University of Modena and Reggio Emilia

Department of Biomedical and Metabolic Sciences and Neurosciences

Clinical and Experimental Medicine PhD program – XXXIII cycle

Coordinator Prof Giuseppe Biagini

Imaging biomarkers of body composition including fatty liver, sarcopenia, and fat distribution: diagnostic and prognostic role in different metabolic, infectious, and oncologic diseases

Thesis submitted for the degree of Doctor in Clinical and Experimental Medicine by

Phd Candidate: Dr Giulia Besutti

Supervisor: Dr Paolo Giorgi Rossi

Co-supervisor: Prof Guido Ligabue

Table of Contents

1. ABBREVIATION LIST

A Fib, atrial fibrillation and flutter

AGREE, Appraisal of Guidelines for Research and Evaluation

ALT, alanine aminotransferase

ARFI, acoustic radiation force impulse

ART, antiretroviral therapy

AST, aspartate aminotransferase

AUROC, Area Under the Receiver Operating Characteristic curve

BMI, Body Mass Index

CAP, Controlled Attenuation Parameter

CEUS, contrast-enhanced ultrasound

CI, Confidence Interval

CKD, chronic kidney disease

CMP, cardiomyopathies

COPD, chronic obstructive pulmonary disease

COVID-19, coronavirus disease 2019

CRC, colorectal cancer

CRP, C-reactive protein

CT, Computed Tomography

DALY, disease-adjusted life year

DLBCL, Diffuse Large B-Cell Lymphoma

DWI, diffusion-weighted imaging

DXA, dual X-ray-absorptiometry

EACS, European AIDS Clinical Society

EASL, European Association for the Study of the Liver

EAT, Epicardial Adipose Tissue

ECOG, Eastern Cooperative Oncology Group

ELF, Enhanced Liver Fibrosis score

ER, emergency room

EXP, exposed patients

FIB-4, fibrosis-4 score

FLI, Fatty Liver Index

FLIP, Fatty Liver Inhibition of Progression

GGT, gammaglutamil transpeptidase

GRADE, Grading of Recommendations, Assessment, Development and Evaluation

HCC, hepatocellular carcinoma

HDL, high-density lipoprotein

HIV, Human Immunodeficiency Virus

HR, Hazard Ratio

HS, Hepatic Steatosis

HTN HD, hypertensive heart disease

HU, Hounsfield Unit

IDU, injection drug use

IHD, ischemic heart disease

ILD, interstitial lung diseases

IMAT, Intermuscolar Adipose Tissue

IPI, international prognostic index

IU, international units

IVIM, intravoxel incoherent motion

L/S, liver to spleen

LARC, locally advanced rectal cancer

LB, Liver Biopsy

LDH, lactate dehydrogenase

LDL, low-density lipoprotein

LEs, Liver Enzymes

LHivPa, Liver Pathologies in HIV in Palermo

LIVEHIV, Liver Disease in HIV

LM, liver metastases

MET, metachronous liver metastases

MetS, metabolic syndrome

MHMC, Modena HIV Metabolic Clinic

MR, Magnetic Resonance

MRE, Magnetic Resonance Elastography

MRI, Magnetic Resonance Imaging

MRS, Magnetic Resonance Spectroscopy

MSK, musculoskeletal disorders

MSM, men having sex with men

NAFLD, Non-Alcoholic Fatty Liver Disease

NAS, NAFLD Activity Score

NASH, Non-Alcoholic Steatohepatitis

NFS, NAFLD Fibrosis Score

NNRTIs, non-nucleoside reverse transcriptase inhibitors

non-EXP, non-exposed patients

NOS, Newcastle Ottawa Scale

NPC, nasopharyngeal cancer

NRTIs, nucleoside reverse transcriptase inhibitors

OR, Odds Ratio

OS, Overall Survival

PDFF, Proton Density Fat Fraction

PET, Positron Emission Tomography

PFS, Progression Free Survival

PIs, Protease Inhibitors

PPV, Positive Predictive Value

PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PSWE, Point Shear Wave Elastography

PT, proximal thigh

RHD, rheumatic heart disease

ROB, risk of bias

RR, relative risk

SAT, Subcutaneous Adipose Tissue

SMD, Skeletal Muscle Density

SMI, Skeletal Muscle Index

SO2, oxygen saturation level

SYN, synchronous liver metastases

T2D, Type 2 diabetes

TAT, Total Adipose Tissue

TB, tuberculosis

US, ultrasound

US-FLI, ultrasound-Fatty Liver Indicator

VAT, Visceral Adipose Tissue

VCTE, Vibration-controlled Transient Elastography

2. ABSTRACT

Rationale: Imaging tests may be used to obtain different biomarkers describing body composition.

Imaging in Non-Alcoholic Fatty Liver Disease (NAFLD) assessment

To understand which role imaging biomarkers may have in NAFLD assessment, firstly a systematic review was conducted to appraise NAFLD guidelines, focusing on screening approaches in high-risk patients. Of the 14 guidelines included, British and North America guidelines received the highest scores.

A second systematic review with the aim of evaluating accuracy of imaging vs biopsy in diagnosing steatohepatitis (NASH) among NAFLD patients included 58 studies and several imaging techniques, with US and MR elastography and non-elastographic techniques having the most promising results (AUROC 0.80-1), but results derived mostly from single studies without independent validation.

A NAFLD assessment algorithm derived from guidelines was applied to three prospective cohorts of HIV patients (total 1534 patients), 313 (20.4%) with diabetes or obesity, thus requiring NAFLD screening. Among these 313 patients, 123 (39.3%) were referred to the hepatologist. When extending the algorithm also to patients with other metabolic comorbidities (total 1062 patients), 341 (32.1%) would require hepatologist referral.

By applying a similar algorithm, 171 consecutive Type 2 diabetes patients at their first diagnosis were prospectively stratified based on liver function tests and steatosis/fibrosis scores, generating 115 (67.3%) referrals to a hepatologist and 30 (17.5%) referrals for liver biopsy. Of 14 biopsies performed, 12 resulted in steatohepatitis, 1 with significant fibrosis.

In an ongoing prospective study (22 patients included by now) we aim to evaluate the accuracy of different US and MR techniques in the diagnosis of NASH and fibrosis, using liver biopsy as the reference standard, in high-risk NAFLD patients.

Fatty Liver in oncologic patients

In a metanalysis including 19 studies analyzing the association between diffuse liver diseases (steatosis, fibrosis, cirrhosis, viral infections) and the occurrence of liver metastases (LM) in patients with solid tumors, we observed that patients with diffuse liver diseases had lower risk of synchronous LM (RR 0.50 95%CI 0.34-0.76), and a slightly higher risk of metachronous LM (HR 1.11 95%CI 1.03-1.19), with considerable overall heterogeneity.

In a study including 283 stage II-IV rectal cancer patients, 90 (31,8%) patients had baseline CT-defined liver steatosis. The prevalence of synchronous LM was higher in patients with steatosis (19% vs 13%), while the incidence of metachronous LM was similar, without any significant association. In a small analysis of 63 patients without baseline steatosis and treated with neoadjuvant chemotherapy, chemotherapy-induced liver damage (defined based on CT and liver function tests) was associated with higher incidence of LM and worst survival.

Sarcopenia and ectopic fat

-Diffuse large B-cell lymphoma (DLBCL)

In a retrospective study of 116 consecutive DLBCL patients, the skeletal muscle index (SMI) and density (SMD) and the intermuscular adipose tissue area (IMAT) at L3 vertebra and proximal thigh (PT) were obtained from baseline CT scan. Low muscle quality (L3-SMD), but not low muscle quantity (L3-SMI), was associated with early therapy termination, shorter overall and progression-free survivals (OS, PFS). In multivariable models, L3-SMD remained associated with OS, and increasing PT-IMAT was a poor prognostic factor for OS and PFS.

-COVID-19

In a study including 318 consecutive COVID-19 patients who performed chest CT scan at emergency room, pectoralis muscle density, total, visceral, and intermuscular adipose tissue areas (TAT, VAT, and IMAT) at T7-T8 vertebrae were retrospectively measured. In multivariate models, decreased muscle density, increased TAT, VAT, and IMAT were risk factors for hospitalization and mechanical ventilation or death. Part of the effect of age on death was mediated by sarcopenia.

ABSTRACT- Italian

Razionale: gli esami di imaging possono essere utilizzati per ottenere diversi biomarcatori di composizione corporea.

L’imaging nella valutazione della steatosi epatica non alcolica (NAFLD)

Per capire quale ruolo possono avere i biomarcatori di imaging nella valutazione della NAFLD, è stata innanzitutto condotta una revisione sistematica per valutare le linee guida esistenti, concentrandosi sugli approcci di screening nei pazienti ad alto rischio. Delle 14 linee guida incluse, quelle britanniche e quelle del Nord America hanno ricevuto i punteggi più alti.

Una seconda revisione sistematica, con l'obiettivo di valutare l'accuratezza delle metodiche di imaging rispetto alla biopsia nella diagnosi di steatoepatite (NASH) in pazienti con NAFLD, ha incluso 58 studi e diverse tecniche di imaging, tra cui le più promettenti sono risultate l'elastografia (US e RM) e le tecniche RM non elastografiche (AUROC 0.80-1). Questi risultati derivano però per lo più da studi singoli senza validazione indipendente.

Un algoritmo di valutazione della NAFLD derivato dalle maggiori linee guida è stato applicato a tre coorti prospettiche di pazienti HIV (totale 1534 pazienti), 313 (20,4%) con diabete o obesità. Tra questi 313 pazienti, 123 (39,3%) sono stati indirizzati all'epatologo. Se si estendesse l'algoritmo anche a pazienti con altre comorbidità metaboliche (totale 1062 pazienti), 341 (32,1%) pazienti richiederebbero l'invio ad un epatologo.

Applicando un algoritmo simile, 171 pazienti consecutivi affetti da diabete di tipo 2 alla prima diagnosi sono stati stratificati prospetticamente sulla base degli indici di funzionalità epatica e di scores di steatosi / fibrosi, generando 115 (67.3%) invii all'epatologo e 30 (17.5%) invii a biopsia. In 12 delle 14 biopsie eseguite è stata diagnosticata steatoepatite, fibrosi significativa solo in 1/14.

In uno studio prospettico in corso (22 pazienti inclusi ad oggi) ci proponiamo di valutare l'accuratezza di diverse tecniche US e MR nella diagnosi di NASH e fibrosi, utilizzando la biopsia epatica come standard di riferimento, in pazienti NAFLD ad alto rischio.

Steatosi epatica nei pazienti oncologici

In una metanalisi comprendente 19 studi che analizzano l'associazione tra malattie epatiche diffuse (steatosi, fibrosi, cirrosi, infezioni virali) e l'insorgenza di metastasi epatiche (LM) in pazienti con tumori solidi, abbiamo osservato che i pazienti con malattie epatiche diffuse hanno un minor rischio di LM sincrone (RR 0,50 95% CI 0,34-0,76) e un rischio leggermente più elevato di LM metacrone (HR 1,11 95% CI 1,03-1,19), con una considerevole eterogeneità complessiva tra gli studi inclusi.

In uno studio che includeva 283 pazienti con tumore del retto in stadio II-IV, 90 (31,8%) pazienti avevano una steatosi epatica al basale definita alla CT. La prevalenza di LM sincrone era lievemente maggiore nei pazienti con steatosi (19% vs 13%), mentre l'incidenza di LM metacrone era simile. In una piccola analisi di 63 pazienti senza steatosi al basale e trattati con chemioterapia neoadiuvante, il danno epatico indotto dalla chemioterapia (definito sulla base di CT e test di funzionalità epatica) è risultato essere associato a una maggiore incidenza di LM e ad una peggiore sopravvivenza.

Sarcopenia e grasso ectopico

-Linfoma diffuso a grandi cellule B (DLBCL)

In uno studio retrospettivo su 116 pazienti consecutivi con DLBCL, l'indice di massa muscolare scheletrica (SMI), la densità del muscolo scheletrico (SMD) e l'area di tessuto adiposo intermuscolare (IMAT) a livello della vertebra L3 e della coscia prossimale (PT) sono stati ottenuti dalla scansione CT eseguita alla prima stadiazione. La bassa qualità muscolare (L3-SMD), ma non la bassa quantità muscolare (L3-SMI), è risultata essere associata all'interruzione precoce della terapia e a minor sopravvivenza globale e priva di progressione (OS, PFS). Nei modelli multivariati, L3-SMD è rimasto associato a OS e l'aumento di PT-IMAT era un fattore prognostico sfavorevole per OS e PFS.

