

# An Assessment of Benefits and Harms of Hepatocellular Carcinoma Surveillance in Patients With Cirrhosis

Omair Atiq,<sup>1</sup> Jasmin Tiro,<sup>2,3</sup> Adam C. Yopp,<sup>4</sup> Adam Muffler,<sup>1</sup> Jorge A. Marrero,<sup>1</sup> Neehar D. Parikh,<sup>4</sup> Caitlin Murphy,<sup>3</sup> Katharine McCallister,<sup>3</sup> and Amit G. Singal<sup>1-3</sup>

Although surveillance ultrasound and alpha fetoprotein (AFP) tests have minimal direct harm, downstream harms from follow-up tests must be weighed against surveillance benefits when determining the value of hepatocellular carcinoma (HCC) screening programs. Our study's aims were to characterize prevalence and correlates of surveillance benefits and harms in cirrhosis patients undergoing HCC surveillance. We conducted a retrospective cohort study among patients with cirrhosis followed at a safety-net health system between July 2010 and July 2013. We recorded surveillance-related benefits, defined as early tumor detection and curative treatment, and surveillance-related physical harms, defined as computed tomography or magnetic resonance imaging scans, biopsies, or other procedures performed for false-positive or indeterminate surveillance results. Sociodemographic and clinical correlates of surveillance harms were evaluated using multivariable logistic regression. We identified 680 patients with cirrhosis, of whom 78 (11.5%) developed HCC during the 3-year study period. Of the 48 (61.5%) HCCs identified by surveillance, 43.8% were detected by ultrasound, 31.2% by AFP, and 25.0% by both surveillance tests. Surveillance-detected patients had a higher proportion of early HCC (70.2% vs. 40.0%;  $P = 0.009$ ), with no difference in tumor stage between ultrasound- and AFP-detected tumors ( $P = 0.53$ ). Surveillance-related physical harms were observed in 187 (27.5%) patients, with a higher proportion of ultrasound-related harm than AFP-related harm (22.8% vs. 11.4%;  $P < 0.001$ ). Surveillance-related harms were associated with elevated ALT (odds ratio [OR], 1.87; 95% confidence interval [CI], 1.26-2.76), thrombocytopenia (OR, 2.06; 95% CI, 1.26-3.38), and hepatology subspecialty care (OR, 1.63; 95% CI, 1.09-2.42). **Conclusion:** Over one fourth of patients with cirrhosis experience physical harm for false-positive or indeterminate surveillance tests—more often related to ultrasound than AFP. Interventions are needed to reduce surveillance-related harm to increase the value of HCC screening programs in clinical practice. (HEPATOLOGY 2017;65:1196-1205).

Primary liver cancer is the second-leading cause of cancer-related death worldwide and the fifth-leading cause in the United States.<sup>(1)</sup> Incidence of hepatocellular carcinoma (HCC), the most common type of primary liver cancer, is rapidly increasing in the United States, and it is projected to become the third-leading cause of cancer-related death by 2030.<sup>(2)</sup> The prognosis for patients with HCC depends on tumor stage at diagnosis, with curative options only available for patients diagnosed at an early stage.<sup>(3)</sup> Several societies, including the American Association for the Study of Liver Diseases (AASLD) and

*Abbreviations:* AASLD, American Association for the Study of Liver Diseases; AFP, alpha fetoprotein; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; CI, confidence interval; CT, computed tomography; EMR, electronic medical record; FIT, fecal immunochemical test; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; HE, hepatic encephalopathy; ICD-9, International Classification of Diseases, Ninth Revision; INR, international normalized ratio; LT, liver transplantation; MRI, magnetic resonance imaging; NASH, nonalcoholic steatohepatitis; NCCN, National Comprehensive Cancer Network; OR, odds ratio; TACE, transarterial chemoembolization.

Received August 1, 2016; accepted October 13, 2016.

Supported by the AHRQ Center for Patient-Centered Outcomes Research (R24 HS022418). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or AHRQ.

Copyright © 2016 by the American Association for the Study of Liver Diseases

View this article online at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).

DOI 10.1002/hep.28895

*Potential conflict of interest:* Dr. Singal consults, advises, and is on the speakers' bureau for Bayer. He is on the speakers' bureau and received grants from Gilead. He advises Wako Diagnostics. Dr. Kono advises Wako Diagnostics. Dr. Yopp is on the speakers' bureau for Bayer. He received grants from Peregrine, Merck, and Novartis.

National Comprehensive Cancer Network (NCCN), recommend surveillance using ultrasound, with or without alpha fetoprotein (AFP), at 6-month intervals in patients with cirrhosis.<sup>(4,5)</sup> Several studies evaluating HCC surveillance among patients with cirrhosis demonstrate an association with improved early detection and overall survival, but were retrospective in design with inherent limitations, including lead-time bias, length-time bias, and short follow-up duration.<sup>(6)</sup> Notably, these studies only measured HCC surveillance benefits and did not characterize potential physical, financial, and/or psychological harms.<sup>(6,7)</sup>

