**INVITED COMMENTARY**

# Non-invasive Tests of Fibrosis: Does One Size Fit All?

Michael James[1] · William Alazawi[1]

NAFLD is a major challenge at the forefront of hepatology in primary and secondary care today. Early detection of fibrosis, which correlates with risk of inferior health-related quality of life [1], development of cancer, hepatic decompensation, and death [2] is a key strategy of pathways that risk-stratify patients for management in the community versus specialist clinics [3]. Non-invasive tests and scores (collectively NITs) of fibrosis enable clinicians to assess the risk of significant liver disease in the general population or in specific at-risk groups without the need to perform liver biopsy. Calculated from 'routine' blood results, specific biomarkers, or physical characteristics of the liver, NITs are now central to care pathways for the management of liver test abnormalities and for nonalcoholic fatty liver disease (NAFLD) in particular. NITs rely on their high negative predictive value to exclude patients at low risk of advanced fibrosis (NASH Clinical Research Network (CRN) score $\geq$ F3).

Most NITs were developed and validated in populations of mostly white ethnicity [4], leading to questions about their effectiveness in other ethnic groups. Ethnicity represents shared cultural heritage as a compound surrogate that includes genetics, environmental factors, diet, social structure, privilege, and health beliefs and behaviors; all of which are highly relevant to the pathogenesis of NAFLD and fibrosis. Therefore, the assumption that all NITs are generalizable to all populations is potentially problematic.

People of South Asian ethnicity living with NAFLD are younger and have a lower body mass index (BMI) at diagnosis at comparable stages of severity than people of white ethnicity [5]. Given that almost 60% of the world's population live in Asia, understanding the performance of NITs within this vastly heterogenous population is clinically important.

In this issue of *Digestive Diseases and Sciences*, Arora et al. [6] assessed the performance of NITs alone or in combination in four Asian population cohorts (three from India and one from Singapore). Six NITs were assessed in 641 patients with different stages of fibrosis of biopsy-proven NAFLD: NAFLD Fibrosis score (NFS), Fibrosis-4 (FIB-4), aspartate aminotransferase (AST) to alanine aminotransferase (ALT) ratio (AAR), AST-to-platelet ratio (APRI), enhanced liver fibrosis (ELF), BMI/AAR/diabetes (BARD) scores, liver stiffness measurements (LSM) via transient elastography (TE) and the combination scores Agile 3+ and Fibroscan-AST (FAST). To assess optimum thresholds for detecting advanced fibrosis in the study population, area under the receiver operator curve characteristic (AUROC) for all NITs was calculated. The authors reported that although all blood-based NITs had poor diagnostic accuracy (AUROC < 0.7), high diagnostic accuracy was found in the LSM alone or in the Agile 3+ score that incorporates LSM with other markers (AUROC > 0.8). Comparing combination scores with LSM alone, LSM was superior to FAST but had similar accuracy to Agile 3+.

As expected, age, BMI > 25 kg/m$^2$, and the presence of diabetes were associated with advanced fibrosis. In patients with BMI > 25 kg/m$^2$, only LSM and Agile 3+ maintained high diagnostic accuracy. Though existing NIT thresholds yielded lower sensitivities and specificities in the study population that could be optimized by adjusting rule-in and rule-out cut-off points. FIB-4 or NFS (at the study-optimized rule-out thresholds) followed by LSM for the indeterminate and high-risk groups identified a higher percentage of correctly classified patients while reducing unclassified cases compared with LSM alone, suggesting an approach to rationalise LSM use in resource-poor settings.

The authors recognize some of the study's limitations: its retrospective nature in a biopsied population with attendant preselection bias toward at-risk groups and altered pre-test probability for advanced fibrosis. The debate around liver biopsy as a reference standard for fibrosis in NAFLD, with its inherent inter- and intra-observer variability in reporting, sampling variability and error has been well rehearsed [7],

✉ William Alazawi
   w.alazawi@qmul.ac.uk

   Michael James
   michael.james@qmul.ac.uk

1  Centre for Immunobiology, Blizard Institute, Queen Mary
   University of London, 4 Newark Street, London E1 2AT, UK

and it would be helpful to know the time interval between NIT assessment or calculation and biopsy in the study. Nevertheless, these issues are not unique being inherent to much of the NIT versus histology literature.