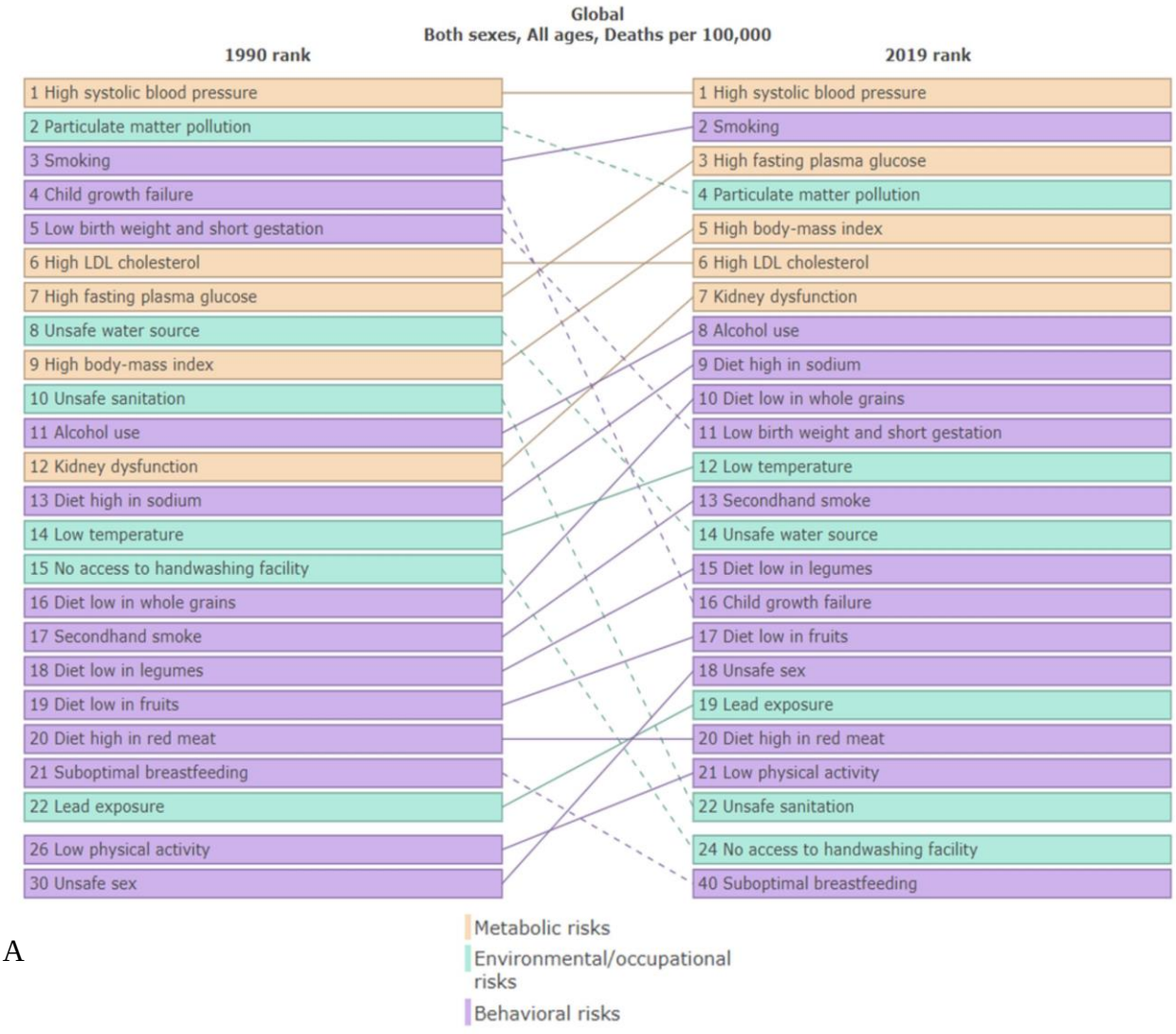
-<u>COVID-19</u>

In uno studio su 318 pazienti COVID-19 consecutivi che hanno eseguito la TC del torace al pronto soccorso, la densità del muscolo pettorale, le aree di tessuto adiposo totale, viscerale e intermuscolare (TAT, VAT e IMAT) sono state misurate retrospettivamente al livello delle vertebre T7-T8. Nei modelli multivariati, la diminuzione della densità muscolare, l'aumento di TAT, VAT e IMAT sono risultati essere fattori di rischio per il ricovero e la ventilazione meccanica o la morte. Parte dell'effetto dell'età sulla morte era mediato dalla sarcopenia.

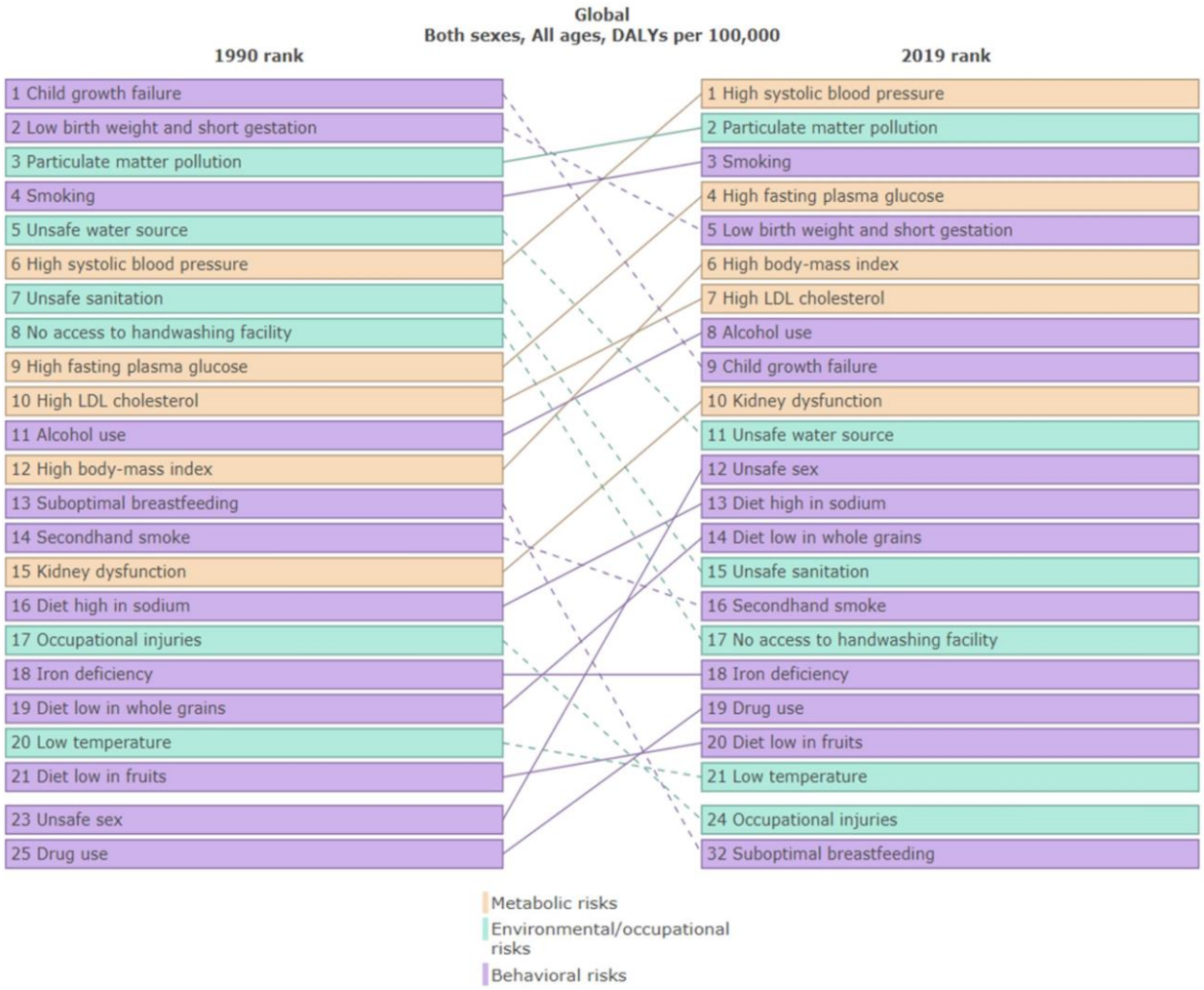
3. INTRODUCTION

3.1. Burden of disease induced by metabolic risk factors

By looking at the impact of different risk factors (including metabolic, environmental/occupational, and behavioral risk factors) on deaths and disability-adjusted life years (DALYs), it is apparent that the rate of events linked to metabolic risk factors has increased in the last thirty years worldwide (Figure 1). When restricting this evaluation to middle-to-high socio-demographic index, or to Italy, this change in time is less evident, since in 1990 metabolic risk factors had already risen among the first determinants of death and morbidity [1].



A



B

Figure 1: Worldwide changes of the impact of different risk factors on deaths (A) and disability-adjusted life years (DALYs) (B), from 1990 to 2019 [1].

However, by evaluating the effect on DALYs of different metabolic risk factors considered separately, also in Italy a slight increase is visible for high fasting plasma glucose, high body mass index, and low bone mineral density (Figure 2).

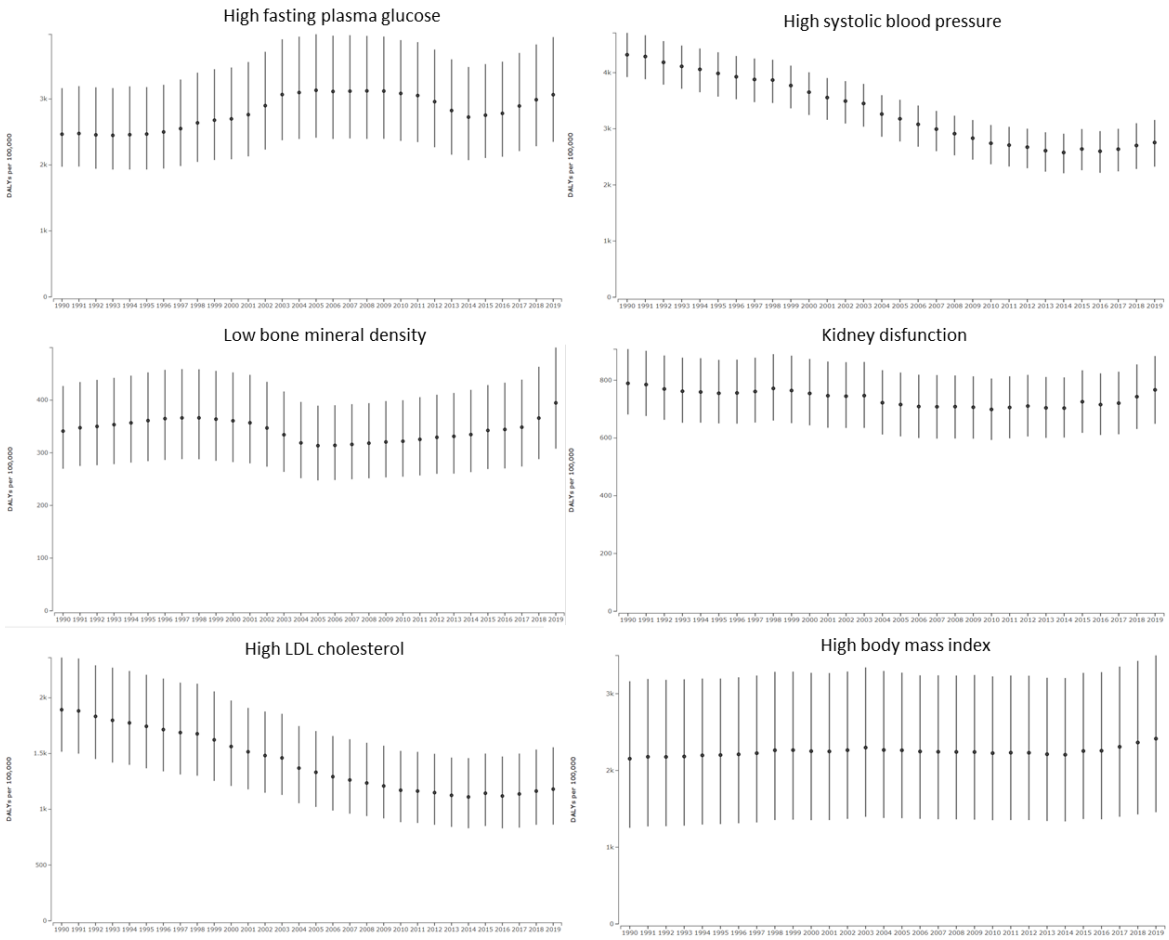
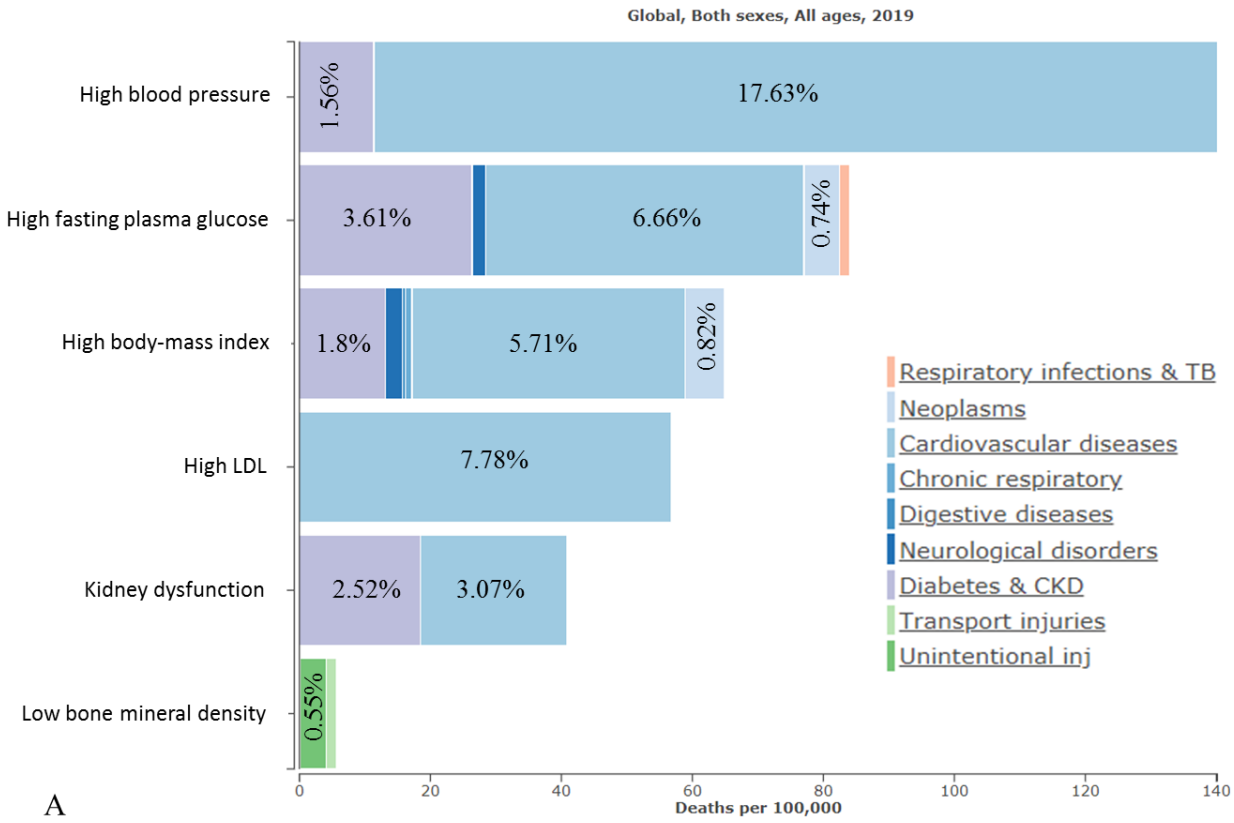
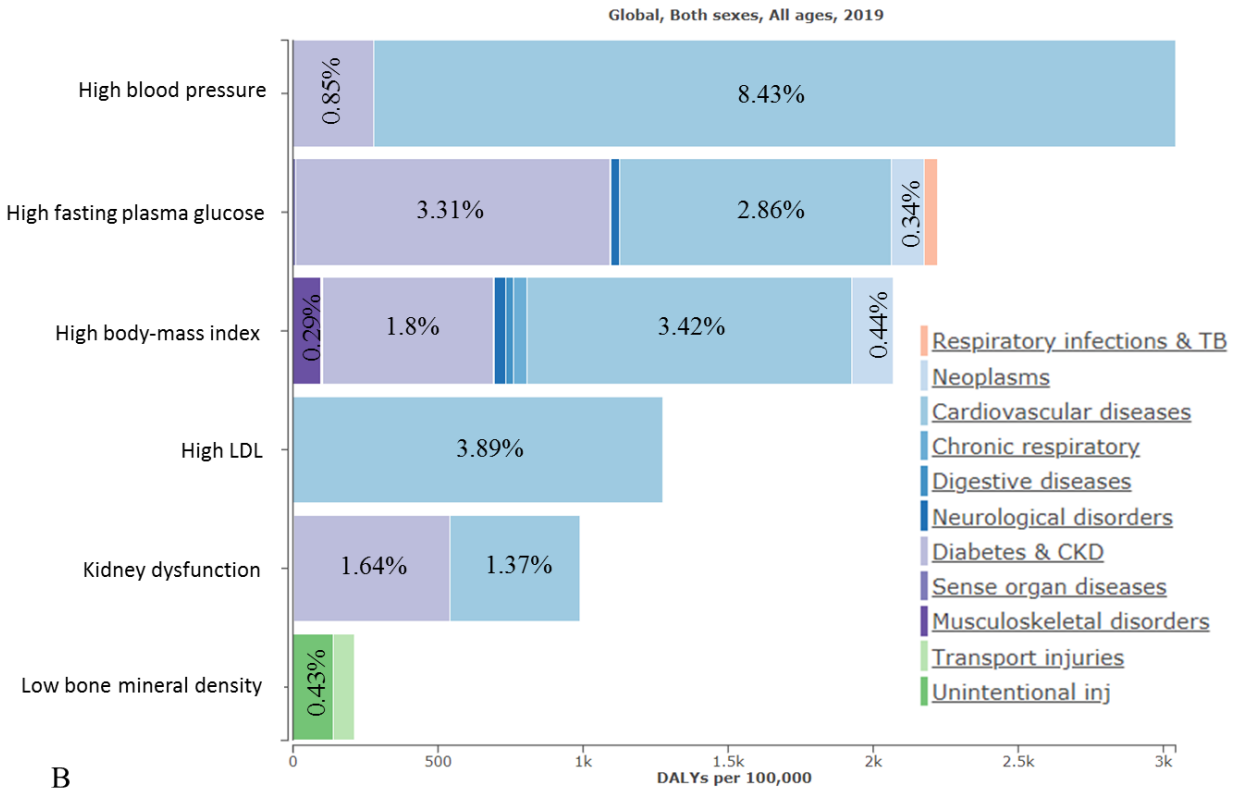