Data for both benefits and harms are needed to determine the value of cancer screening programs.<sup>(8)</sup> Experience with other cancer screening programs demonstrates the potential for significant physical and financial harms. For example, use of the fecal immunochemical test (FIT) in colorectal cancer screening has minimal direct harms, but follow-up colonoscopy among those with abnormal FIT is associated with risk of perforation, bleeding, and anesthesia complications.<sup>(9,10)</sup> Similarly, HCC surveillance using ultrasound and AFP has minimal discomfort and no direct physical harms; however, there are potential “downstream” harms associated with diagnostic evaluation protocols. Liver lesions found on ultrasound are typically evaluated with computed tomography (CT) and/or magnetic resonance imaging (MRI), which are associated with radiation exposure, contrast injury, and cost.<sup>(11,12)</sup> If a liver lesion cannot be definitively characterized on cross-sectional imaging, patients may undergo biopsy, which is associated with risks of bleeding, tumor seeding, and injury to nearby organs.<sup>(13,14)</sup> Although the imperfect sensitivity (~60%-65%) and specificity (~70%-95%) of surveillance tests and potential for physical harms from these procedures have been acknowledged, no study has quantified the

frequency or severity of these harms as adverse outcomes directly related to HCC surveillance in clinical practice.<sup>(4,6,7)</sup> Therefore, the aim of our study was to characterize prevalence and correlates of HCC surveillance benefits and physical harms related to follow-up diagnostic testing in patients with cirrhosis.

## Patients and Methods

### STUDY POPULATION

We conducted a retrospective cohort study of patients with cirrhosis followed at Parkland Health and Hospital System, the safety-net health system for Dallas County. Parkland is an integrated health system comprised of 12 primary care provider clinics in low-income neighborhoods, a hepatology outpatient clinic, a multidisciplinary HCC clinic, and a tertiary hospital—all sharing the same comprehensive electronic medical record (EMR). Parkland currently provides inpatient and outpatient care for over 2000 patients with cirrhosis in Dallas. Parkland offers a sliding fee scale program, which provides access to primary and subspecialty care, including HCC surveillance and diagnostic testing, at low cost for uninsured Dallas County residents.

Patients with cirrhosis were identified by a set of International Classification of Diseases, Ninth Revision (ICD-9) codes, which are highly sensitive and specific for cirrhosis (456.0, 456.1, 456.2, 456.21, 567.23, 571.2, 571.5, 572.2, 572.3, and 572.4).<sup>(15)</sup> One author (O.A.) adjudicated cases to confirm they met diagnostic criteria for cirrhosis, defined as stage 4 fibrosis on liver biopsy or a cirrhotic-appearing liver on abdominal imaging with signs of portal hypertension (e.g., varices, ascites, or splenomegaly). All patients were required to have at least one outpatient clinic visit

#### ARTICLE INFORMATION:

From the <sup>1</sup>Department of Internal Medicine, UT Southwestern Medical Center and Parkland Health Hospital System, Dallas, TX; <sup>2</sup>Harold C. Simmons Cancer Center UT Southwestern Medical Center, Dallas, TX; <sup>3</sup>Department of Clinical Sciences, UT Southwestern Medical Center, Dallas, TX; <sup>4</sup>Department of Surgery, UT Southwestern Medical Center, Dallas, TX; and Department of Medicine, University of Michigan, Ann Arbor, MI.

#### ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Amit G. Singal, M.D., M.S.  
Division of Digestive and Liver Diseases,  
University of Texas Southwestern  
5959 Harry Hines Boulevard  
POB 1

Suite 420  
Dallas TX 75390-8887  
E-mail: amit.singal@utsouthwestern.edu  
Tel: +1-214-645-6029

and one HCC surveillance test between July 2010 and July 2011 to demonstrate that Parkland was their medical home. This study was approved by the Institutional Review Board of UT Southwestern Medical Center (Dallas, TX).

## DATA COLLECTION

We manually abstracted information on patient demographics, clinical history, laboratory data, and imaging results from the EMR. All records were reviewed by one investigator (O.A.) and independently verified by a second investigator (A.S.). Discrepancies were resolved through discussion to establish consensus.

### HCC Surveillance Receipt

Dates of all HCC surveillance tests between July 2010 and July 2013 were abstracted. HCC surveillance at Parkland is typically performed using ultrasound, with or without AFP, per the AASLD guidelines<sup>(4)</sup> with low use of surveillance CT or MRI. We manually reviewed imaging orders, imaging reports, and associated clinical notes to determine intent of ultrasound exams and AFP (surveillance vs. diagnostic) and test results. Ultrasounds with indications including “surveillance,” “screening,” “rule out HCC,” and “cirrhosis” were classified as surveillance exams. Imaging exams performed for diagnostic reasons, for example, abdominal pain or elevated liver enzymes, were classified as nonsurveillance cases. We recorded whether ultrasounds were normal (no suspicious masses), positive (suspicious liver mass  $\geq 1$  cm), or indeterminate (mass  $< 1$  cm or unclear if mass is present, e.g., coarse echo texture). AFP results were considered positive if  $\geq 20$  ng/mL, the most common cutoff used for HCC surveillance in clinical practice,<sup>(16)</sup> and indeterminate if  $\geq 11$  ng/mL, the upper limit of normal, but  $< 20$  ng/mL.

### Benefits of HCC Surveillance

Benefits of HCC surveillance included the: (1) proportion of HCC patients detected at an early tumor stage and 2) proportion of HCC patients eligible for curative treatment. Patients with cirrhosis diagnosed with HCC during the study period were identified using ICD-9 codes for HCC (155.0) and a prospectively maintained list of all HCC patients seen in the Parkland Multidisciplinary Liver Tumor Clinic.<sup>(17)</sup> All HCC cases were adjudicated to confirm that they met diagnostic criteria based on AASLD guidelines.<sup>(4)</sup>

Tumor characteristics, including tumor nodules, maximum diameter, and presence of vascular invasion or distant metastases, were determined by imaging studies interpreted by radiologists at our institution, and early-stage HCC was defined using Milan Criteria, the most common criteria for liver transplantation (LT) in the United States. Treatment of HCC was categorized as LT, surgical resection, local ablative therapy, transarterial chemoembolization (TACE), systemic chemotherapy, or best supportive care. HCC treatment was considered curative if it consisted of LT, surgical resection, or local ablative therapy. In patients who received multiple treatments, we used a trumping algorithm based on survival benefit (LT  $>$  surgical resection  $>$  local ablative therapy  $>$  TACE  $>$  systemic chemotherapy).