Performance of NITs in patients of Asian ethnicity has previously been assessed, with inconsistent findings. Work by our group [5] in London, showed ethnic disparity in accuracy in blood-based NITs, but not TE, similar to Arora et al. In contrast, a post hoc analysis of the STELLAR trial cohorts reported NIT accuracy to be comparable in white and Asian populations [8], concluding that separate cut-offs were unnecessary. Despite the benefits of a large population, prospective clinical trial design, and the use of a central pathologist, the STELLAR population was skewed toward large numbers of people with advanced disease. Reflecting the trial design and probably since investigators had, quite-rightly, heavily pre-selected patients for the studies, 76% of this cohort had advanced fibrosis stages F3 and F4 and 71% had very active NASH (NAS > 3); very different from the proportions observed in the general population.

Self-reported ethnicity may also have limitations as a phenotypic marker when assessing the performance of NITs in subgroups, given this information may not be as accurate at determining individual level risk. As stated above, ethnicity is a compound surrogate for many factors that can vary within as well as among the labels given to and chosen by patients. Therefore, to better assess efficacy of NITs, the use of more specific markers for disease phenotype such as those inferred by genetic susceptibility may add precision to identifying populations that in turn could enable a more personalized assessment of the optimum threshold for tests, rather than cut-offs based on the assumption of homogeneity in a particular group. Significant effort is being invested to develop and evaluate novel biomarkers and multi-marker scores, such as in the LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) project, using the prospective European NAFLD registry [9] and NIMBLE (Non-invasive biomarkers for metabolic liver disease) [10]. These consortia are already generating useful longitudinal data for evaluating the prognostic power of these tests. Ensuring broad representation of ethnic groups and considering ethnicity and the relevant factors associated with it will be essential in these endeavors.

As drug development in NAFLD moves forward, many are asking whether liver biopsy and histologically defined treatment eligibility or outcome measures are really appropriate given the large numbers of patients who could potentially benefit. Histologically defined stages are clinically useful as they help stratify patients according to their risk of progressing to liver cancer, decompensation, and death. If the same risk prediction can be delivered using NITs, this raises the possibility that patient selection and determining response can both be determined by the same NITs; a concept being tested in new trials and in analyses of existing data [11]. As NITs become more widespread and the basis for early detection, treatment selection and response determination, understanding accuracy in specific populations becomes even more vital.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Lazarus JV, Mark HE, Anstee Q et al. Advancing the global public health agenda for NAFLD: a consensus statement. *Nat Rev Gastroenterol Hepatol* 2022;19:60–78.
2. Sanyal AJ, Van Natta ML, Clark J et al. Prospective study of outcomes in adults with Nonalcoholic fatty liver disease. *N Engl J Med* 2021;385:1559–1569.
3. Macpherson I, Abeysekera KWM, Harris R et al. Identification of liver disease: why and how. *Frontline Gastroenterol* 2022;13:367–373.
4. Angulo P, Hui JM, Marchesini G et al. The NAFLD fibrosis score: a non-invasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* 2007;45:846–854.
5. De Silva S, Li W, Kemos P et al. Non-invasive markers of liver fibrosis in fatty liver disease are unreliable in people of South Asian descent. *Frontline Gastroenterol* 2018;9:115–121.
6. Arora U, Goyal RM, Teh KKJ et al. Poor performance of non-invasive tests for advanced fibrosis in nonalcoholic fatty liver disease: a multicentric Asian study. *Dig Dis Sci.* (Epub ahead of print). https://doi.org/10.1007/s10620-023-08085-y.
7. Ratziu V, Charlotte F, Heurtier A et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 2005;128:1898–1906.
8. Wong VWS, Tak WY, Goh GBB et al. Performance of non-invasive tests of fibrosis among Asians, Hispanic and non-Hispanic Whites in the STELLAR trials. *Clin Gastroenterol Hepatol* 2023;21:90–102.
9. Vali Y, Lee J, Boursier J et al. Biomarkers for staging fibrosis and nonalcoholic steatohepatitis in nonalcoholic fatty liver disease (the LITMUS project): a comparative diagnostic accuracy study. *Lancet Gastroenterol Hepatol.* 2023. https://doi.org/10.1016/S2468-1253(23)00017-1.
10. Sanyal A, Shankar S, Yates K et al. The Nimble stage 1 study validates diagnostic circulating biomarkers for nonalcoholic steatohepatitis. *Res Sq.* 2023. https://doi.org/10.21203/rs.3.rs-2492725/v1.
11. Stine JG, Munaganuru N, Barnard A et al. Change in MRI-PDFF and histological response in patients with nonalcoholic steatohepatitis: a systematic review and metanalysis. *Clin Gastroenterol Hepatol* 2021;19:2274–2283.