Figure 2: Rate of disability-adjusted life years (DALYs) from 1990 to 2019 in Italy. Each graph represents the time course of the rate attributable to one of six metabolic risk factors. [1]

The proportion of deaths which is attributable to metabolic risk factors is especially high when considering deaths due to cardiovascular and cerebrovascular diseases, diabetes, and chronic kidney disease [1]. When evaluating each single metabolic risk factor (Figure 3), high blood pressure alone is responsible for almost 20% of total death and 10% of total DALYs, acting mostly through cardiovascular diseases, and in lesser part through diabetes and chronic kidney diseases. High fasting plasma glucose and high body mass index (BMI) are responsible for approximately 11% and 8.5% of total deaths, respectively, and each of them for approximately 6% of total DALYs. Again, causative diseases for these two risk factors are cardiovascular diseases, diabetes and chronic kidney diseases, but a smaller role is also played by neoplasms, neurological disorders, respiratory infection and other disorders (respiratory, digestive, and musculoskeletal). Increased low-density lipoprotein (LDL) cholesterol is another important metabolic risk factor, that, acting through cardiovascular diseases, is responsible for almost 8% of total deaths and almost 4% of total DALYs.



A



B

Figure 3: Graphs representing death (A) and disability-adjusted life years (DALYs) (B) rates attributable to each metabolic risk factor, further divided by different causative diseases [1].

Reported percentages are calculated on total deaths (A), and total DALYs (B) worldwide in year 2019. TB, tuberculosis; CKD, chronic kidney diseases; LDL, low-density lipoprotein.

3.2. Metabolic - behavioral risk factors and body composition

Body composition is a general term used to indicate the percentages in which different tissues are represented in a person's body. The simpler and most used characterization of body composition is limited to the percentages of fat mass and fat-free mass or lean mass [2]. However, especially when focusing on its relationship with cardiometabolic risk, a more detailed description of body composition is necessary. Indeed, different fat depots in the body are known to exert different effects in terms of systemic inflammation and cardiovascular risk [3-4], and also lean mass should be better characterized, because not only quantity but also quality of the muscle may play an important role [5].

The aforementioned high burden of disease induced by metabolic risk factors underlines the importance of body composition, since these factors are mostly biomarkers of or at least closely linked to body composition. For example, visceral obesity may favor insulin resistance, increase of fasting plasma glucose and LDL cholesterol [6], and high BMI is obviously a measure of obesity, even if limited in discriminating the fat mass from the fat-free mass and between different fat compartments [7].

Moreover, body composition is also strongly influenced by behavioral risk factors including dietary factors and low physical activity [8] (Figure 4), which are responsible for another significant proportion of deaths and DALYs [1]. In Italy, dietary risk accounts for almost 14% of deaths and 8% of DALYs, caused by cardiovascular diseases and, to a lesser degree, diabetes, renal diseases, and neoplasms, while low physical activity is responsible for approximately 3% and 1.5% of deaths and DALYs, respectively (Figure 5) [1].

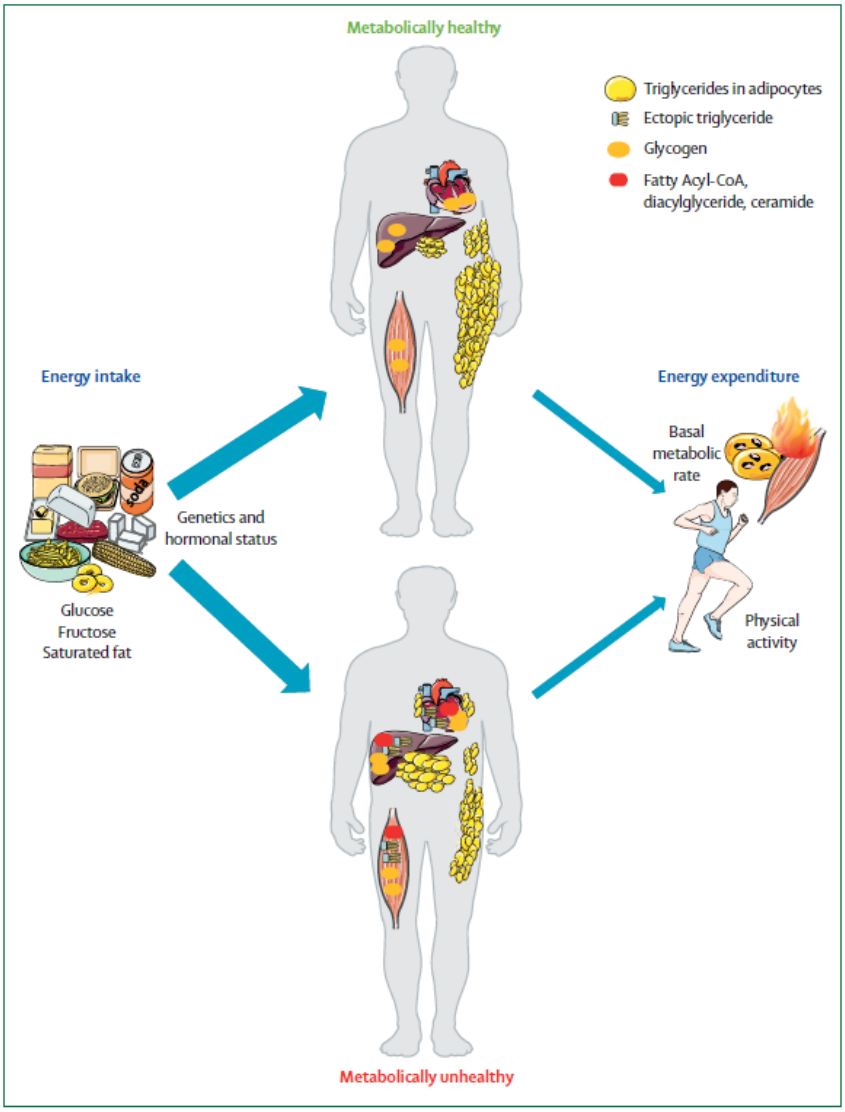
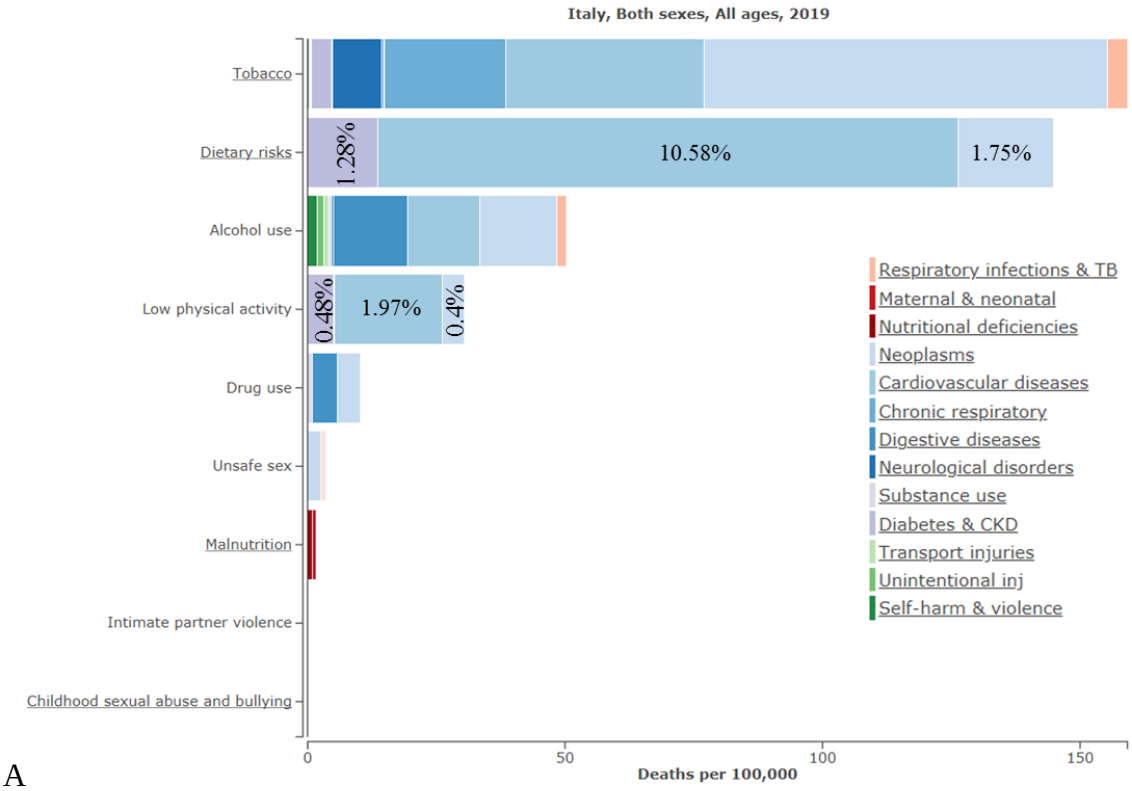
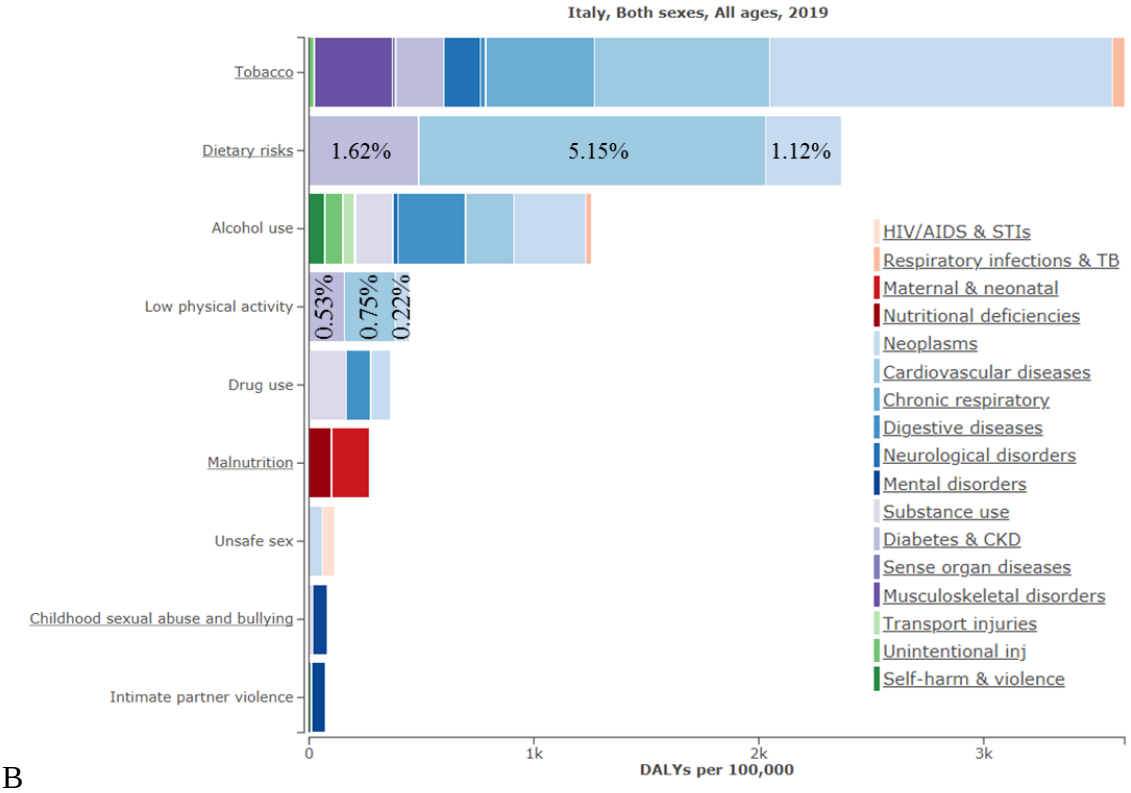


Figure 4: Diet and physical activity effect on body composition and metabolic health. Increased energy intake deriving from sugar and saturated fat, combined with low energy expenditure, results in triglycerides storage in adipose tissue and ectopic fat depots. The metabolically unhealthy phenotype is characterized by fat storage in visceral area and in ectopic sites including muscle and liver and is linked to increased subclinical inflammation and insulin resistance [8].



A



B

Figure 5: Graphs representing death (A) and disability-adjusted life years (DALYs) (B) rates attributable to each behavioral risk factor, further divided by different causative diseases [1]. Reported percentages are calculated on total deaths (A), and total DALYs (B) in Italy in year 2019. TB, tuberculosis; CKD, chronic kidney diseases; LDL, low-density lipoprotein.