### Physical Harms of HCC Surveillance

Using test indication and test results, we identified the subset of patients who had a surveillance test that was classified as abnormal. A binary outcome of physical harm was defined for each surveillance test result per person. Physical harms included any follow-up tests (CT, MRI, liver biopsy, or angiogram) performed for false-positive or indeterminate surveillance results. AASLD and European Association for the Study of the Liver guidelines both recognize the low yield of diagnostic testing and recommend short-interval repeat ultrasound for indeterminate results; therefore, follow-up diagnostic tests for indeterminate surveillance results (e.g., mass  $< 1$  cm or nodular coarse echo texture without definite mass) were classified as physical harms. We recorded all tests performed for follow-up of surveillance results during the study period, so it was possible for patients to have more than one follow-up test and physical harm.

There is variation in clinical significance among measured physical harms. For example, a liver biopsy complication is more clinically significant than theoretical radiation harm from a single CT scan. To account for different degrees of harm based on exposure to radiation and invasive procedures, we also described surveillance-related harm as an ordinal variable (no harm, mild harm, moderate harm, and severe harm). “No harm” was defined as patients without any follow-up CT, MRI, or biopsy for positive or indeterminate surveillance tests; “mild harm” as those who have a single diagnostic CT or MRI encounter without complications; “moderate harm” as those who underwent multiple CT and/or MRI exams; and “severe harm”

was defined as those who undergo invasive procedures, such as liver biopsy or angiogram, for false-positive or indeterminate tests.

## Correlates of Surveillance Harms

Age, sex, race, and ethnicity were recorded for each patient. Body mass index (BMI) was calculated using height and weight at the index visit and dichotomized (obese vs. nonobese) using a cutoff of 30. Data regarding underlying liver disease etiology, presence of decompensation (ascites or hepatic encephalopathy [HE]), and receipt of hepatology care were abstracted from laboratory data and clinical notes. We classified patients according to etiology of liver disease, including hepatitis C virus (HCV), hepatitis B virus (HBV), alcohol-related liver disease, nonalcoholic steatohepatitis (NASH), and other. HCV infection was defined by the presence of a positive HCV antibody, viral load, or genotype. HBV infection was defined by the presence of HBV surface antigen or viral load. Patients were determined to have alcohol-related cirrhosis if they had a documented history of heavy alcohol use in the clinical notes. Patients were classified as NASH if they had evidence of metabolic syndrome in the absence of HCV infection, HBV infection, or a heavy alcohol history. Degree of ascites and HE was categorized as none, mild, or controlled on medications or severe/uncontrolled per clinical notes. Laboratory data of interest from time of index visit included platelet count, creatinine, aspartate aminotransferase (AST), alanine aminotransferase (ALT), bilirubin, albumin, and international normalized ratio (INR).

## STATISTICAL ANALYSIS

We reported point estimates of surveillance-related benefit and physical harms for the whole cohort and stratified by surveillance test type (ultrasound vs. AFP) and test result (false positive vs. indeterminate). In recognition of the debate concerning whether to use AFP in conjunction with ultrasound,<sup>(20)</sup> we report stratified estimates by surveillance type (ultrasound vs. AFP). Furthermore, the stratified analysis can be informative because ultrasound and AFP may be done at different times, with follow-up testing recommended if either surveillance test is positive. We estimated the proportion of physical harms by test result (false positive vs. indeterminate) because rationale for follow-up testing would likely differ and may require intervention strategies. For example, harms attributed to false-positive

results may require surveillance tests with higher specificity, but harms attributed to indeterminate results could potentially be minimized by provider education to discourage nonguideline concordant care. Multivariable logistic regression models were constructed to identify patient-level factors associated with physical harm. In a secondary analysis, we performed multivariable ordinal logistic regression to define patient-level correlates of harm when defined as a four-level outcome (none, mild, moderate, and severe). Final models included covariates significant on univariable analysis and those considered clinically important *a priori* (obesity, cirrhosis etiology, Child Pugh score, and hepatology care). Statistical significance was defined as  $P < 0.05$ . All statistical analysis was performed using the statistical software, Stata (version 11.2; StataCorp LP, College Station, TX).

## Results

### PATIENT CHARACTERISTICS AND RECEIPT OF SURVEILLANCE

A total of 680 patients with cirrhosis met inclusion criteria (Table 1). Mean age of patients was 54.3 years, and two thirds (64.7%) were men. The cohort was racially diverse, consisting of 32.5% non-Hispanic whites, 22.9% blacks, and 42.1% Hispanic Caucasians. The most common etiologies of cirrhosis were HCV infection (56.2%), alcohol-induced liver disease (25.7%), and NASH (11.6%). Median Child Pugh score was 7 (interquartile range, 6-8), with 29.9% of patients having Child Pugh A cirrhosis and 57.1% Child Pugh B cirrhosis. Patients were followed for a mean of  $26.7 \pm 11.7$  months. At least one surveillance ultrasound had been performed in 523 (76.9%) patients, and 640 (94.1%) had  $\geq 1$  serum AFP measurement; however, only 179 (26.3%) patients had  $\geq 3$  surveillance ultrasound exams and only 11 (1.6%) had  $\geq 6$  surveillance ultrasound exams during the 3-year follow-up period. Overall, 78 (11.5%) patients developed HCC during the 3-year study period.

