

BIostatISTICS STATEMENT

Name of Journal: World Journal of Gastroenterology

ESPS Manuscript NO: 35566

Manuscript Type: ORIGINAL ARTICLE

Title: Prediction of hepatocellular carcinoma development by APRI in primary biliary cholangitis

Authors: Ka-Shing Cheung¹, Wai-Kay Seto^{1,2}, James Fung^{1,2}, Lung-Yi Mak¹, Ching-Lung Lai^{1,2}, Man-Fung Yuen^{1,2}

¹Department of Medicine, The University of Hong Kong, Queen Mary Hospital, Hong Kong

²State Key Laboratory for Liver Research, The University of Hong Kong, Hong Kong

Correspondence to: Prof. Man-Fung Yuen, Department of Medicine, The University of Hong Kong

Queen Mary Hospital, Pokfulam Road, Hong Kong

E mail: mfyuen@hkucc.hku.hk Fax: 852 28162863 Tel: 852 22553984

Biostatistics statement:

All statistical analyses were performed using R version 3.2.3 (R Foundation for Statistical Computing) statistical software. Continuous variables were expressed as median and interquartile range (IQR). Spearman's bivariate correlation was used to test the correlation between continuous variables. Mann-Whitney U-test was used to compare continuous variables of two groups. Chi-square test or Fisher's exact test when appropriate, was applied for comparing categorical variables. Cox proportional hazards model was used to determine the hazard ratio (HR) HCC with different variables. Patients not meeting the clinical endpoint (HCC) were censored at latest follow-up or death. The duration of follow-up was the period of observation from the date of diagnosis to the censored date. To deal with missing data in the Cox model, multiple imputation was used to construct 50 complete datasets by imputing the missing variables.^[32] Kaplan-Meier method was used to analyze the development of HCC, and statistical significance was determined by log-rank test. The receiver operating curve was generated by plotting 'sensitivity' against '1 - specificity' at different values. The performances of various prognostic models were expressed in terms of AUROC (area under receiver operating curve), with the 95% confidence interval (95% CI) being derived from bootstrapping by sampling with replacement from the original sample and repeating the process by 1000 times. A two-sided p-value of < 0.05 was used to define statistical significance.

Name: Ka Shing Cheung

Date: 28.07.2017