3.3. Imaging techniques for body composition assessment

The high and increasing prevalence of risk factors and disease burden associated with body composition, and the limited value of traditional anthropometric measures such as BMI and waist circumference in describing the complex spectrum of body composition phenotypes, make the research on body composition biomarkers a very active research area.

In this regard, imaging techniques may provide both anatomical and functional information on body components, allowing to study the distribution of body fat in different compartments, including the increased triglycerides storage in hepatocytes (fatty liver), and the quantity as well as quality of lean mass, providing biomarkers of sarcopenia.

Imaging methods that can be used for the global assessment body composition are dual X-ray-absorptiometry (DXA), whole body magnetic resonance imaging (MRI) or whole body computed tomography (CT) [9]. To minimize radiation dose and/or acquisition time, single abdominal slices may be acquired with CT or MRI to measure tissue areas and density or signal intensity, which can be used to estimate whole-body composition (e.g. abdominal visceral and subcutaneous adipose tissue, and abdominal skeletal muscle at L3 level) [10-11]. Finally, Ultrasound (US) may have a role in evaluating the thickness of specific body components, as abdominal fat compartments or thigh skeletal muscle, or in the evaluation of fatty liver [12].

3.3.1. Distribution of fat in different body compartments

The distribution pattern of fat is closely linked to metabolic profile and cardiovascular risk. Ectopic fat, including visceral, hepatic, cardiac (intramyocardial, pericardial, and epicardial), and muscular fat, is more associated with systemic inflammation and metabolic derangement [13], with increased cardiovascular risk also favored by a paracrine effect on atherosclerotic plaque [14].

Both CT and MRI may differentiate between subcutaneous adipose tissue (SAT), intermuscular adipose tissue (IMAT), and visceral adipose tissue (VAT) located in the abdominal cavity, moreover they can be used to quantify epicardial adipose tissue (EAT) and fatty liver. Of course, their use in clinical practice is limited by costs and radiation exposure (for CT), they are not free of contraindications and usually unsuitable for heavily overweight patients. In the research setting, CT has been frequently used to assess body fat distribution, and body composition in general, in patients with cancer or other clinical conditions which require CT scan execution, so that images are available to measure body composition parameters without creating ethical concerns [15].

To evaluate abdominal fat compartment, a single slice is commonly used, and different slice levels have been tested [10-11]. The best correlation of adipose tissue area with total body adipose tissue volume measured with MRI was found for a slice located 5 cm below the L4/L5 level [11], however, the corresponding to 5 cm above the transition L4/L5, approximately at L3 level, is frequently used as landmark to obtain comprehensive body composition measures since it resulted in more accurate estimates of total body skeletal muscle volume [2]. To measure fat compartments on CT images, predetermined Hounsfield Unit (HU) thresholds are used to select fat (from -190 to -30 HU) [15-16], while MRI contrast resolution allows the identification of different compartments based on signal intensity difference between tissues. Then, VAT, IMAT, and SAT may be segmented manually or through specific software to obtain cross-sectional areas. By using predefined formulas accounting for segmented tissue areas, slice thickness, interval, and number, compartment volumes may be obtained too [2]. Furthermore, MRI through chemical shift techniques may also provide a quantification of lipid content in different tissue and organs [17]. Finally, cardiac or thoracic CT scan, by using post-processing techniques similar to those described for abdominal fat compartments, may also be used to obtain volumes of EAT or pericoronary fat, which have been studied for their supposed paracrine activity favoring ischemic events [18].

The DXA scan differentiates between fat mass, non-bone lean mass and bone mineral content. These measures may be obtained at a whole-body level, or regionally (trunk, arms, legs, supraumbilical abdomen, gluteo-femoral region). Moreover, in the last years new software has emerged allowing DXA estimation of VAT, based on SAT mapping and subtraction of SAT from total fat mass of the supraumbilical abdomen (android fat mass) [19]. Major advantages of DXA when compared to MRI and CT are low costs, low radiation exposure, wide availability and short scan time. The main limitation is the confounding effect of different hydration states of soft tissues. To create a significant variation in estimates, tissue overhydration should be severe and not commonly found in clinical practice, however regional fat estimates may be affected by large local water accumulations as in case of subcutaneous edema or ascites [20]. Moreover, DXA accuracy may change according to thickness and size of the patient's body, technical aspects such as machine calibration or software version, and regions of interest definition [2].

Low costs, large availability and harmlessness are main advantages of US, too. In the abdomen, the combined use of a convex probe and a linear probe allows the measurements of different fat compartment thicknesses, including the most used intraabdominal fat thickness and maximum subcutaneous fat thickness (Figure 6), but also mesenteric fat thickness, pre-peritoneal fat thickness, peri and pararenal fat thickness [12]. Finally, cardiac US may be used to measure the thickness of

EAT, which should be assessed on the free wall of the right ventricle in long and short axis parasternal views [21].

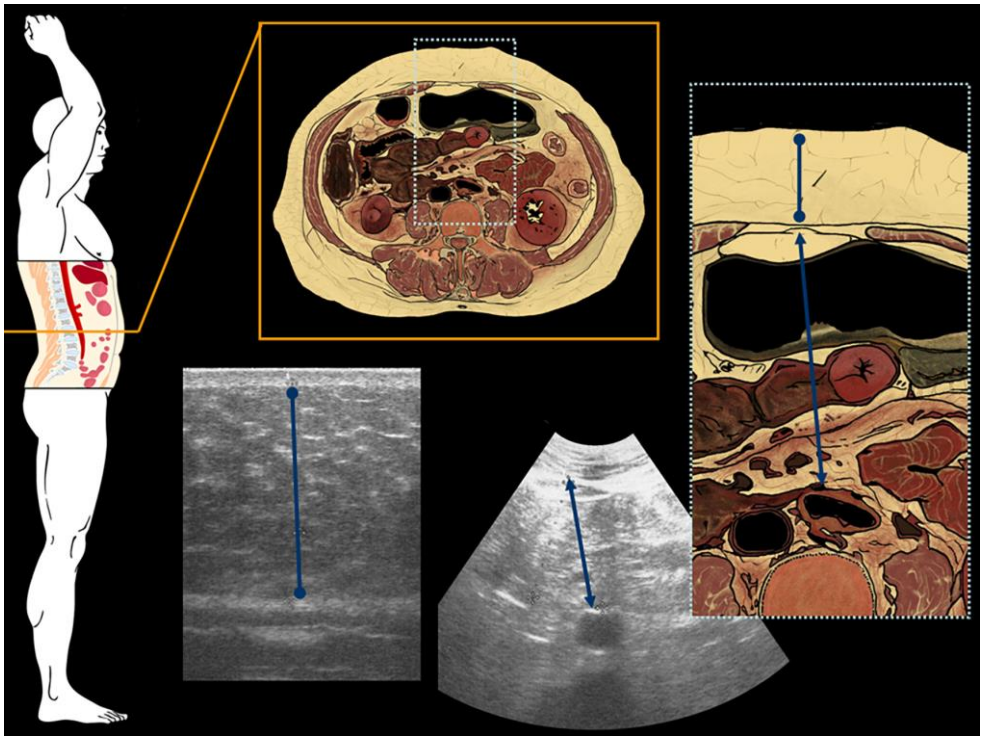


Figure 6: Axial section of the abdomen at a lumbar level, 2 cm above the umbilicus, graphically represented and in US B-mode. The blue line with rounded tips represent the maximum abdominal subcutaneous fat thickness, measured between the linea alba and the fat-skin interface, while the blue line with arrowheads represent the intra-abdominal fat thickness, measured between the anterior aortic wall and the linea alba [12].

3.3.2. Fatty Liver

The accumulation of triglycerides within the hepatocytes, or steatosis, is found in different pathologic condition, where steatosis may be secondary to alcohol or steatogenic drugs, or to metabolic dysregulation and insulin resistance, as in most cases of Non-Alcoholic Fatty Liver Disease (NAFLD). The role of imaging in fatty liver assessment is well-established. In fact, while the gold standard for Non-Alcoholic Fatty Liver Disease (NAFLD) diagnosis is still considered liver biopsy, non-invasive imaging techniques are largely used for steatosis diagnosis and quantification of liver fat content [22]. However, the real goal in NAFLD assessment is not to

determine the amount of fat in the hepatocytes, but to diagnose the associated hepatocyte damage in patients with Non-Alcoholic Steatohepatits (NASH) or the development of fibrosis [23].

US is used as a screening technique for liver steatosis, as it is largely available, harmless, and costless, and the accumulation of triglycerides within hepatocytes gives a typical hyperechoic appearance to the liver parenchyma, which may be homogeneous or heterogeneous [24]. To diagnose and estimate the severity of liver steatosis, the right kidney is usually used as an internal reference. However, only a qualitative judgement is possible, while the quantification of liver fat content with US is not possible. Moreover, US is operator-dependent, and its diagnostic accuracy is highly variable, especially when including milder cases, with sensitivity ranging from 53% to 100% and specificity between 77% and 98% [25]. Besides steatosis diagnosis, US scores have been also proposed for the diagnosis of NASH, including US fatty liver indicator (US-FLI), which include liver/kidney contrast, posterior attenuation of US beam, vessel blurring, difficult visualization of the gallbladder wall and diaphragm, and areas of focal sparing [26]. Finally, US-based methods measuring liver stiffness, named shear wave elastography (SWE), are used to estimate liver fibrosis, which is the first determinant of mortality in patients with NAFLD [27]. Different shear wave-based elastographic techniques are available, including point SWE and two dimensional (2D) SWE, which have been reported to be equivalent or superior when compared to transient elastography (TE) [28-29-30].

Computed tomography (CT) scan, preferably without contrast media administration, is able to detect liver steatosis, since the accumulation of fat within the hepatocytes leads to a decrease in attenuation, measured in Hounsfield units (HU). For CT scan too, besides the absolute attenuation value of liver parenchyma, the comparison with an internal reference is useful, and spleen is generally used. Different cut-off values have been used to diagnose liver steatosis, including hepatic attenuation 10 or 0 HU less than splenic attenuation at unenhanced CT, absolute liver attenuation <40 or <48 HU, liver-to-spleen ratio <0.9 or 1.1 [31-32-33]. Reported sensitivity and specificity are good (88% - 95% and 90% - 99%), however the detection may be hindered by increased liver attenuation due to iron overload, and most studies have been conducted only in patients with moderate-to-severe steatosis [25]. Finally, CT scan is not able to detect steatohepatitis or fibrosis, and, since it involves radiation exposure, it is not suitable as a screening or follow-up tool for the detection or quantitative assessment of liver steatosis.

Magnetic Resonance (MR) techniques, including imaging (MRI), and spectroscopy (MRS) are the most accurate non-invasive methods to diagnose and quantitatively evaluate liver fat infiltration, to the extent that they are now considered as gold standards. The detection of steatosis by means of

MR is mostly based on chemical shift, which may be defined as the shift in the oscillation frequency of different protons based on the chemical environment in which they are located: protons in water and fat have different precession frequencies when they are situated in a magnetic field [34]. Dual-phase and multiecho MRI use this phenomenon. In fact, multiple images are acquired with different and predefined echo times in order to register the signal of water and fat protons both when they are in phase (aligned in the same direction) and when they are out of phase (aligned in opposite direction). In these sequences, increasing liver fat content results in a signal drop in out-of-phase images [25]. While dual-phase MRI allows to estimate liver fat content only by using an internal reference standard, which is usually the spleen [35], multiecho imaging may be used to accurately measure proton density fat fraction (PDFF), i.e. the proportion of hepatic proton density which is attributable to fat, by using multiple couples of in- and out-of-phase echo times [36-37]. Modern multiecho sequences allow to address confounding factors such as T1 bias (by using a low flip angle) and T2* effect (especially important in patients with iron overload), as well as the multifrequency interference effects of fat protons (by incorporating spectral modeling of fat) [36-38-39-40]. Similarly, also ^{1}H-MRS, allowing the in-vivo study of liver molecular composition, permits the measurement of PDFF [41]. However, MRS has several drawbacks, including the scarce availability, and the need for complex post-processing requiring time and the support of a medical physicist. PDFF measured both by MRS and MRI has been shown to strongly correlate with histological steatosis grade, and, as opposed to liver biopsy, it can be used to follow patients longitudinally [42-43]. Apart from liver fat content estimation, MR may also provide an accurate measure of liver fibrosis by means of MR elastography (MRE), which allows to image liver stiffness based on the propagation of shear wave that are generated by a dedicated driver [25]. This technique has the advantages of being accurate [44], not influenced by obesity, and able to represent the stiffness of the whole liver parenchyma, as opposed to TE, US elastography, or liver biopsy. On the other hand, it is costly and time-consuming. Finally, several MR techniques have been proposed also to evaluate disease activity and inflammation in NAFLD, with immature and inconsistent results [45].

3.3.3. Sarcopenia

The term sarcopenia refers to an acute or chronic progressive skeletal muscle disorder characterized by reduction of muscle quantity, quality, and strength, associated with adverse outcomes including falls, functional decline, frailty, and mortality [46]. Even if it is strongly associated with age, many genetic and lifestyle risk factors have a role in its pathogenesis, along with different chronic

pathological conditions such as diabetes, cancer, cirrhosis, rheumatoid arthritis, and chronic obstructive pulmonary diseases. Moreover, muscle mass loss may occur along with fat mass increase, two combined processes that may culminate in sarcopenic obesity. Both sarcopenia and sarcopenic obesity have been associated with poor outcomes in a variety of pathological conditions [47-50].

The reported general definition of sarcopenia does not correspond to well-defined diagnostic criteria, which are debated in terms of which diagnostic methods and which cut-off values should be used [51].

Skeletal muscle cross-sectional areas or thickness, usually in the anterior portion of the thigh (quadriceps or rectus femoris), may be measured with US. Moreover, US may provide information about muscle quality, by looking at both muscle echogenicity (increased in case of fat accumulation), and stiffness by means of elastography (as a measure of fibrosis) [4]. Finally, the pennation angle, defined as the angle between muscle fibers and their inserting intramuscular tendon, is another US biomarker of muscle quality and strength [12].

The most widely used technique for sarcopenia assessment in research and clinical setting is DXA. The appendicular lean mass, i.e. the soft tissue found in arms and legs, is used as a surrogate marker of total body skeletal muscle mass [12]. More frequently, measures are indexed according to squared height or total weight, in order to be comparable among subjects independently of their size, obtaining lean mass index (total lean mass/squared height), appendicular lean mass index (appendicular lean mass/squared height), and skeletal muscle index (appendicular lean mass/total weight) [19].