### BENEFITS OF HCC SURVEILLANCE

Tumors were detected by surveillance in 48 (61.5%) of the 78 patients who developed HCC during the follow-up. Of these, 21 (43.8%) were detected by ultrasound alone, 15 (31.2%) by AFP alone, and 12 (25.0%) by both ultrasound and AFP (Fig. 1). The

**TABLE 1. Patient Sociodemographic and Clinical Characteristics, Overall and Stratified by Occurrence of Surveillance-Related Physical Harm**

Characteristic	All Patients (n = 680)	Patients Without Physical Harm (n = 493)	Patients With Physical Harm (n = 187)	P Value*
Age, years	54.3 ± 9.4	54.3 ± 9.7	54.3 ± 8.6	0.95
Sex (% male)	440 (64.7)	316 (64.1)	124 (66.3)	0.59
Race/ethnicity				0.49
Non-Hispanic white	221 (32.5)	168 (34.1)	53 (28.4)	
Black	156 (22.9)	109 (22.1)	47 (25.1)	
Hispanic	286 (42.1)	203 (41.2)	83 (44.4)	
Other/unknown	17 (2.5)	13 (2.6)	4 (2.1)	
BMI				0.38
<25	166 (24.5)	121 (24.6)	45 (24.2)	
25.0-29.9	225 (33.2)	161 (32.7)	64 (34.4)	
30.0-34.9	158 (23.3)	122 (24.8)	36 (19.4)	
≥35	129 (19.0)	88 (17.9)	41 (22.0)	
Etiology of liver disease				0.30
Hepatitis C	382 (56.2)	265 (53.7)	117 (62.6)	
Hepatitis B	22 (3.2)	16 (3.3)	6 (3.2)	
Alcohol-related	175 (25.7)	135 (27.4)	40 (21.4)	
NASH	79 (11.6)	59 (12.0)	20 (10.7)	
Other	22 (3.2)	18 (3.6)	4 (2.1)	
Child Pugh class				0.93
A	203 (29.8)	148 (30.0)	55 (29.4)	
B	388 (57.1)	282 (57.2)	106 (56.7)	
C	89 (13.1)	63 (12.8)	26 (13.9)	
Presence of HE (%)	154 (22.7)	110 (22.3)	44 (23.5)	0.74
Presence of ascites (%)	270 (39.7)	204 (41.4)	66 (35.3)	0.15
Receipt of hepatology care	441 (65.5)	307 (62.7)	134 (73.2)	0.01
Platelet count (×10 <sup>9</sup> /L)	110 ± 64	114 ± 67	103 ± 53	0.06
Thrombocytopenia	522 (78.0)	362 (74.8)	160 (86.5)	0.001
Creatinine (mg/dL)	1.0 ± 0.8	1.1 ± 0.9	0.9 ± 0.5	0.02
AST (U/L)	78 ± 72	73 ± 64	92 ± 89	0.002
AST >35 U/L	564 (83.7)	399 (81.9)	165 (88.2)	0.05
ALT (U/L)	58 ± 50	53 ± 45	73 ± 61	<0.001
ALT >35 U/L	412 (61.0)	277 (56.8)	135 (72.2)	0.001
Bilirubin (mg/dL)	1.8 ± 2.6	1.9 ± 2.8	1.6 ± 2.1	0.35
INR	1.3 ± 0.4	1.3 ± 0.5	1.3 ± 0.4	0.65
No. of HCCs	78 (11.5)	63 (12.8)	15 (8.0)	0.08

\*P value comparing patients with and without any surveillance-related physical harm.

remaining 30 HCC cases were detected incidentally or presented symptomatically. The majority (70.2%) of HCC patients detected by surveillance ultrasound and/or AFP had early HCC, compared to only 40.0% for those detected symptomatically or incidentally ( $P = 0.009$ ). There was not a significant difference in the proportion of HCC within Milan criteria between ultrasound-detected and AFP-detected tumors (76.2% vs. 66.7%;  $P = 0.53$ ; Fig. 2). Similarly, patients detected with surveillance were more likely to undergo curative treatment than non-surveillance-detected patients (22.9% vs. 0%;  $P = 0.005$ ), with no difference

in curative treatment receipt by surveillance modality ( $P = 0.43$ ).

### PHYSICAL HARMS OF HCC SURVEILLANCE

Physical harms related to false-positive or indeterminate surveillance results are shown in Table 2 and Fig. 1. Of all 680 patients, physical harm was observed in 187 (27.5%) patients, with 22 (3.2%) subjected to multiple CT scans, 8 (1.2%) multiple MRI scans, and 36 (5.3%) a combination of CT and MRI scans.

