35. Hu CC, Weng CH, Hua MC, Chang PH, Lin CL, Chen YT, et al. New scoring method to predict risk of hepatocellular carcinoma occurrence in patients with chronic hepatitis C after pegylated interferon and ribavirin therapy. J Interferon Cytokine Res 2020;40(2):82–91.

36. Tani J, Morishita A, Sakamoto T, Nakahara M, Fujita K, et al. Simple scoring system for prediction of hepatocellular carcinoma occurrence after hepatitis C virus eradication by direct-acting antiviral treatment: all Kagawa Liver Disease Group Study. Oncol Lett 2020;19(3):2205–2212.

37. Fan R, Papatheodoridis G, Sun J, Innes H, Toyoda H, Xie Q, et al. aMAP risk score predicts hepatocellular carcinoma development in patients with chronic hepatitis. J Hepatol 2020;S0168–8278(20):30478–30485.

38. Hiraoka A, Kumada T, Ogawa C, Kariyama Y, Morita M, Nousu K, et al. Proposed a simple score for recommendation of scheduled ultrasonography surveillance for hepatocellular carcinoma after direct acting antivirals: multicenter analysis. J Gastroenterol Hepatol 2019;34(2):436–441.

39. Abe K, Wakabayashi H, Nakayama H, Suzuki T, Kuroda M, Yoshida N, et al. Factors associated with hepatocellular carcinoma occurrence after HCV eradication in patients without cirrhosis or with compensated cirrhosis. PLoS One 2020;15(12):e0243473.

40. Shiha G, Waked I, Soliman R, Elbasiony M, Gomaa A, Mikhail NNH, et al. GES: a validated simple score to predict risk of HCC in patients with HCV-GT4 associated advanced liver fibrosis after oral antivirals. Liver Int 2020;40:2828–2833.

41. Shiha G, Soliman R, Mikhail NNH, Hassan A, Elsam L. Development of a simple dynamic algorithm for individualized HCC risk-based surveillance using pre- and post-treatment GES score. Liver Int 2021;41(11):2768–2776.

42. Chun HS, Kim BK, Park JY, Kim DY, Ahn SH, Han KH, et al. Design and validation of risk prediction model for hepatocellular carcinoma development after sustained virological response in patients with chronic hepatitis C. Eur J Gastroenterol Hepatol 2020;32(3):378–385.

43. Voulgaris T, Papatheodoridi M, Lampertico P, Papatheodoridis GV. Clinical utility of hepatocellular carcinoma risk scores in chronic hepatitis B. Liver Int 2020;40(3):484–495.

44. Fan R, Yin X, Hou J. Reply to “External validation of aMAP risk score in Chronic hepatitis C genotype 4 patients with liver cirrhosis who achieved SVR following DAAs”. J Hepatol 2021;74(4):996–997.

45. Bruix J, Sherman M, American Association for the Study of Liver Diseases. Management of hepatocellular carcinoma: an update. Hepatology 2011;53(3):1020–1022.

46. European Association for the Study of the Liver; European Organisation for Research and Treatment of Cancer. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol 2012;56(4):908–943.

47. Chang TS, Wu YC, Tung SY, Wei KL, Hsieh YY, Huang HC, et al. Alpha-fetoprotein measurement benefits hepatocellular carcinoma surveillance in patients with cirrhosis. Am J Gastroenterol 2015;110(6):836–844.

48. Biselli M, Conti F, Gramenzi A, Frigerio M, Cucchiatti A, Fatti G, et al. A new approach to the use of alpha-fetoprotein as surveillance test for hepatocellular carcinoma in patients with cirrhosis. Br J Cancer 2015;112(1):69–76.

49. Kanwal F, Kramer JR, Ilyas J, Duan Z, El-Serag HB. HCV genotype 3 is associated with an increased risk of cirrhosis and hepatocellular cancer in a national sample of U.S. Veterans with HCV. Hepatology 2014;60(1):98–105.

50. Lee MH, Hsiao TI, Subramaniam SR, Le AK, Vu VD, Trinh HN, et al. HCV genotype 6 increased the risk for hepatocellular carcinoma development after sustained virological response in patients with chronic hepatitis C. J Hepatol 2017;69(5):1088–1098.