After applying adequate density thresholds (from -30 HU to +150 HU), segmentation of CT images may provide skeletal muscle cross sectional area, which may be divided by squared height to obtain skeletal muscle index (SMI). Usually, the assessment is conducted on images acquired at the level of the L3 vertebrae, where cross sectional area has been shown to better correlate with total body skeletal muscle volume [11]. CT analysis may also provide information on skeletal muscle quality, both by assessing skeletal muscle density (SMD) which is influenced by intramuscular adipose tissue (fat droplets within muscle fibers), and by measuring intermuscular adipose tissue area (IMAT), i.e. fat between muscle fibers and within the fascia (see paragraph 1.3.1.) [2]. Disadvantages of CT scan have been discussed already, including radiation exposure which limits its use for the sole purpose of evaluating body composition.

MRI can be used to measure cross-sectional muscle area at L3 or mid-thigh level, while whole body protocols, even if considered the gold standard, are less used since they are time-consuming. MRI may also provide the fat content within skeletal muscle by means of chemical shift imaging [52].

References

1. https://vizhub.healthdata.org/gbd-compare/, accessed on 08/02/2021.

2. Prado CM, Heymsfield SB. Lean tissue imaging: a new era for nutritional assessment and intervention. JPEN J Parenter Enteral Nutr. 2014 Nov;38(8):940-53. doi: 10.1177/0148607114550189. Erratum in: JPEN J Parenter Enteral Nutr. 2016 Jul;40(5):742.

3. Ibrahim MM. Subcutaneous and visceral adipose tissue: structural and functional differences. Obes Rev. 2010 Jan;11(1):11-8. doi: 10.1111/j.1467-789X.2009.00623.x.

4. Lee MJ, Wu Y, Fried SK. Adipose tissue heterogeneity: implication of depot differences in adipose tissue for obesity complications. Mol Aspects Med. 2013 Feb;34(1):1-11. doi: 10.1016/j.mam.2012.10.001.

5. Heymsfield SB, Gonzalez MC, Lu J, Jia G, Zheng J. Skeletal muscle mass and quality: evolution of modern measurement concepts in the context of sarcopenia. Proc Nutr Soc. 2015 Nov;74(4):355-66. doi: 10.1017/S0029665115000129.

6. Tchernof A, Després JP. Pathophysiology of human visceral obesity: an update. Physiol Rev. 2013 Jan;93(1):359-404. doi: 10.1152/physrev.00033.2011.

7. Bosy-Westphal A, Braun W, Geisler C, Norman K, Müller MJ. Body composition and cardiometabolic health: the need for novel concepts. Eur J Clin Nutr. 2018 May;72(5):638-644. doi: 10.1038/s41430-018-0158-2.

8. Stefan N. Causes, consequences, and treatment of metabolically unhealthy fat distribution. Lancet Diabetes Endocrinol. 2020 Jul;8(7):616-627. doi: 10.1016/S2213-8587(20)30110-8.

9. Müller MJ, Braun W, Pourhassan M, Geisler C, Bosy-Westphal A. Application of standards and models in body composition analysis. Proc Nutr Soc. 2016 May;75(2):181-7. doi: 10.1017/S0029665115004206.

10. Shen W, Punyanitya M, Wang Z, Gallagher D, St-Onge MP, Albu J, Heymsfield SB, Heshka S. Visceral adipose tissue: relations between single-slice areas and total volume. Am J Clin Nutr. 2004 Aug;80(2):271-8. doi: 10.1093/ajcn/80.2.271.

11. Shen W, Punyanitya M, Wang Z, Gallagher D, St-Onge MP, Albu J, Heymsfield SB, Heshka S. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. J Appl Physiol (1985). 2004 Dec;97(6):2333-8. doi: 10.1152/japplphysiol.00744.2004.

12. Ponti F, De Cinque A, Fazio N, Napoli A, Guglielmi G, Bazzocchi A. Ultrasound imaging, a stethoscope for body composition assessment. Quant Imaging Med Surg. 2020 Aug;10(8):1699-1722. doi: 10.21037/qims-19-1048.

13. Mathieu P, Poirier P, Pibarot P, Lemieux I, Després JP. Visceral obesity: the link among inflammation, hypertension, and cardiovascular disease. Hypertension. 2009 Apr;53(4):577-84. doi: 10.1161/HYPERTENSIONAHA.108.110320.

14. Mazzoccoli G, Copetti M, Dagostino MP, Grilli M, Fontana A, Pellegrini F, Greco A. Epicardial adipose tissue and idiopathic deep venous thrombosis: an association study. Atherosclerosis. 2012 Aug;223(2):378-83. doi: 10.1016/j.atherosclerosis.2012.05.033.

15. Yip C, Dinkel C, Mahajan A, Siddique M, Cook GJ, Goh V. Imaging body composition in cancer patients: visceral obesity, sarcopenia and sarcopenic obesity may impact on clinical outcome. Insights Imaging. 2015 Aug;6(4):489-97. doi: 10.1007/s13244-015-0414-0.

16. Fosbøl MØ, Zerahn B. Contemporary methods of body composition measurement. Clin Physiol Funct Imaging. 2015 Mar;35(2):81-97. doi: 10.1111/cpf.12152.

17. Hong CW, Fazeli Dehkordy S, Hooker JC, Hamilton G, Sirlin CB. Fat Quantification in the Abdomen. Top Magn Reson Imaging. 2017 Dec;26(6):221-227. doi: 10.1097/RMR.0000000000000141.

18. Monti CB, Codari M, De Cecco CN, Secchi F, Sardanelli F, Stillman AE. Novel imaging biomarkers: epicardial adipose tissue evaluation. Br J Radiol. 2020 Sep 1;93(1113):20190770. doi: 10.1259/bjr.20190770.

19. Bazzocchi A, Ponti F, Albisinni U, Battista G, Guglielmi G. DXA: Technical aspects and application. Eur J Radiol. 2016 Aug;85(8):1481-92. doi: 10.1016/j.ejrad.2016.04.004.

20. Pietrobelli A, Wang Z, Formica C, Heymsfield SB. Dual-energy X-ray absorptiometry: fat estimation errors due to variation in soft tissue hydration. Am J Physiol. 1998 May;274(5):E808-16. doi: 10.1152/ajpendo.1998.274.5.E808.

21. Iacobellis G, Assael F, Ribaudo MC, Zappaterreno A, Alessi G, Di Mario U, Leonetti F. Epicardial fat from echocardiography: a new method for visceral adipose tissue prediction. Obes Res. 2003 Feb;11(2):304-10. doi: 10.1038/oby.2003.45.

22. Zhang YN, Fowler KJ, Hamilton G, Cui JY, Sy EZ, Balanay M, Hooker JC, Szeverenyi N, Sirlin CB. Liver fat imaging-a clinical overview of ultrasound, CT, and MR imaging. Br J Radiol. 2018 Sep;91(1089):20170959. doi: 10.1259/bjr.20170959.

23. European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. Diabetologia. 2016 Jun;59(6):1121-40. doi: 10.1007/s00125-016-3902-y.

24. Hertzberg BS, Middleton WD. Ultrasound: the requisites. Philadelphia: Elsevier - Health Sciences Division; 2015.

25. Esterson YB, Grimaldi GM. Radiologic Imaging in Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. Clin Liver Dis. 2018 Feb;22(1):93-108. doi: 10.1016/j.cld.2017.08.005.

26. Ballestri S, Lonardo A, Romagnoli D, Carulli L, Losi L, Day CP, Loria P. Ultrasonographic fatty liver indicator, a novel score which rules out NASH and is correlated with metabolic parameters in NAFLD. Liver Int. 2012 Sep;32(8):1242-52. doi: 10.1111/j.1478-3231.2012.02804.x.

27. Dulai PS, Singh S, Patel J, Soni M, Prokop LJ, Younossi Z, Sebastiani G, Ekstedt M, Hagstrom H, Nasr P, Stal P, Wong VW, Kechagias S, Hultcrantz R, Loomba R. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. Hepatology. 2017 May;65(5):1557-1565. doi: 10.1002/hep.29085.

28. Ferraioli G, Tinelli C, Lissandrin R, Zicchetti M, Dal Bello B, Filice G, Filice C. Point shear wave elastography method for assessing liver stiffness. World J Gastroenterol. 2014 Apr 28;20(16):4787-96. doi: 10.3748/wjg.v20.i16.4787.

29. Herrmann E, de Lédinghen V, Cassinotto C, Chu WC, Leung VY, Ferraioli G, Filice C, Castera L, Vilgrain V, Ronot M, Dumortier J, Guibal A, Pol S, Trebicka J, Jansen C, Strassburg C, Zheng R, Zheng J, Francque S, Vanwolleghem T, Vonghia L, Manesis EK, Zoumpoulis P, Sporea I, Thiele M, Krag A, Cohen-Bacrie C, Criton A, Gay J, Deffieux T, Friedrich-Rust M. Assessment of biopsy-proven liver fibrosis by two-dimensional shear wave elastography: An individual patient data-based meta-analysis. Hepatology. 2018 Jan;67(1):260-272. doi: 10.1002/hep.29179.

30. Crossan C, Tsochatzis EA, Longworth L, Gurusamy K, Davidson B, Rodríguez-Perálvarez M, Mantzoukis K, O'Brien J, Thalassinos E, Papastergiou V, Burroughs A. Cost-effectiveness of non-invasive methods for assessment and monitoring of liver fibrosis and cirrhosis in patients with chronic liver disease: systematic review and economic evaluation. Health Technol Assess. 2015 Jan;19(9):1-409, v-vi. doi: 10.3310/hta19090.

31. Boyce CJ, Pickhardt PJ, Kim DH, Taylor AJ, Winter TC, Bruce RJ, Lindstrom MJ, Hinshaw JL. Hepatic steatosis (fatty liver disease) in asymptomatic adults identified by unenhanced low-dose CT. AJR Am J Roentgenol. 2010 Mar;194(3):623-8. doi: 10.2214/AJR.09.2590.

32. Lawrence DA, Oliva IB, Israel GM. Detection of hepatic steatosis on contrast-enhanced CT images: diagnostic accuracy of identification of areas of presumed focal fatty sparing. AJR Am J Roentgenol. 2012 Jul;199(1):44-7. doi: 10.2214/AJR.11.7838.

33. Pickhardt PJ, Park SH, Hahn L, Lee SG, Bae KT, Yu ES. Specificity of unenhanced CT for non-invasive diagnosis of hepatic steatosis: implications for the investigation of the natural history of incidental steatosis. Eur Radiol. 2012 May;22(5):1075-82. doi: 10.1007/s00330-011-2349-2.

34. Mehta SR, Thomas EL, Bell JD, Johnston DG, Taylor-Robinson SD. Non-invasive means of measuring hepatic fat content. World J Gastroenterol. 2008 Jun 14;14(22):3476-83. doi: 10.3748/wjg.14.3476.

35. van Werven JR, Marsman HA, Nederveen AJ, Smits NJ, ten Kate FJ, van Gulik TM, Stoker J. Assessment of hepatic steatosis in patients undergoing liver resection: comparison of US, CT, T1-weighted dual-echo MR imaging, and point-resolved 1H MR spectroscopy. Radiology. 2010 Jul;256(1):159-68. doi: 10.1148/radiol.10091790.

36. Yokoo T, Bydder M, Hamilton G, Middleton MS, Gamst AC, Wolfson T, Hassanein T, Patton HM, Lavine JE, Schwimmer JB, Sirlin CB. Nonalcoholic fatty liver disease: diagnostic and fat-grading accuracy of low-flip-angle multiecho gradient-recalled-echo MR imaging at 1.5 T. Radiology. 2009 Apr;251(1):67-76. doi: 10.1148/radiol.2511080666.

37. Yokoo T, Shiehmorteza M, Hamilton G, Wolfson T, Schroeder ME, Middleton MS, Bydder M, Gamst AC, Kono Y, Kuo A, Patton HM, Horgan S, Lavine JE, Schwimmer JB, Sirlin CB. Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T. Radiology. 2011 Mar;258(3):749-59. doi: 10.1148/radiol.10100659.

38. Westphalen AC, Qayyum A, Yeh BM, Merriman RB, Lee JA, Lamba A, Lu Y, Coakley FV. Liver fat: effect of hepatic iron deposition on evaluation with opposed-phase MR imaging. Radiology. 2007 Feb;242(2):450-5. doi: 10.1148/radiol.2422052024.

39. Meisamy S, Hines CD, Hamilton G, Sirlin CB, McKenzie CA, Yu H, Brittain JH, Reeder SB. Quantification of hepatic steatosis with T1-independent, T2-corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. Radiology. 2011 Mar;258(3):767-75. doi: 10.1148/radiol.10100708.

40. Yu H, Shimakawa A, McKenzie CA, Brodsky E, Brittain JH, Reeder SB. Multiecho water-fat separation and simultaneous R2* estimation with multifrequency fat spectrum modeling. Magn Reson Med. 2008 Nov;60(5):1122-34. doi: 10.1002/mrm.21737.

41. Reeder SB, Cruite I, Hamilton G, Sirlin CB. Quantitative Assessment of Liver Fat with Magnetic Resonance Imaging and Spectroscopy. J Magn Reson Imaging. 2011 Oct;34(4):729-749. doi: 10.1002/jmri.22775.

42. Noureddin M, Lam J, Peterson MR, Middleton M, Hamilton G, Le TA, Bettencourt R, Changchien C, Brenner DA, Sirlin C, Loomba R. Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. Hepatology. 2013 Dec;58(6):1930-40. doi: 10.1002/hep.26455.

43. Tang A, Tan J, Sun M, Hamilton G, Bydder M, Wolfson T, Gamst AC, Middleton M, Brunt EM, Loomba R, Lavine JE, Schwimmer JB, Sirlin CB. Nonalcoholic fatty liver disease: MR imaging of liver proton density fat fraction to assess hepatic steatosis. Radiology. 2013 May;267(2):422-31. doi: 10.1148/radiol.12120896.

44. Singh S, Venkatesh SK, Loomba R, Wang Z, Sirlin C, Chen J, Yin M, Miller FH, Low RN, Hassanein T, Godfrey EM, Asbach P, Murad MH, Lomas DJ, Talwalkar JA, Ehman RL. Magnetic resonance elastography for staging liver fibrosis in non-alcoholic fatty liver disease: a diagnostic accuracy systematic review and individual participant data pooled analysis. Eur Radiol. 2016 May;26(5):1431-40. doi: 10.1007/s00330-015-3949-z.

45. Besutti G, Valenti L, Ligabue G, Bassi MC, Pattacini P, Guaraldi G, Giorgi Rossi P. Accuracy of imaging methods for steatohepatitis diagnosis in non-alcoholic fatty liver disease patients: A systematic review. Liver Int. 2019 Aug;39(8):1521-1534. doi: 10.1111/liv.14118.