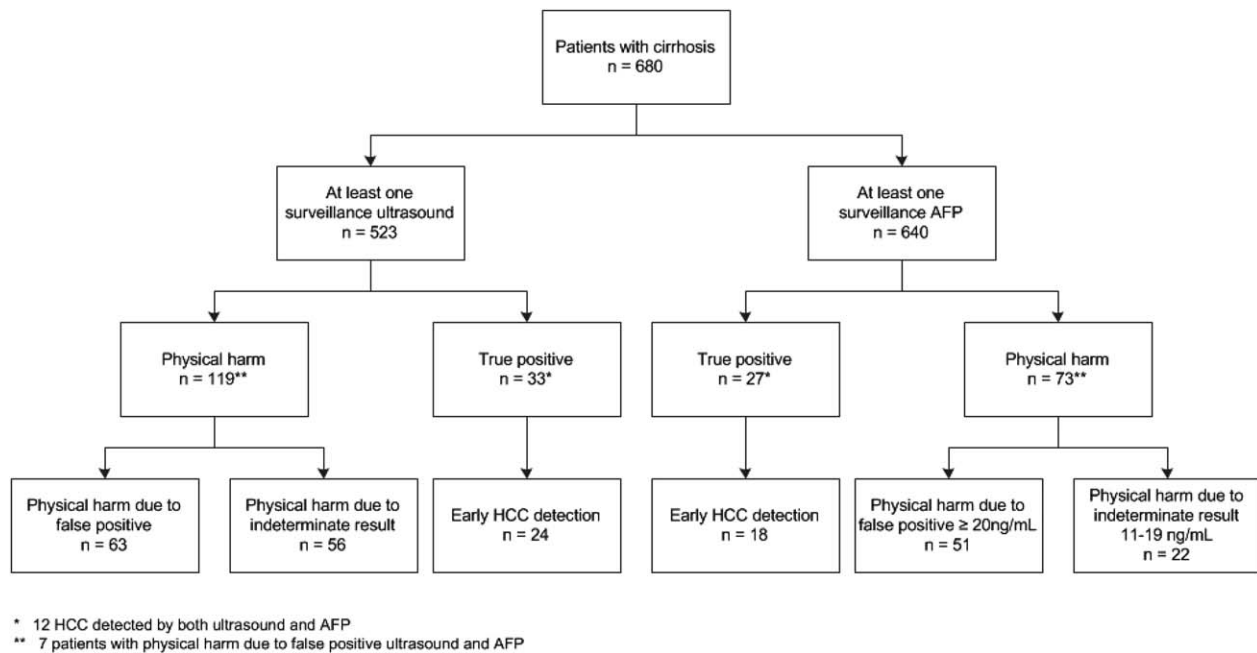


FIG. 1. Flow diagram of HCC surveillance benefits and harms in a cohort of patients with cirrhosis.

Although most harm was mild to moderate, 2 patients underwent biopsy of liver lesions (after two and four MRI exams) and 1 underwent an angiogram (after five MRI exams). As expected, the proportion of patients experiencing physical harm increased with the number

of surveillance exams from 11.9% among those with one surveillance exam to 29.6% among those with two to nine exams to 61.0% among those with  $\geq 10$  surveillance exams.

There were differences in the proportion of patients experiencing physical harm by surveillance modality, with a significantly higher proportion of ultrasound-related physical harm than AFP-related harm ( $P < 0.001$ ; Fig. 2). Of the 523 patients with  $\geq 1$  surveillance ultrasound, ultrasound-related physical harms were observed in 119 (22.8%) patients—73 with mild harm, 44 with moderate harm, and 2 with severe harm (both liver biopsies). Diagnostic evaluation was triggered by false-positive ultrasounds in 63 of these cases, and an additional 56 underwent diagnostic evaluation for indeterminate results. Indeterminate results included 35 patients with heterogeneous, nodular liver echotexture and 21 with subcentimeter liver nodules. Among patients with  $\geq 1$  serum AFP measurement ( $n = 640$ ), 73 (11.4%) experienced AFP-related physical harms—49 with mild harm, 23 with moderate harm, and 1 with severe harm (angiogram). Similar to ultrasound, AFP-related harm was attributed to a combination of false positives and indeterminate results. Only 51 patients with AFP-related harm had AFP levels exceeding 20 ng/mL, with 22 undergoing diagnostic evaluation for intermediate AFP elevations between 11

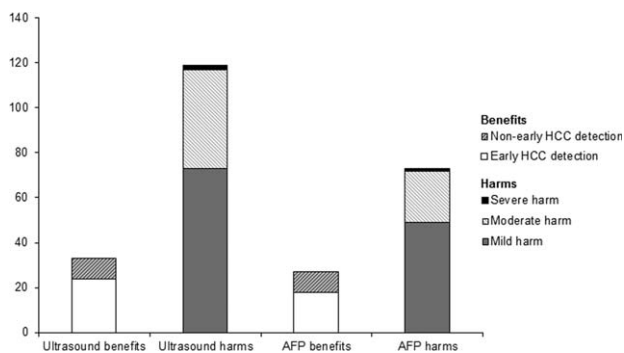


FIG. 2. Benefits and physical harms of HCC surveillance, stratified by surveillance modality. There was not a significant difference in the proportion of HCC detected at an early stage by surveillance modality (76.2% vs. 66.7% for ultrasound and AFP respectively;  $P = 0.53$ ); however, the proportion of patients experiencing ultrasound-related physical harm was significantly higher than AFP-related physical harm (22.8% vs. 11.4%;  $P < 0.001$ ).

**TABLE 2. Mild, Moderate, and Severe Physical Harms of HCC Surveillance, Stratified by Surveillance Modality and False Positive Versus Indeterminate Result**

Characteristic	AFP (n = 640)		Ultrasound (n = 523)	
	False Positive	Indeterminate	False Positive	Indeterminate
Any harm	51*	22	63*	56
Mild harm, n				
Single four-phase CT	26	18	32	28
Single MRI	3	2	6	7
Moderate harm, n				
Multiple four-phase CT	4	1	7	10
Multiple MRI	1	1	5	1
Four-phase CT and MRI	16	0	11	10
Severe harm				
Biopsy of liver mass, n	0	0	2	0
Hepatic angiogram, n	1	0	0	0

\*Seven patients with physical harm related to false-positive ultrasound and AFP were included in both groups.

and 20 ng/mL. Of note, 6 of 7 patients with both false-positive ultrasound and AFP had moderate harm with multiple CT and/or MRI exams performed for diagnostic evaluation.

