

Presenter: Matthew Davenport, M.D.

Title of Abstract: **Repeatability of Major Diagnostic Features and Scoring Systems for Hepatocellular Carcinoma**

Abstract: **Using Dynamic Contrast-Enhanced Magnetic Resonance Imaging**

Institution: University of Michigan Health System

Authors: Matthew S. Davenport M.D., Shokoufeh Khalatbari M.S., Peter S.C. Liu M.D., Katherine E. Maturen M.D., Ravi K. Kaza M.D., Ashish P. Wasnik M.D., Mahmoud M. Al-Hawary M.D., Daniel I. Glazer M.D., Erica Stein M.D., Jeet Patel M.D., Deepak K. Somashekar M.D., Benjamin L. Viglianti M.D. Ph.D., William J. Weadock M.D., Hero K. Hussain M.D.

Modality: MR

Organ System: GI

Intro: An imaging-based diagnosis of hepatocellular carcinoma (HCC) can lead to locoregional therapy and/or liver transplantation without tissue confirmation. Recent efforts led by the American College of Radiology (ACR) and organ procurement and transplantation network (OPTN) have created liver observation scoring systems to improve the specificity of imaging for the diagnosis of HCC. The repeatability of these scoring systems (Liver Imaging Reporting and Data System [LI-RADS], OPTN) and the constituent components that comprise them (e.g., arterial enhancement) have not been assessed. Good repeatability is critical if these systems are to be successfully implemented.

Purpose: To determine the repeatability of major diagnostic features and scoring systems for hepatocellular carcinoma (HCC) using magnetic resonance imaging (MR).

Methods Used: Institutional review board approval was obtained and patient consent waived for this HIPAA-compliant, retrospective study. One hundred dynamic contrast-enhanced (gadobenate dimeglumine) liver MRI studies with and without comparison studies performed between 2/25/2011 and 1/12/2013 demonstrating liver observations preliminarily assigned LI-RADS (Liver Imaging Reporting and Data System) 1.1 scores of 1 to 5 (n=20 for each integer 1-5) were included. Ten blinded readers (n=5 expert, n=5 novice) reviewed the target observations. Maximum diameter (by post-contrast phase) and presence of major HCC features (arterial enhancement, washout appearance, pseudocapsule) were recorded. LI-RADS and OPTN (organ procurement and transplantation network) scores were assigned. Inter-reader agreement was assessed using intra-class correlation coefficients and Kappa statistics. Scoring rates were compared using McNemar's test.

Results of Abstract: Overall inter-reader agreement was good for arterial enhancement (0.67 [95%CI: 0.65-0.69]), and moderate for washout appearance (0.48 [95%CI: 0.46-0.50]) and pseudocapsule (0.52 [95%CI: 0.50-0.54]). Overall inter-reader agreement was fair for LI-RADS (0.35 [95%CI: 0.335-0.37]) and moderate for OPTN (0.53 [95%CI: 0.51-0.55]). Inter-reader size agreement was excellent for all post-contrast imaging phases (0.95-0.97). Expert readers agreed more than novice readers (LI-RADS: 0.43 [95%CI: 0.41-0.45] vs. 0.35 [95%CI: 0.33-0.37]; OPTN: 0.64 [95%CI: 0.60-0.68] vs. 0.50 [95%CI: 0.46-0.54]). Novices were less likely to assign a diagnosis of hepatocellular carcinoma than experts (LI-RADS 5: 17% [85/500] vs. 28% [140/500], p<0.0001; OPTN 5: 17% [83/500] vs. 29% [143/500], p<0.0001).

Discussion: Major qualitative MR imaging features for HCC have fair to moderate inter-reader agreement. Expert readers agree more than novice readers, but only experts using OPTN have good agreement. Novices are less likely to assign LI-RADS 5 and OPTN 5 scores than experts.

Scientific and/or Clinical Significance? Our study demonstrates that major qualitative diagnostic criteria for hepatocellular carcinoma are not easily repeatable across readers, regardless of training level. We also show that the new LI-RADS 1.1 scoring system for liver observation reporting is not consistently applied by readers, regardless of experience level.

Relationship to existing work This work helps define the critically important topic of repeatability for HCC liver observation scoring systems.