46. Cruz-Jentoft AJ, Sayer AA. Sarcopenia. Lancet. 2019 Jun 29;393(10191):2636-2646. doi: 10.1016/S0140-6736(19)31138-9. Erratum in: Lancet. 2019 Jun 29;393(10191):2590.

47. Ooi PH, Hager A, Mazurak VC, Dajani K, Bhargava R, Gilmour SM, Mager DR. Sarcopenia in Chronic Liver Disease: Impact on Outcomes. Liver Transpl. 2019 Sep;25(9):1422-1438. doi: 10.1002/lt.25591.

48. Ryan E, McNicholas D, Creavin B, Kelly ME, Walsh T, Beddy D. Sarcopenia and Inflammatory Bowel Disease: A Systematic Review. Inflamm Bowel Dis. 2019 Jan 1;25(1):67-73. doi: 10.1093/ibd/izy212.

49. Yin J, Lu X, Qian Z, Xu W, Zhou X. New insights into the pathogenesis and treatment of sarcopenia in chronic heart failure. Theranostics. 2019 May 31;9(14):4019-4029. doi: 10.7150/thno.33000.

50. Shachar SS, Williams GR, Muss HB, Nishijima TF. Prognostic value of sarcopenia in adults with solid tumours: A meta-analysis and systematic review. Eur J Cancer. 2016 Apr;57:58-67. doi: 10.1016/j.ejca.2015.12.030.

51. Waters DL, Baumgartner RN. Sarcopenia and obesity. Clin Geriatr Med. 2011 Aug;27(3):401-21. doi: 10.1016/j.cger.2011.03.007.

52. Karlsson A, Rosander J, Romu T, Tallberg J, Grönqvist A, Borga M, Dahlqvist Leinhard O. Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water-fat MRI. J Magn Reson Imaging. 2015 Jun;41(6):1558-69. doi: 10.1002/jmri.24726.

4. EXPERIMENTAL PLAN AND ORGANIZATION OF DISSERTATION

This dissertation is organized in two principal sections, each of them composed by different manuscripts. The first section covers the diagnostic and prognostic use of different imaging biomarkers of **fatty liver**, while the second section focuses on the potential prognostic role of CT biomarkers of body composition, including **muscle quantity and quality and fat distribution.**

1) Fatty liver biomarkers

This section has two main research objectives: the first is to explore the role of imaging biomarkers in NAFLD **diagnosis and staging**; the second is to assess imaging biomarkers of fatty liver as potential **prognostic factors** in oncology.

To assess the role of NAFLD diagnostic tools, firstly we analyzed the existing guidelines on NAFLD screening and diagnosis in order to identify which are the clinical needs and to understand the possible utility of the candidate tests. Then, we mapped the available methods through a scoping review, and we summarized the evidences on their accuracy through a systematic review.

Since the actual impact of the application of NAFLD guidelines in high-risk patients in clinical practice is debated, we prospectively evaluated the feasibility and efficiency of the application of NAFLD guidelines in two different categories of patients at increased risk for advanced NAFLD: HIV-positive patients and patients with type 2 diabetes.

Finally, we registered a prospective study to evaluate diagnostic accuracy of US and MR techniques for the diagnosis of NASH and fibrosis in patients with NAFLD, which is currently ongoing.

As for the prognostic value of fatty liver biomarkers in oncologic patients, since liver steatosis alters liver microenvironment possibly influencing the development of liver metastases, we focused on the relationship between liver steatosis and the occurence of liver metastases in patients with solid tumors. Firstly, a systematic review and metanalysis have been conducted to evaluate the association of diffuse liver diseases (including steatosis) with the occurrence of synchronous and metachronous liver metastases in patients with solid tumors. Then, a retrospective study has been conducted to evaluate the impact of CT-assessed liver steatosis on liver metastases and overall survival in patients with rectal cancer.

2) Muscle quantity and quality and fat distribution

Since body composition is associated with long-term health outcomes in many oncologic diseases, particularly in elderly patients, we conducted a retrospective study to evaluate the association of CT biomarkers of muscle quantity and quality and fat distribution with disease outcomes in patients

with diffuse large B-cell lymphoma (DLBCL). Actually, DLBCL is a perfect model to study the impact of body composition on survival, because a baseline Positron Emission Tomography (PET)-CT scan is recommended, thus the reference test (unenhanced CT scan) to evaluate fat distribution and lean mass quality and quantity is available in all patients.

Body composition is likely to have an impact on COVID-19 course, since ectopic fat is linked to systemic inflammation which may favor the COVID-19 cytokine storm, and sarcopenia may affect respiratory efficiency. Moreover, elderly patients, typically sarcopenic, are known to have a very poor prognosis. Therefore, we conducted a retrospective study to evaluate the impact of CT biomarkers of body composition on COVID-19 outcomes, also performing a mediation analysis in order to understand if body composition is a mediator in the effect of age on death.

5. FATTY LIVER

5.1. The role of imaging in fatty liver diagnosis and staging

5.1.1. Systematic review of existing guidelines for NAFLD assessment*

Summary: To assess the role of imaging biomarkers in NAFLD diagnosis and staging, firstly we analyzed the existing guidelines on NAFLD screening and diagnosis in order to identify which are the clinical needs and to understand the possible utility of the candidate tests.

Aim: In this systematic review, guidelines on non-alcoholic fatty liver disease (NAFLD) were evaluated, aiming at a guideline synthesis focusing on diagnosis and staging.

Methods: A systematic literature search was conducted on any relevant database or institutional website to find guidelines on NAFLD assessment intended for clinical use on humans, in English, published from January 2010 to August 2020. Included guidelines were appraised using the AGREE II Instrument; those with higher scores and intended for use in adult patients were included in a comparative analysis.

Results: Fourteen guidelines were included in the systematic review, eight of which reached an AGREE II score sufficiently high to be recommended for clinical use, of which one developed for pediatric patients only. British and North American guidelines received the highest scores. Most guidelines recommend a screening or case-finding approach in patients with metabolic risk factors who are at increased risk of steatohepatitis or fibrosis. Ultrasound is mostly recommended to confirm steatosis, while the presence of metabolic syndrome, liver function tests, fibrosis scores and elastographic techniques may help selecting high-risk patients to be referred to the hepatologist, who may consider liver biopsy, although referral criteria for liver biopsy are not clearly defined. Most guidelines identify the development of noninvasive tests to replace liver biopsy as a research priority.

Conclusion: Several high-quality guidelines exist for NAFLD assessment, with no complete agreement on whether to screen high-risk patients and on the tests and biomarkers suggested to stratify patients and select those to be referred to liver biopsy.

***** Monelli F, Venturelli F, Bonilauri L, Manicardi E, Manicardi V, Rossi PG, Massari M, Ligabue G, Riva N, Schianchi S, Bonelli E, Pattacini P, Bassi MC, Besutti G. Systematic review of existing guidelines for NAFLD assessment. Hepatoma Res 2021;7:25. http://dx.doi.org/10.20517/2394-5079.2021.03

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is associated with metabolic disorders such as obesity, type 2 diabetes (T2D), hypertension and dyslipidemia [1]. NAFLD is a continuum of clinical entities, from simple hepatic steatosis to inflammatory non-alcoholic steatohepatitis (NASH). In this process there is an increasing risk of developing liver fibrosis and cirrhosis [2], ultimately leading to a higher risk of end-stage liver disease and hepatocellular carcinoma. NAFLD is characterized by an accumulation of triglycerides in lipid droplets inside the hepatocytes [3] in the absence of other causes of liver injury [4], including excessive alcohol intake. The transition from NAFLD to NASH can be assessed accurately only with histology on liver biopsy samples. Histologically, NAFLD is defined by the presence of steatosis in more than 5% of hepatocytes [5], with no evidence of cellular injury such as hepatocyte ballooning. NASH, instead, is characterized by the presence of steatosis and the presence of inflammation and hepatocytic injury, which is frequently associated with liver fibrosis [6].

The incidence of NAFLD is increasing worldwide and is currently the most common liver disease [7]. Its global prevalence of 25% varies across different geographical areas, from 14% in Africa to 32% in the Middle East [8]. The distribution of NAFLD appears to be linked to socioeconomic status, with a higher prevalence in industrialized countries [9], although incidence is increasing in every social and ethnic group [10].

The high prevalence and the increased risk of developing severe liver diseases, including HCC, make NAFLD a major health issue worldwide, and thus a potentially considerable burden on health care systems. Among the questions associated with this evolving epidemiology, two impact the management of NAFLD patients the most. The first is how to select patients at high risk of progression to severe disease; the second concerns which tests to perform on these patients, including noninvasive examinations and liver biopsy. This second problem is especially important given that the very first drugs to treat NASH are in phase II and III trials and are expected to be accepted soon for clinical use [11, 12].

In recent years, many national and international guidelines have been developed to provide recommendations on how to select patients to be referred for diagnosis and which procedure should be used for the assessment of NAFLD. The aim of this systematic review was to produce a comprehensive and updated review of existing guidelines focusing on NAFLD diagnosis and staging.

METHODS

Guideline eligibility

Inclusion criteria were: 1) guidelines on NAFLD assessment intended for clinical use on humans; 2) English language; 3) publication date from January 2010 to August 2020. Exclusion criteria were: 1) outdated guidelines, i.e., guidelines for which an updated version is included in the review; 2) review of guidelines. Practice guidance released by national or international associations, which are different from guidelines and report statements instead of recommendations, were formally excluded from guideline selection, but were used to integrate and/or update recommendations extracted from formal guidelines released by the same association.

Guideline search and selection

A systematic review of the literature was carried out according to PRISMA guidelines [13]. MEDLINE, Embase and CINAHL databases were explored focusing on guidelines on NAFLD diagnosis and management produced in the last 10 years. The search string originally used for MEDLINE and then adapted to every database was: ["Fatty Liver"[Mesh] OR "Non-alcoholic Fatty Liver Disease"[Mesh] OR “fatty liver” or “NAFLD or “non-alcoholic fatty liver disease”)].

The literature search was completed by searching relevant websites such as Trip Database, Australia National Health and Medical Research Council, American College of Physicians, American Medical Association, Institute for Clinical Systems Improvement, Institute of Medicine, National Guidelines Clearinghouse, National Institute for Health and Clinical Excellence, Royal College of Physicians, Scottish Intercollegiate Guidelines Network and World Health Organization and by exploring the bibliography of relevant studies and reviews on the topic.

One reviewer (FM) screened the search results based on title/abstract, and then two reviewers (FM and GB) independently examined eligibility based on the full text of the relevant articles. When unclear, inclusion was decided by group consensus with the other authors involved (FM, GB, PGR).

Recommendations extraction and synthesis and guideline quality appraisal

One reviewer (FM) extracted the recommendations of each guideline with respect to the definition of NAFLD and alcohol intake thresholds for defining a non-alcoholic etiology, criteria for eligibility for screening of at-risk population groups, screening modality for at-risk groups, procedures for NAFLD assessment and staging, including diagnosis of NASH and fibrosis, assessment tools and criteria of secondary liver disease, the definition of advanced NAFLD, the role of imaging and of liver biopsy and research priorities. A cross-check of the extracted data for

accuracy was conducted by another reviewer (GB). Every guideline was evaluated using the Appraisal of Guidelines for Research and Evaluation (AGREE) II Instrument [14], a tool designed to appraise the methodological rigor, transparency and applicability of clinical guidelines. The tool provides a rate in percentage for each of six domains, a comprehensive evaluation of the appraised guideline (from 1 to 7) and a final judgment on recommendations for clinical use. Two authors (FM and GB) applied the tool, when scores differed between the two reviewers, a final score was defined by consensus. Guidelines reporting recommendations for adult population and achieving an overall score ≥ 6 using the AGREE II tool were included in a synthesis of extracted recommendations. A synopsis of the extracted topics is reported.

RESULTS

Included guidelines and quality appraisal

Of the 140 extracted records after duplicate removal, 107 were excluded by title and abstract screening and 18 by full text reading, while 14 guidelines were included in this systematic review (Figure 1). One of the selected guidelines, the US multi-societal guidelines, USA 2012 [2], was further integrated with some additional recommendations published in a practice guidance in 2018 [15]. This guidance was excluded by the formal selection but was used to integrate USA 2012; thus, in the following synthesis, we refer to the combination of both as USA 2012/2018. Among the 14 included guidelines, 5 were produced in Europe [5,16,17,18,19], 5 in Asia and Oceania [20,21,22,23,24], 2 in USA [2,25], including the only one specifically developed for the pediatric population [25], one in South America [26] and one from a supranational institution [27].

The results of the consensus appraisal according to the AGREE II Instrument are presented in Table 1. The domains with lower scores on average were the rigor of development and stakeholder involvement. The rigor of development domain contains seven subfields which are related to the methods adopted in the production of the guideline. Many of the guidelines analyzed did not use any clear systematic method to search and select the evidence, did not provide a direct connection between scientific works and recommendations, did not describe the strengths and limitations of the body of evidence and did not provide a procedure or a timeline for guideline updating. For example, WORLD-WGO guideline [27] has low scores in every of those cited domains, while being an excellent guideline in terms of clinical applicability. As concerns the stakeholder involvement domain, many guidelines failed to present how and if the views and preferences of the target population have been evaluated, while most of the guidelines included in the final analysis excelled in taking into consideration the point of view of involved healthcare professionals. Among every

analyzed guideline, only UK [16] has a specific section and demonstrates that the patient's point of view was considered in its production.

Among the included guidelines, two reached a score of 7 [2,15,16], and other six the score of 6 [5,19,21,22,25,27] and were selected as recommended for clinical use by the authors who performed the appraisal, one of these intended for use in pediatric population only.

Publication year ranged from 2010 to 2019, i.e., almost the entire range of our search, with 6 documents released before 2015. Methodological quality of the guidelines was not associated with publication year.