### CORRELATES OF PHYSICAL HARMS

In univariable analyses, physical harm from false-positive or indeterminate surveillance results was significantly associated with elevated ALT level, thrombocytopenia, receipt of hepatology care, and viral etiology of cirrhosis. In multivariable analysis, physical harm was associated with elevated ALT level (odds ratio [OR], 1.87; 95% confidence interval [CI], 1.26-2.76), thrombocytopenia (OR, 2.06; 95% CI, 1.26-3.38), and receipt of hepatology care (OR, 1.63; 95% CI, 1.09-2.42). A secondary analysis evaluating harm as an ordinal outcome similarly found an association with elevated ALT (OR, 1.92; 95% CI, 1.30-2.83), thrombocytopenia (OR, 2.18; 95% CI, 1.34-3.55), and receipt of hepatology care (OR, 1.74; 95% CI, 1.17-2.57).

In exploratory subgroup analyses, we evaluated whether these associations were driven by false-positive/indeterminate AFP or ultrasound results. AFP-related harm was associated with viral etiology of cirrhosis (OR, 5.25; 95% CI, 2.31-11.92) and elevated ALT (OR, 2.84; 95% CI, 1.39-5.80), whereas ultrasound-related harm was associated with nonviral etiologies of cirrhosis (OR, 1.59; 95% CI, 1.03-2.44) and thrombocytopenia (OR, 2.14; 95% CI, 1.17-3.90).

### Discussion

This study quantified and weighed physical harms of HCC surveillance against HCC early detection in a large cohort of patients with cirrhosis. Although HCC surveillance detected over 60% of HCC and nearly doubled early tumor detection rates, over one fourth of patients experienced surveillance harms for false-positive or indeterminate results and nearly 10% had moderate-to-severe harm. Prevalence of surveillance harms increased steadily over time, increasing from ~10% among those with one surveillance test to >50% among those with 10 or more surveillance exams. Although surveillance harms were largely related to false-positive ultrasound or AFP results, harms were compounded by diagnostic imaging for indeterminate surveillance results, including nonguideline concordant follow-up for subcentimeter lesions or intermediate AFP elevations.

Complementary data regarding benefits and harms are essential to determine the value of cancer screening programs.<sup>(8)</sup> Experiences with breast and prostate cancer screening, in which evolving data about screening-related harms created controversy about published screening guidelines and altered clinical practice, highlight the importance of evaluating screening-related harms in advance of guideline recommendations and widespread use.<sup>(18)</sup> However, similar to the early evaluations for breast, colon, and prostate cancer screening programs, data for HCC surveillance have focused on surveillance-related benefits to date.<sup>(19)</sup> A meta-analysis identified nearly 50 studies characterizing the association between HCC surveillance and early detection, curative treatment, and overall survival among

patients with cirrhosis; however, the researchers noted a lack of data regarding surveillance-related harms as a high-priority area for research.<sup>(7)</sup> These data are of particular importance given that the benefits of HCC surveillance appear to be modest in patients with cirrhosis.<sup>(6,20)</sup> Our study begins to address this need by characterizing physical harms of HCC surveillance.

HCC surveillance was responsible for tumor detection in approximately 60% of HCC patients and increased early tumor detection rates from 40% to 70%. Tumor detection was attributed to ultrasound alone in nearly half of cases, AFP in one quarter, and both tests in the one quarter of patients. Early detection and curative treatment receipt did not significantly differ between ultrasound and AFP, although this may have been related to small sample size. Similar to previous studies,<sup>(20)</sup> our data suggest that AFP is complementary to ultrasound and increases the effectiveness of HCC surveillance for early tumor detection in clinical practice.

In terms of surveillance harms, ultrasound and AFP had a similar proportion of false-positive results; however, the harms of ultrasound were compounded by a high number of indeterminate findings, including nodular coarse echotexture that precluded definite exclusion of any liver masses and nonguideline concordant management of subcentimeter lesions. We noted radiologists often recommending diagnostic imaging with multiphase CT or MRI for cases with nodular coarse echotexture. Further data and guidance for what constitutes an inadequate ultrasound examination are likely needed to help radiologists distinguish cases in which ultrasound is sufficient, despite liver nodularity, and cases in which further imaging would be beneficial. We also observed high utilization of diagnostic CT and MRI in patients with subcentimeter lesions despite guidelines recommending repeat short-interval ultrasound given the low risk of HCC. This provider behavior may stem from several causes, including lack of knowledge about the guidelines, fear of medicolegal liability, hypervigilance to find HCC at an early stage, and perceived higher positive predictive value for ultrasound than AFP.<sup>(21,22)</sup> Many studies have discussed the suboptimal specificity of AFP, resulting in providers using clinical judgment when interpreting “low-level” positive AFP values.<sup>(23,24)</sup> However, there are less data discussing false-positive results related to ultrasound imaging, so providers may place more importance on following up any liver lesions, including those that are subcentimeter, given fear of potentially missing HCC at an early stage.<sup>(25)</sup>

