Review

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

Non-invasive diagnosis of non-alcoholic fatty liver disease. A critical appraisal

Mariana V. Machado, Helena Cortez-Pinto*

Departamento de Gastrenterologia, Hospital Santa Maria, CHLN, Unidade de Nutrição e Metabolismo, Faculdade de Medicina de Lisboa, IMM, Lisbon, Portugal

Summary

Non-alcoholic fatty liver disease (NAFLD) affects one in every three subjects in the occidental world. The vast majority will not progress, but a relevant minority will develop liver cirrhosis and its complications. The classical gold standard for diagnosing and staging NAFLD and assessing fibrosis is liver biopsy (LB). However, it has important sample error issues and subjectivity in the interpretation, apart from a small but real risk of complications. The decision to perform an LB is even harder in a condition so prevalent such as NAFLD, in which the probability of finding severe liver injury is low. In an attempt to overcome LB and to subcategorize patients with NAFLD in different prognoses allowing better management decisions, several non-invasive methods have been studied in the last decade. The literature is vast and confusing. This review will summarize which methods have been tested and how they perform, which tests are adequate for clinical practice and how they can change the management of these patients.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is the hepatic pandemic of the XXI century, being the number one cause of chronic hepatic disease in the occidental world [1]. Although usually benign, fatty liver may associate with serious injury, with inflammation and hepatocyte necro-apoptosis, non-alcoholic steatohepatitis (NASH), in 20–30% of subjects [2]. Those patients are at risk of developing fibrosis, one fifth progressing to liver cirrhosis [2]. It is apparently more slowly progressive than other chronic liver diseases, such as alcohol or viral-induced disease [3]. However, because NAFLD is so common, occurring in one out of three persons in the developed world [1], it is the third cause of liver transplantation in the United States [4]. Moreover, the problem of hepatocytes being fatty, overcomes the liver itself, as it increases the risk for cardiovascular disease and death and duplicates the risk for type 2 diabetes mellitus (T2DM), independently of the severity of liver injury [5].

The gold standard for the diagnosis and staging of NAFLD is liver biopsy (LB), although as it will be discussed later, it may have been dethroned by more accurate methods in what concerns steatosis. However, it is the only way to directly diagnose NASH and fibrosis, even if several assays and models try to predict it with reasonably good accuracy. LB has several drawbacks. It is an invasive procedure, frequently associated with distress and discomfort. Although generally safe, it comes with a risk for major complications in 1–3% and even death in 0.01% [6]. The second issue relates to sampling problems, which results in misdiagnosis in a very significant number of cases. In fact, NASH may be wrongly excluded in up to one fourth of the cases and fibrosis severity misclassified in up to one third of the patients [7]. That propensity for sampling error relates to the procedure and to the disease. Even an adequate LB will show only 0.05 cm^3 from an organ whose volume ranges between 800 and 1000 cm^3, corresponding to less than 1:50,000 of the total volume [6]. Also, in NAFLD, lesions are not evenly distributed [7]. Lastly, diagnosis is dependent on the subjectivity and experience of the pathologist, mostly in identifying ballooning and grading necro-inflammation.

Several non-invasive methods aim at diagnosing and quantifying hepatic steatosis, while others were designed to predict NASH or significant/advanced fibrosis. In this review, the rationale for pursuing each diagnosis and instruments available will be discussed. The reliability and importance of diagnostic tests depend on the disease, the population where it is applied and the change in management induced by the test's result. A good screening test should have a high sensitivity (Se) even at expense of specificity (Sp), whereas a diagnostic test that selects patients for invasive procedures, therapy or clinical trials should have

Keywords: Non-alcoholic fatty liver disease; Non-alcoholic steatohepatitis; Fibrosis; Non-invasive diagnosis.
Received 31 August 2012; received in revised form 4 November 2012; accepted 7 November 2012
* Corresponding author. Address: Serviço de Gastrenterologia, Hospital de Santa Maria, Av. Prof. Egas Moniz, 1649-035 Lisboa, Portugal. Tel.: +351 217985187; fax: +351 217985142.
E-mail address: hlcortezpinto@netcabo.pt (H. Cortez-Pinto).
Abbreviations: NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; T2DM, type 2 diabetes mellitus; LB, liver biopsy; Se, sensitivity; Sp, specificity; ROC, receiver operating characteristic; AUROC, area under ROC; NPV, negative predictive value; PPV, positive predictive value; FLI, fatty liver index; US, ultrasonography; BMI, body mass index; GGT, γ-glutamyltranspeptidase; LAP, lipid accumulation product; MRS, magnetic resonance spectroscopy; AST, aspartate aminotransferase; ALT, alanine aminotransferase; US-FLI, ultrasonographic fatty liver indicator; CT, computed tomography; CAP, controlled attenuation parameter; BMI, body mass index; CK18, cytokeratin 18; CBT, C-caffeine breath test; HA, hyaluronic acid; TIMP-1, tissue inhibitor of metalloproteinases-1; ARFI, acoustic radiation force impulse.

high Sp. The most common approach to evaluate a test has been the analysis of the Receiving Operating Characteristic (ROC) curves and the area under ROC (AUROC), which evaluates the probability of a test discriminating a true positive (Se) against the probability of finding a false positive result (1-Sp). When the AUROC is higher than 0.8, it suggests good accuracy. It is a valuable tool but must be analysed carefully, particularly when comparing different tests. Although Se and Sp are invariant for a diagnostic test, they may depend on the characteristics of the population, such as age, gender and severity of disease. Thus, it may not be accurate to compare AUROC of different tests in different studies, with different populations and no statistical work done comparing them.