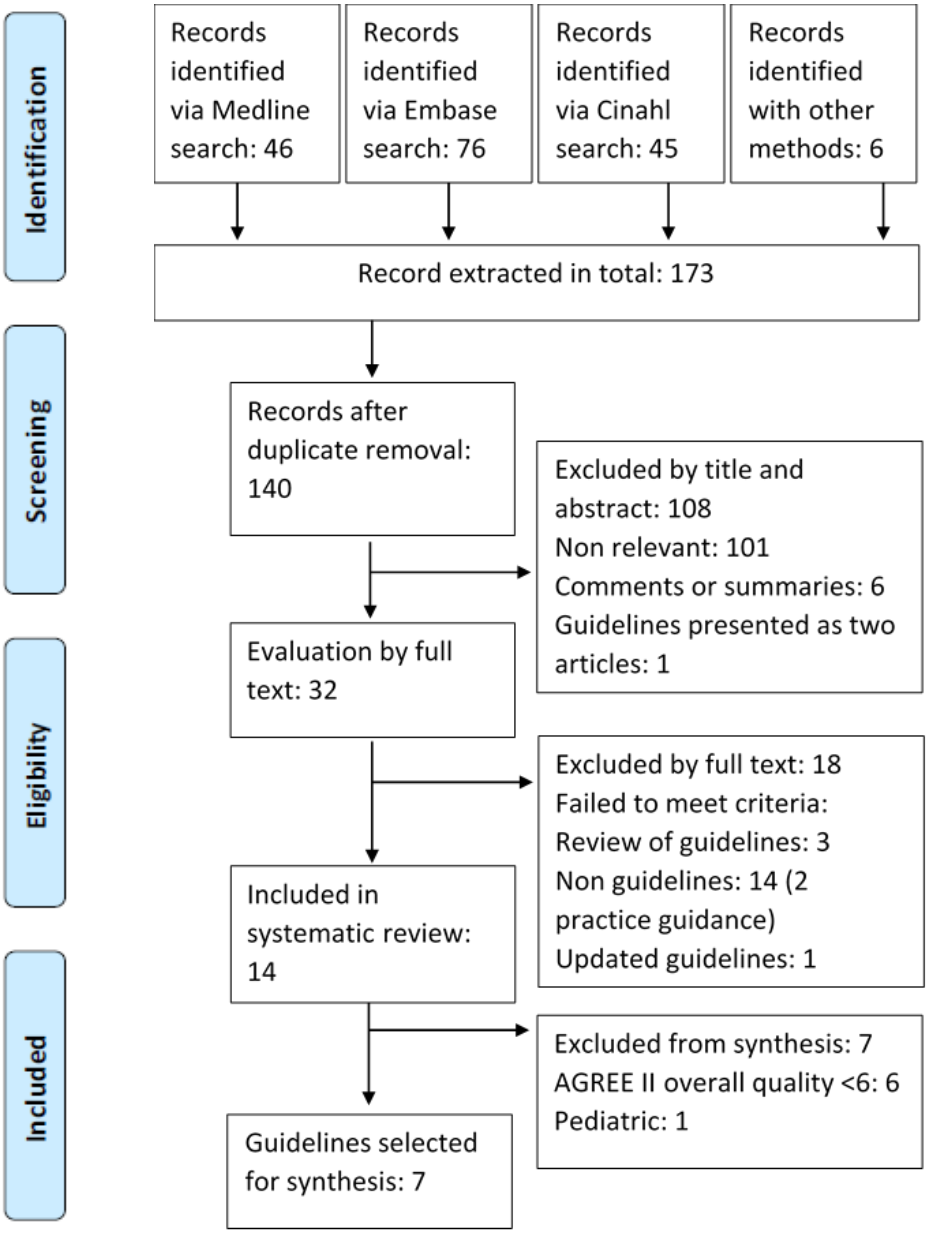


Figure 1: PRISMA flowchart of guidelines searching.

Country / Year Agree Domain	EUROPE [5]	UK [16]	ASIA PACIFIC [20]	CHINA [21]	KOREA [22]	JAPAN [23]	WORLD [27]	POLAND [17]	SPAIN [18]	ITALY [19]	USA [2/15]	USA PED [25]	CHILE [26]	CHINA [24]
	2016	2016	2017	2018	2013	2015	2014	2018	2018	2010	2012	2017	2014	2013
1 SCOPE AND PURPOSE	83	94	61	83	94	78	78	89	83	89	94	78	61	44
2 STAKEHOLDER INVOLVEMENT	50	83	39	33	61	50	50	50	44	67	61	56	39	33
3 RIGOUR OF DEVELOPMENT	73	79	29	38	77	40	21	15	40	69	77	71	31	29
4 CLARITY OF PRESENTATION	72	89	44	61	78	56	83	72	89	89	94	83	56	56
5 APPLICABILITY	75	92	29	25	71	67	79	42	63	58	83	67	38	42
6 EDITORIAL INDEPENDENCE	75	100	83	67	83	75	42	50	58	67	75	67	33	58
Overall	6	7	4	6	6	5	6	3	5	6	7	6	4	4
Recommended	YES	YES	NO	YES	YES	NO	YES	NO	NO	YES	YES	YES	NO	NO

Table 1: Appraisal of included guidelines using AGREE II tool. Domain boxes are colored depending on the score they received, and they are expressed in percentage: red 0% - 35%; yellow 36% - 54%; blue 55% - 74%; light green 75% - 90%; dark green 91% - 100%. Overall grading is expressed in points from 1 (lowest) to 7 (highest).

Recommendation synthesis

Table 2 presents a summary of the recommendations reported by the guidelines on the adult population and selected as recommended for clinical use according to the quality appraisal. Here, we focus on critical points with higher impact on patient management and safety as well as on the workload for health systems.

Who should be screened for NAFLD: None of the included guidelines recommends a screening for NAFLD in the general population, although most of them recommend a screening or case-finding approach among high-risk patients, mostly defined as patients with metabolic risk factors, including T2D, obesity and metabolic syndrome (MetS) [2,5,16,17,18,19,21,22,23,24,26,27]. US guidelines (USA 2012/2018) [2,15] still do not recommend such screening because of uncertainties surrounding diagnostic tests and treatment options, although in the last update (USA2018) [15], the use of clinical decision aids such as fibrosis scores or transient elastography is suggested to find T2D patients at higher risk of fibrosis. Analyzing included guidelines and respective references, the cause of the discrepancy on strategies of high-risk patient screening may be caused by a lack of large studies on the topic. As to the identification of patients at risk of NAFLD, all guidelines require the exclusion of patients with secondary causes of liver steatosis, including steatogenic drugs, and above all, excessive alcohol consumption. In this regard, alcohol thresholds are similar in most guidelines but not identical. In fact, USA 2012/2018 [2/15] uses alcohol units instead of grams, resulting in higher thresholds (>196 g in women and >294 g in men weekly) as opposed to other guidelines (>20 g in women and >30 g in men daily or >140 g in women and >210 g in men weekly).

Which noninvasive tests should be used: Ultrasonography is the preferred first-line diagnostic procedure to screen for NAFLD, while steatosis biomarkers, including fatty liver index (FLI) [28], are considered acceptable alternatives. Liver enzymes are used to guide the identification of patients at higher risk of advanced disease (NASH, fibrosis), even if in some patient categories, especially T2D, advanced disease cannot be ruled out in the presence of normal plasma levels of liver enzymes (EUROPE 2016) [5]. Fibrosis scores (e.g., NFS [29], FIB-4, ELF [enhanced liver fibrosis] score), VCTE or magnetic resonance elastography (MRE) are acceptable noninvasive procedures to identify patients with a low risk of advanced fibrosis (>F2 = portal fibrosis with few septa). If >F2 cannot be excluded, patients should be referred to the hepatologist, who may decide to perform liver biopsy based on the patient's baseline risk and general clinical conditions, and the opportunity for therapies, frequently within clinical trials in selected settings (EUROPE 2016) [5], (USA 2012/2018) [2,15].

When to perform liver biopsy: The hepatologist should consider liver biopsy in NAFLD patients with high risk of advanced disease (NASH or advanced fibrosis). Hence, the presence of metabolic syndrome, abnormal liver function tests, fibrosis biomarkers and/or liver stiffness measurements can be used to target patients for liver biopsy. In a recent Chinese guideline (CHINA 2018) [21], high serum levels of CK-18 fragments (M30 and M65) were also introduced as a possible reason for liver biopsy. Finally, liver biopsy is generally considered when other concurrent chronic liver diseases cannot be excluded. All this considered, the decision to perform a liver biopsy is ultimately made by the hepatologist, without common and specific referral criteria, since the definition of high-risk NAFLD patient may be variously interpreted.

Moreover, some of the included guidelines [16,19,21,22] also focus on research priorities related to these points, underlying the necessity to develop non-invasive tests and biomarkers to diagnose NASH and fibrosis, or to assess disease progression, in order to replace liver biopsy or limit its use to a restrict number of patients.

Country	EUROPE [5]	WORLD [27]	UK [16]	CHINA [21]	KOREA [22]	ITALY [19]	USA [2/15]
Year	2016	2014	2016	2018	2013	2010	2012/2018
NAFLD definition	HS involving >5% of the hepatocytes, associated with insulin resistance without other causes of HS	Fat accumulation in the liver exceeding 5% of its weight	HS in >5% of the liver without other causes of HS	HS involving ≥5% of the hepatocytes without other causes of HS	Fat infiltration of the liver on imaging or histology (>5% of the liver weight) without other causes of HS	HS defined by imaging or histology without other causes of HS	HS defined by imaging or histology without other causes of HS
Alcohol threshold value	Daily >20 g in women, >30 g in men	Daily >20 g in women, >30 g in men	Daily >20 g in women, >30 g in men	Weekly >140 g in women and >210 g in men for the previous 12 months	Weekly >140 g in women and >210 g in men for the previous 2 years	Daily >20 g in women, >30 g in men	Weekly >196 g in women and >294 g in men for the previous 2 years
People at risk	Insulin resistance, T2D, metabolic risk factors and persistently abnormal LEs	Central obesity, T2D, dyslipidemia, MetS, altered LEs, fatty liver on US	T2D (glycaemia≥1.1 g/l) and MetS (NCEP criteria)	T2D (glycaemia≥1.1 g/l), MetS (NCEP criteria), obesity, hypertriglyceridemia, elevated ALT and gGT	T2D, obesity, dyslipidemia and MetS (NCEP criteria)	Obesity or MetS. Male at a higher risk	Obesity, T2D mellitus, dyslipidemia, MetS
Population screening	No	No	No	No	No	No	No
Risk patient screening	Yes, people at risk	Yes, people at risk	Yes, people at risk	Yes, patients at risk	Equivocal. Attention if AST and ALT increased or risk factors	No	No
Screening methods	US, LEs, dietary and physical assessment.	LEs, US	FLI, US	LEs, US	LEs, US	No	No
NAFLD assessment	US, LEs and assessment of dietary and physical activity habits. Scores (FLI, SteatoTest and NAFLD liver fat score) or other imaging (CT, MRI or H-MRS) can replace US. If no HS and no abnormal LEs, follow up in 3/5 years.	Assess liver and pancreatic function, test hepatic viruses and assess eventual MetS. If risk factors, evaluate liver disease. If not treat MetS and follow up in 6 months.	FLI ≥60 or US. No non-invasive tools are recommended for diagnosis of severe NAFLD (HS >30%). Do not use routine liver blood tests to rule out NAFLD.	US, LEs and investigate dietary and exercise habits. TE is a good alternative to US in assessing HS quantitatively.	LEs, if positive, US. Secondarily CT, MRI, and MRS could be helpful in the evaluation of the amount of fat in the liver. Patients with NAFLD should be screened for advanced liver fibrosis and NASH.	LEs, US and assessment of MetS.	USA2012: not recommended USA2018: Suspect NAFLD in people with HS on imaging with symptoms of liver disease or abnormal LEs. Exclude competing etiologies and assess liver status: liver imaging and serum

							ferritin, iron saturation, autoantibodies.
Diagnosis of advanced NAFLD	Assess serum marker (ALT, AST, gGT) and fibrosis (NFS, FIB-4, ELF or FibroTest). In case of Fibrosis (>=F2) or abnormal LEs (any increase in ALT, AST or gGT) refer patient to a specialist. Otherwise, follow up in 2 years.	Evaluate potential sign of cirrhosis with clinical visit, and measuring LEs, albumin and platelets.	Test the presence of advanced liver fibrosis: ELF>10,51 is positive. Refer adults and young people with advanced liver fibrosis to a hepatologist and reassess adults every 3 years and children/young people every 2 years.	Do not use imaging or laboratory tests to exclude NASH. The presences of MetS, persistent elevated serum ALT level, increased serum cytokeratin (CK)-18 fragment (M30 and M65) levels are suggestive of NASH in NAFLD patients.	NAFLD FS, TE, and magnetic resonance elastography could be helpful in the estimation of advanced liver fibrosis in patients with NAFLD.	Perform: (1) Laboratory tests (AST, ALT, lipid profile, insulin sensitivity), (2) US and (3) Physical examination. In case of no risk factors for NASH/Fibrosis, Lifestyle modifications and control of MetS and follow up at 6/12 months.	USA2012: perform NAFLD FS to assess the risk of fibrosis and or cirrhosis. Evaluate the presence of MetS. USA2018: in patients with T2D NASH/FS, fibrosis-4 index or VCTE stratify patients' risk.
Diagnosis of NASH/Fibrosis	In NAFLD patients with Fibrosis or abnormal LEs, consider LB and start monitoring and/or therapy.	If there are sign of cirrhosis or LEs altered after 6 months of follow up and MetS treatment, consider LB.	Monitor cirrhosis in adults and young people with ELF>16. The gold standard for NASH diagnosis is LB but studies on its cost-effectiveness are inconclusive.	Perform LB in case of MetS, persistently high serum aminotransferases, and/or high serum levels of CK-18 fragments (M30 and M65).	LB should be considered in cases in which NASH or advanced liver fibrosis is suspected in NAFLD patients and when concurrent other chronic liver diseases cannot be excluded.	Perform LB in case of NASH/Fibrosis risk factors (age>45, obesity, diabetes or insulin resistance, MetS, low platelets, low albumin, AST>ALT and imaging sign of portal hypertension) and if altered liver tests without. If simple HS, follow-up. If NASH/Fibrosis treatment.	Consider LB: (1) presence of MetS, (2) altered NFS>0.676 and (3) to exclude competing etiologies for HS and other chronic liver diseases. USA2018 adds: (4) suspected NAFLD, high serum ferritin and increased iron saturation with homozygote or heterozygote C282Y HFE mutation.
Role of Imaging	US as first line. CT, MRI and H-MRS are optional or in case US cannot be performed.	US as first line in the assessment of fatty liver, suspect of NAFLD. Abdominal CT can be performed if US is not informative.	US as first line in presence of altered live enzymes.	US for screening, TE as an alternative.	US as first line. CT, MRI and MRS to obtain a quantification of liver fat. In patients with NAFLD transient elastography and MR	US has the role of first line diagnosis of fatty liver and second line in the assessment of the risk of fibrosis and	US, CT, and MR do not reliably assess NASH and fibrosis in patients with NAFLD; VCTE and other tools such as MR elastography, are rarely used in clinical practice

					elastography for the study of liver fibrosis.	NASH by evaluating portal hypertension.	and cannot be recommended.
Role of biopsy	NAFLD patients with Fibrosis or abnormal LEs	LB rule out other diseases, grade and stage disease; cannot reliably distinguish NASH from alcoholic steatohepatitis. Performed in every patient with suspected advanced liver disease who do not improve with lifestyle modification.	No clear indication, although it remains the gold standard for NAFLD diagnosis.	NAFLD patients with MetS, persistently high serum aminotransferases, and/or high serum levels of CK-18 fragments.	Suspected NASH/Fibrosis or when concurrent other chronic liver disease cannot be excluded. LB remains the gold standard for diagnosing NAFLD.	Risk of NASH/Fibrosis or no normalization of liver functional tests at follow-up.	Risk of NASH/Fibrosis or in case of suspected NAFLD if competing etiologies of HS and co-existing chronic liver diseases cannot be excluded. USA 2018 adds: alteration of iron metabolism with mutations of C282Y HFE indicates a biopsy.
Research recommend ations	X	X	To identify the most accurate non-invasive tests to diagnose NAFLD in adults.	To develop biomarkers (serum, genomics, proteomics, glycomics, metabolomics and new imaging techniques) to replace liver biopsy for the diagnosis of NASH and liver fibrosis	To evaluate the availability and usefulness of non-invasive biomarkers in the assessment of disease progression and treatment response.	Study of scores, laboratory and imaging, transient elastography to be used as non-invasive predictors of early cirrhosis	X

Table 2: Synthesis of recommendations from guidelines on diagnosis and management of NAFLD and NASH/Fibrosis in adult population, including guidelines recommended for clinical use after quality appraisal. US, Ultrasound; FLI, Fatty Liver Index; LEs, Liver Enzymes; NFS, NAFLD fibrosis score, Metabolic Syndrome, MetS; T2D, Type 2 Diabetes; Liver Biopsy, LB; HS, Hepatic Steatosis; TE, Transient elastography; VCTE, Vibration Controlled Transient Elastography; ELF, Enhanced Liver Fibrosis score.