Although surveillance-related harms were observed in nearly one fourth of patients, harms were particularly likely in some subsets, including patients receiving hepatology subspecialty care, patients with elevated ALT levels, and those with portal hypertension and thrombocytopenia. The association between hepatology care and surveillance harms may be mediated by higher provider awareness of HCC risk and a lower threshold for ordering diagnostic imaging.<sup>(26,27)</sup> Previous studies have reported higher rates of false-positive AFP in patients with viral hepatitis, hepatic inflammation, and elevated liver enzymes.<sup>(28)</sup> Elevated AFP levels should be cautiously interpreted in these patients, although AFP-adjusted algorithms or tailoring AFP cutoff by liver disease etiology may help reduce rates of unnecessary diagnostic imaging.<sup>(23,29)</sup> Increased liver nodularity in patients with advanced Child Pugh class and thrombocytopenia can impair radiologists’ ability to definitively exclude liver lesions, leading to recommendations for cross-sectional imaging.<sup>(30,31)</sup> Alternative surveillance tools for these patients are particularly needed given both lower sensitivity and specificity related to poor visualization. Although viral etiology was not associated with increased physical harms in multivariable analysis, we noted an association with increased AFP-related harm and lower ultrasound-related harm in exploratory subgroup analyses. Given that the epidemiology of HCC shifts from HCV-related to NASH-related, it is possible that the proportion of physical harm attributed to ultrasound may increase further.

Our study adds to the literature highlighting a need for better surveillance tools, with improved sensitivity for early tumor detection, and improved specificity to avoid unnecessary diagnostic tests. Over one third of HCC cases in our study presented incidentally or symptomatically. Suboptimal surveillance tool sensitivity is one of the most common reasons for late-stage tumor presentation in academic centers, prompting some to adopt CT and MRI as surveillance modalities despite a lack of supporting data.<sup>(21,32)</sup> Our study also highlights the potential for physical harms from both ultrasound and AFP, in part related to suboptimal surveillance test specificity. Although some may argue the physical harms of CT or MRI imaging is minimal, some patients experienced severe harm with biopsy and/or angiogram. Furthermore, patients may have also experienced psychological harms while awaiting diagnostic evaluation, although this was not measured in our study. Several biomarkers are currently being evaluated, but most have yet to undergo phase III or



IV biomarker studies and may be years removed from being fully validated and ready for routine clinical use.<sup>(33)</sup> Furthermore, data evaluating any harm related to these biomarkers is also largely unknown. While awaiting newer surveillance tools, our study suggests that over 40% of surveillance harms is related to non-guideline concordant management of indeterminate surveillance results, so provider education may be a simple intervention reduce surveillance-related harms in the interim.

Our study had limitations that must be taken into consideration when interpreting the results. First, the study was conducted in a single safety-net health system and its results may not be generalizable to other health systems. Second, surveillance can result in physical, financial, and psychosocial harms, but our study was limited to retrospective data available in the EMR and therefore focused on physical harms. Furthermore, physical harms were largely limited to receipt of diagnostic testing, with less data available to assess downstream harms such as contrast-induced renal failure. Third, patients may have potentially received HCC surveillance and/or diagnostic tests at outside institutions, although this is unlikely because many patients did not have insurance and thus would have to pay out of pocket to get care outside of the safety-net health system in Dallas. Finally, only one fourth of patients in our study had three or more surveillance exams during the 3-year study period, and it is possible, if not likely, that the magnitude of surveillance benefits and harms would be greater in settings with higher surveillance rates. However, the low surveillance rates observed in our study are consistent with previous studies.<sup>(27,34,35)</sup> Overall, we feel these limitations are outweighed by the strengths of the study, particularly in its characterization of HCC surveillance-related harms in patients with cirrhosis.

In summary, HCC surveillance is associated with early tumor detection and increased curative treatment receipt; however, these benefits must be weighed against surveillance harms. Nearly one fourth of non-HCC patients underwent diagnostic testing for false-positive or indeterminate surveillance results and nearly 10% had multiple diagnostic tests. Although false-positive ultrasound and AFP results were the most common causes, nonguideline concordant management of indeterminate ultrasound results accounted for nearly one third of cases with surveillance-related harm. While awaiting more accurate surveillance tools for early tumor detection, provider education may help

reduce surveillance-related harms and improve the value of HCC surveillance in patients with cirrhosis.

## REFERENCES

- 1) American Cancer Society. Global Cancer Facts & Figures, 3rd ed. Atlanta, GA: American Cancer Society; 2015
- 2) Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* 2014;74:2913-2921.
- 3) Singal AG, Marrero JA. Recent advances in the treatment of hepatocellular carcinoma. *Curr Opin Gastroenterol* 2010;26:189-195.
- 4) Bruix J, Sherman M. management of hepatocellular carcinoma: an update. *HEPATOLOGY* 2010;53:1-35.
- 5) 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;56:908-943.
- 6) Singal AG, Pillai A, Tiro J. Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: a meta-analysis. *PLoS Med* 2014;11:e1001624.
- 7) Kansagara D, Papak J, Pasha AS, O'Neil M, Freeman M, Relevo R, et al. Screening for hepatocellular carcinoma in chronic liver disease: a systematic review. *Ann Intern Med* 2014;161:261-269.
- 8) Harris RP, Wilt TJ, Qaseem A. A value framework for cancer screening: advice for high-value care from the American College of Physicians. *Ann Intern Med* 2015;162:712-717.
- 9) Denis B, Gendre I, Sauleau EA, Lacroute J, Perrin P. Harms of colonoscopy in a colorectal cancer screening programme with faecal occult blood test: a population-based cohort study. *Dig Liver Dis* 2013;45:474-480.
- 10) Marshall KG. Population-based fecal occult blood screening for colon cancer: will the benefits outweigh the harm? *CMAJ* 2000; 163:545-546.
- 11) Forner A, Vilana R, Ayuso C, Bianchi L, Sole M, Ayuso JR, et al. Diagnosis of hepatic nodules 20 mm or smaller in cirrhosis: prospective validation of the noninvasive diagnostic criteria for hepatocellular carcinoma. *HEPATOLOGY* JAN 2008;47:97-104.
- 12) Sangiovanni A, Manini MA, Iavarone M, Romeo R, Forzenigo LV, Raquelli M, et al. The diagnostic and economic impact of contrast imaging techniques in the diagnosis of small hepatocellular carcinoma in cirrhosis. *Gut* 2010;59:638-644.
- 13) Buscarini L, Fornari F, Bolondi L, Colombo P, Livraghi T, Magnolfi F, et al. Ultrasound-guided fine-needle biopsy of focal liver lesions: techniques, diagnostic accuracy and complications. A retrospective study on 2091 biopsies. *J Hepatol* 1990;11:344-348.
- 14) Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD. Liver biopsy. *HEPATOLOGY* 2009;49:1017-1044.
- 15) Nehra MS, Ma Y, Clark C, Amarasingham R, Rockey DC, Singal AG. Use of administrative claims data for identifying patients with cirrhosis. *J Clin Gastroenterol* 2013;47:e50-e54.
- 16) Gupta S, Bent S, Kohlwes J. Test characteristics of alpha-fetoprotein for detecting hepatocellular carcinoma in patients with hepatitis C. A systematic review and critical analysis. *Ann Intern Med* 2003;139:46-50.
- 17) Yopp AC, Mansour JC, Beg MS, Arenas J, Trimmer C, Reddick M, et al. Establishment of a multidisciplinary hepatocellular carcinoma clinic is associated with improved clinical outcome. *Ann Surg Oncol* 2014;21:1287-1295.