Identification and quantification of hepatic steatosis

The first challenge is when to suspect NAFLD. Suspicion will not be driven by clinical manifestations, since most patients are asymptomatic. Symptomatic patients present unspecific complaints such as fatigue, abdominal discomfort and, only seldom, manifestations of advanced liver disease. There are, however, high-risk populations in whom the prevalence is so high that *per se* is enough to raise the hypothesis of NAFLD. In fact, up to two thirds of patients with obesity and T2DM, present with hepatic steatosis [8]. Also, liver tests, namely aminotransferases, are usually normal, and when increased, typically present mild elevation with a fluctuant pattern [9]. Isolated increase in alkaline phosphatase is not frequent, but it has been reported in 10% of patients referred to tertiary care [10].

We should ask whether it is worth searching and diagnosing NAFLD in asymptomatic subjects with normal liver tests, since the majority will have non-progressive simple steatosis. Then again, hepatic steatosis is linked to an increase in cardiovascular risk and death. Particularly in diabetic patients, steatosis increases by more than 3-fold the risk for overall death [11] and cardiovascular disease [12]. Even in these high-risk patients, there is controversy regarding screening, among endocrinologists. Some do not recommend it, advocating that traditional scores should assess cardiovascular risk, while others consider diagnosis and evaluation of NAFLD as part of the management of DM being an indication for more intense monitoring and therapeutic intervention [13]. The guidelines on NAFLD from the American Association for the Study of Liver Diseases discourage screening for hepatic steatosis, even in high-risk patients such as diabetics. There is not enough evidence regarding diagnostic tools and treatment options, and there are no studies on the cost-effectiveness of a screening program [14]. There are however, situations in which an active search for NAFLD in asymptomatic subjects is warranted: liver donors for liver transplant, as steatosis is a risk for graft primary non-function, and in major hepatic resection [15], in which steatosis increases the risk for post-operative morbi-mortality [16].

The second question, if it is meaningful to quantify steatosis, is difficult to answer, since it has not been consistently demonstrated that the amount of fat influences prognosis. Also, it is yet to be elucidated if decreasing the amount of hepatic fat with therapeutic interventions will favourably affect the cardio-metabolic risk and the risk for progression to advanced liver disease.

Several diagnostic panels have been proposed to predict steatosis (Table 1). Steatotest incorporates 12 variables in an undisclosed formula, including α2-macroglobulin, haptoglobin, and apolipoprotein A1 [17]. In a French cohort of more than 700 patients, it showed reasonable accuracy, with 0.79 AUROC for moderate-severe steatosis, good negative predictive value (NPV), 93%, but small positive predictive value (PPV), 63%. Another French group found similar results in 288 morbid obese subjects [18]. The Poynard' group conducted a meta-analysis in morbid obese subjects, obtaining the same conclusions [19]. We have to acknowledge that AUROC was suboptimal, it has been validated only in French cohorts, it incorporates tests not used routinely and because the formula is not public, a fee is imposed for each test applied.

Bedogni *et al.* first proposed Fatty Liver Index (FLI) in 2006, as an algorithm derived from the population of the Dionysos Nutrition & Liver Study [20]. The gold standard was ultrasonography (US). The index varies between 0 and 100 and is calculated through a formula incorporating: body mass index (BMI), waist circumference, triglycerides and γ-glutamyltranspeptidase (GGT). It showed good accuracy in detecting NAFLD and it has been used in several population studies [21,22]. However, the gold standard used is far from ideal, and as such, the results should be interpreted carefully. Its main indication is for epidemiological studies, in an attempt to avoid US. Recently, a study on 2075 middle-aged Caucasians from the Regional Health Registry, followed for 15 years, showed that FLI independently associates with overall, cardiovascular and cancer-related mortality [23].

The same group also proposed Lipid Accumulation Product (LAP) that incorporates gender, waist circumference and triglycerides. After log-transformation, for each log unit increment, the risk for steatosis increased more than 4 folds [24]. Although this is a very simple test to apply, it needs validation by independent groups.

Recently, NAFLD Liver Fat Score [25] derived from a Finnish population. Gold standard was magnetic resonance spectroscopy (MRS). The score incorporates simple variables: presence of the metabolic syndrome and T2DM, fasting serum insulin, aspartate aminotransferase (AST) and AST/alanine aminotransferase (ALT) ratio. It yielded 95% Se and Sp. Information on *PNPLA3* gene (rs738409) improved the accuracy for the prediction in less than 1%. A Netherlands' group confirmed these results [26]. It may be a test to take into account when assessing steatosis easily on the bench without recurring to radiology.

The best non-invasive tests for the diagnosis of steatosis are the imaging ones. US should be the first method to be used in a clinical setting. It is inexpensive, widely available and it has 60–94% Se and 66–97% Sp for hepatic steatosis [27–29]. However, US acuity decreases dramatically for mild steatosis. In a study on 100 living donors for liver transplant, US could not detect steatosis when present in less than 10% of hepatocytes, and detected only 55% and 72% of patients, with steatosis 10–19% and 20–29%, respectively [30]. As it is a subjective evaluation, several attempts have tried to make it quantitative. A hepato-renal index contrast above 7.0 dB presented 91% Se and 84% Sp for hepatic steatosis [31]. A semi-quantitative score has been proposed, the Ultrasonographic Fatty Liver Indicator (US-FLI) [32]. It requires the presence of liver/kidney contrast (brighter liver than kidney) among other parameters. A score of at least 2 is highly indicative of NAFLD. US has several limitations: it is unreliable in the detection of mild steatosis, it has only up to 67% PPV [33], it is operator dependent with low inter and intra-observer agreement for steatosis, around 70% and 50% for steatosis presence and severity,