- 18) Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998;338:1089-1096.
- 19) Heleno B, Thomsen MF, Rodrigues DS, Jorgensen KJ, Brodersen J. Quantification of harms in cancer screening trials: literature review. *BMJ* 2013;347:f5334.
- 20) Singal AG, Conjeevaram HS, Volk ML, Fu S, Fontana RJ, Askari F, et al. Effectiveness of hepatocellular carcinoma surveillance in patients with cirrhosis. *Cancer Epidemiol Biomarkers Prev* 2012;21:793-799.
- 21) Joshi K, Mendler M, Gish R, Loomba R, Kuo A, Patton H, Kono Y. Hepatocellular carcinoma surveillance: a national survey of current practices in the USA. *Dig Dis Sci* 2014;59:3073-3077.
- 22) Dalton-Fitzgerald E, Tiro J, Kandunoori P, Halm EA, Yopp A, Singal AG. Practice patterns and attitudes of primary care providers and barriers to surveillance of hepatocellular carcinoma in patients with cirrhosis. *Clin Gastroenterol Hepatol* 2015;13:791-798.
- 23) Gopal P, Yopp AC, Waljee AK, Chiang J, Nehra M, Kandunoori P, Singal AG. Factors that affect accuracy of alpha-fetoprotein test in detection of hepatocellular carcinoma in patients with cirrhosis. *Clin Gastroenterol Hepatol* 2014;12:870-877.
- 24) Sherman M. Alpha-fetoprotein: an obituary. *J Hepatol* 2001;34:603-605.
- 25) Singal A, Volk ML, Waljee A, Salgia R, Higgins P, Rogers MA, Marrero JA. Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. *Aliment Pharmacol Ther* 2009;30:37-47.
- 26) Patel N, Yopp AC, Singal AG. Diagnostic delays are common among patients with hepatocellular carcinoma. *J Natl Compr Canc Netw* 2015;13:543-549.
- 27) Singal AG, Yopp A, Skinner CS, Packer M, Lee WM, Tiro JA. Utilization of hepatocellular carcinoma surveillance among American patients: a systematic review. *J Gen Intern Med* 2012;27:861-867.
- 28) Di Bisceglie AM, Sterling RK, Chung RT, Everhart J, Dienstag JL, Bonkovsky HL, et al. Serum alpha-fetoprotein levels in patients with advanced hepatitis C: results from the HALT-C Trial. *J Hepatol* 2005;43:434-441.
- 29) 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;146:1249-1255.
- 30) Del Poggio P, Olmi S, Ciccarese F, Di Marco M, Rapaccini GL, Benvegna L, Borzio F, et al. Factors that affect efficacy of ultrasound surveillance for early-stage hepatocellular carcinoma in patients with cirrhosis. *Clin Gastroenterol Hepatol* 2014;12:1927-1933.
- 31) Simmons O, Fetzter DT, Yokoo T, Marrero JA, Yopp A, Kono Y, et al. Predictors of adequate ultrasound quality for hepatocellular carcinoma surveillance in patients with cirrhosis. *Alimentary Pharmacology and Therapeutics* 2016 November 8. doi: 10.1111/apt.13841. [Epub ahead of print]
- 32) Singal AG, Nehra M, Adams-Huet B, Yopp AC, Tiro JA, Marrero JA, et al. Detection of hepatocellular carcinoma at advanced stages among patients in the HALT-C trial: where did surveillance fail? *Am J Gastroenterol* 2013;108:425-432.
- 33) Rich N, Singal AG. Hepatocellular carcinoma tumour markers: current role and expectations. *Best Pract Res Clin Gastroenterol* 2014;28:843-853.
- 34) Singal AG, Li X, Tiro JA, Kandunoori P, Adams-Huet B, Nehra MS, Yopp A. Racial, social, and clinical determinants of hepatocellular carcinoma surveillance. *Am J Med* 2015;128:90e1-90e7.
- 35) Davila JA, Morgan RO, Richardson PA, Du XL, McGlynn KA, El-Serag HB. Utilization of surveillance for hepatocellular carcinoma among patients with cirrhosis in the United States. *HEPATOLOGY* 2010;52:132-141.