DISCUSSION

Fourteen guidelines have been released on the assessment of NAFLD in recent years and were included in this systematic review. Of the 14, eight obtained a high score on the AGREE II Instrument and are consequently recommended for clinical use, one in the pediatric setting and the other seven for adult management.

The main limitation of this systematic review is that we only included guidelines that were available in English, including translations into English. As clinical recommendations are usually developed for the local clinical community, making them accessible to those practitioners is often preferred rather than to an international audience. Thus, we cannot rule out that some relevant documents have been excluded. Furthermore, as recommendations on NAFLD may be included in guidelines on wider health problems of internal medicine, we may have missed some important documents containing some information relevant to the scope of our review. Nevertheless, the documents retrieved include recent recommendations from a broad range of countries, thereby permitting a meaningful synopsis of the quality and contents of current guidelines. Although a similar review has been published [30], it includes fewer guidelines, probably due to the different search criteria and literature sources (one database only and limited to years 2016 to 2018). Moreover, while the previous revision generally addressed NAFLD assessment and treatment, the present review specifically focuses on NAFLD screening in high-risk patients.

The British and North American guidelines received the highest scores. Notably, among the guidelines with high scores, the most problematic domains in the AGREE II appraisal were Stakeholder Involvement, Rigour of Development and Applicability. In particular, only the British guidelines involved patients and gave clear evidence of this involvement.

While not all the guidelines agree on the appropriateness of NAFLD screening in high-risk patients, NAFLD assessment methods and the role of biopsy as the only definitive diagnostic tool are substantially similar, despite the criteria for referring patients to biopsy not being clearly defined. It is worth noting that the issue of screening, with its impact on health service resources and organization and targeting people who have risk factors but not necessarily any disease, requires the involvement of all stakeholders, including citizens, patients and policy makers. The weakness of the included guidelines on this point may be justified by the difficulties in identifying any patient organization for this health problem. Nevertheless, any future effort made towards developing recommendations on this topic should start with the awareness of the need to involve citizens and patients, from the scoping of the guidelines to their dissemination.

Despite being one of the leading causes of chronic liver disease, with an increasing incidence worldwide [31,32], NAFLD is still frequently overlooked, even as a cause of hepatocellular

carcinoma. In Europe, the presence of NAFLD as a cause of liver disease has been shown to double the risk of not receiving appropriate surveillance [33]. This dictates the need for a more systematic approach to NAFLD in order to prevent the transition to advanced stages.

While the definition of NAFLD based on the presence of hepatic steatosis and the exclusion of other causes of fatty liver are identical across guidelines, there are slight differences in terms of the secondary causes of steatosis. In particular, the definition of excessive alcohol consumption is slightly different in terms of time span and volumes. Currently, no guideline suggests a screening of the whole population for the diagnosis of NAFLD, although increasing emphasis is being placed on the early identification of NAFLD in patients with risk factors. As for which risk factors should be taken into consideration, it is generally agreed that T2D and MetS are red flags for NAFLD, though only one guideline (WGO 2014) [27] includes elevated liver enzymes as a factor to select patients to screen. Every guideline selected and included in the synthesis agrees on the adoption of noninvasive methods to diagnose NAFLD in patients at risk, the main one being ultrasound. Nevertheless, one guideline (UK 2016) [16] underlines that sonography can reliably detect steatosis when its histological grade is over 30%, a threshold well over the 5% consensually indicated by every guideline for the diagnosis of NAFLD.

Other methods, such as different elastography techniques including TE, MRE and US-based methods, serum markers and clinical scores, have been proposed to identify patients at higher risk of advanced NAFLD; the choice depends on the patient's characteristics and the hospital setting to which he/she is referred.

Currently, liver biopsy is still considered the only diagnostic tool that can consistently assess both NAFLD and NASH/fibrosis, especially due to the inability of imaging techniques to accurately diagnose NASH [34]. Nevertheless, the adoption of liver biopsy in all patients at high risk of advanced NAFLD to confirm the diagnosis would refer a large number of patients to this invasive test, given the high and increasing prevalence of this disease. Thus, the opportunity of such a strategy for assessing the disease is debated, mainly because of doubts concerning the cost-benefit ratio (UK 2016) [16] and the medical risks stemming from its invasive nature (EUROPE 2016) [5]. Another role of liver biopsy is its ability to detect other underlaying hepatic diseases that can be the cause of hepatic steatosis or mimic it. Because of this, USA 2012/2018 [2,15] specifically suggests performing liver biopsy on NAFLD patients whenever a coexisting liver disease cannot be excluded. Given these premises, it is not surprising that criteria for referring are unclear in the reviewed guidelines, essentially conferring the individual hepatologist with great discretion.

The limitations of liver biopsy and the lack of clarity of its referral criteria result in the need to focus research efforts on the development of noninvasive tests able to correctly assess advanced NAFLD

in order to limit or avoid the use of liver biopsy in clinical practice [16,19,21,22]. In regard to this, some recent studies have demonstrated a reduction in referral rates to secondary care when using a combination of noninvasive serum or imaging biomarkers of advanced disease is used [35, 36, 37, 38]. Since some of the included guidelines are somewhat outdated [27, 22, 19], it is possible that upcoming guideline revisions will also focus on a better definition of the use of combined noninvasive biomarkers in clinical practice.

In conclusion, several high-quality guidelines exist for NAFLD assessment. The main area of discrepancy between recommendations from different guidelines is whether screening high-risk patients is opportune and if so, what are the best strategies to do so. A screening limited to patients with metabolic risk factors is mostly recommended, preferably through liver US to confirm steatosis and by means of liver function tests, fibrosis scores and elastographic techniques to identify patients at risk of advanced disease, who should be considered for liver biopsy.

REFERENCES

1 Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. Nat Rev Gastroenterol Hepatol. 2013 Jun;10(6):330-44. [PMID: 23507799 DOI: 10.1038/nrgastro.2013.41]

2 Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, et. al. The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology. 2012 Jun;55(6):2005-23. [PMID: 22488764 doi: 10.1002/hep.25762.]

3 Sunny NE, Bril F, Cusi K. Mitochondrial Adaptation in Nonalcoholic Fatty Liver Disease: Novel Mechanisms and Treatment Strategies. Trends Endocrinol Metab. 2017 Apr;28(4):250-260. [PMID: 27986466 DOI: 10.1016/j.tem.2016.11.006]

4 Marchesini G, Petta S, Dalle Grave R. Diet, weight loss, and liver health in nonalcoholic fatty liver disease: Pathophysiology, evidence, and practice. Hepatology. 2016 Jun;63(6):2032-43. [PMID: 26663351 DOI: 10.1002/hep.28392]

5 European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. J Hepatol. 2016 Jun;64(6):1388-402. [PMID: 27062661 DOI: 10.1016/j.jhep.2015.11.004]

6 Kleiner DE, Makhlouf HR. Histology of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis in Adults and Children. Clin Liver Dis. 2016 May;20(2):293-312. [PMID: 27063270 DOI: 10.1016/j.cld.2015.10.011]

7 Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018 Jan;67(1):123-133. [PMID: 28802062 DOI: 10.1002/hep.29466]

8 Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology. 2016 Jul;64(1):73-84. [PMID: 26707365 DOI: 10.1002/hep.28431]

9 Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to HCC: current concepts and future challenges. Nat Rev Gastroenterol Hepatol. 2019 Jul;16(7):411-428. [PMID: 31028350 DOI: 10.1038/s41575-019-0145-7]

10 Zhu JZ, Dai YN, Wang YM, Zhou QY, Yu CH, Li YM. Prevalence of Nonalcoholic Fatty Liver Disease and Economy. Dig Dis Sci. 2015 Nov;60(11):3194-202. doi: 10.1007/s10620-015-3728-3. [PMID: 26017679 DOI: 10.1007/s10620-015-3728-3]

11 Issa D, Patel V, Sanyal AJ. Future therapy for non-alcoholic fatty liver disease. Liver Int. 2018 Feb;38 Suppl 1:56-63. [PMID: 29427492 DOI: 10.1111/liv.13676]

12 Petroni ML, Brodosi L, Bugianesi E, Marchesini G. Management of non-alcoholic fatty liver disease. BMJ. 2021;372:m4747. Published 2021 Jan 18. [PMID: 33461969 DOI :10.1136/bmj.m4747]

13 Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A et. al. PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015 Jan 1;4(1):1. [PMID: 25554246 DOI: 10.1186/2046-4053-4-1]

14 Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, et. al. AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ. 2010 Dec 14;182(18):E839-42. [PMID: 20603348 DOI: 10.1503/cmaj.090449]

15 Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, et. al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology. 2018 Jan;67(1):328-357. [PMID: 28714183 DOI: 10.1002/hep.29367]

16 National Institute for Health and Care Excellence (UK). Non-Alcoholic Fatty Liver Disease: Assessment and Management. Available from: https://www.nice.org.uk/guidance/ng49 [Last accessed on 30 Aug 2020]

17 Tomasiewicz K, Flisiak R, Halota W, Jaroszewicz J, Lebensztejn D, et. al. Recommendations for the management of non-alcoholic fatty liver disease (NAFLD). Clin Exp Hepatol. 2018 Sep;4(3):153-157. [PMID: 30324139 DOI: 10.5114/ceh.2018.78118]

18 Aller R, Fernández-Rodríguez C, Lo Iacono O, Bañares R, Abad J, Carrión JA, et. al. Consensus document. Management of non-alcoholic fatty liver disease (NAFLD). Clinical practice guideline. Gastroenterol Hepatol. 2018 May;41(5):328-349. [DOI: 10.1016/j.gastrohep.2017.12.003]

19 Loria P, Adinolfi LE, Bellentani S, Bugianesi E, Grieco A, et. al. NAFLD Expert Committee of the Associazione Italiana per lo studio del Fegato. Practice guidelines for the diagnosis and management of nonalcoholic fatty liver disease. A decalogue from the Italian Association for the Study of the Liver (AISF) Expert Committee. Dig Liver Dis. 2010 Apr;42(4):272-82. [PMID: 20171943 DOI: 10.1016/j.dld.2010.01.021]

20 Wong VW, Chan WK, Chitturi S, Chawla Y, Dan YY, et. al. Asia-Pacific Working Party on Non-alcoholic Fatty Liver Disease guidelines 2017-Part 1: Definition, risk factors and assessment. J Gastroenterol Hepatol. 2018 Jan;33(1):70-85. [PMID: 28670712 DOI: 10.1111/jgh.13857]

21 Fan JG, Wei L, Zhuang H. National Workshop on Fatty Liver and Alcoholic Liver Disease, Chinese Society of Hepatology, Chinese Medical Association; Fatty Liver Disease Expert Committee, Chinese Medical Doctor Association. Guidelines of prevention and treatment of nonalcoholic fatty liver disease (2018, China). J Dig Dis. 2019 Apr;20(4):163-173. [PMID: 30444584 DOI: 10.1111/1751-2980.12685]

22 Korean Association for the Study of the Liver (KASL). KASL clinical practice guidelines: management of nonalcoholic fatty liver disease. Clin Mol Hepatol. 2013 Dec;19(4):325-48. [PMID: 24459637 DOI: 10.3350/cmh.2013.19.4.325]

23 Watanabe S, Hashimoto E, Ikejima K, Uto H, Ono M, et. al. Japanese Society of Gastroenterology; Japan Society of Hepatology. Evidence-based clinical practice guidelines for nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. J Gastroenterol. 2015 Apr;50(4):364-77. [PMID: 25708290 DOI: 10.1007/s00535-015-1050-7]

24 Gao X, Fan JG; Study Group of Liver and Metabolism, Chinese Society of Endocrinology. Diagnosis and management of non-alcoholic fatty liver disease and related metabolic disorders: consensus statement from the Study Group of Liver and Metabolism, Chinese Society of Endocrinology. J Diabetes. 2013 Dec;5(4):406-15. [PMID: 23560695 DOI: 10.1111/1753-0407.12056]

25 Vos MB, Abrams SH, Barlow SE, Caprio S, Daniels SR, et. al. NASPGHAN Clinical Practice Guideline for the Diagnosis and Treatment of Nonalcoholic Fatty Liver Disease in Children: Recommendations from the Expert Committee on NAFLD (ECON) and the North American Society

of Pediatric Gastroenterology, Hepatology and Nutrition (NASPGHAN). J Pediatr Gastroenterol Nutr. 2017 Feb;64(2):319-334. [PMID: 28107283 DOI: 10.1097/MPG.0000000000001482]

26 Arab JP, Candia R, Zapata R, Muñoz C, Arancibia JP, et. al. Management of nonalcoholic fatty liver disease: an evidence-based clinical practice review. World J Gastroenterol. 2014 Sep 14;20(34):12182-201. [PMID: 25232252 DOI: 10.3748/wjg.v20.i34.12182]

27 LaBrecque DR, Abbas Z, Anania F, Ferenci P, Khan AG, et. al. World Gastroenterology Organisation. World Gastroenterology Organisation global guidelines: Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. J Clin Gastroenterol. 2014 Jul;48(6):467-73. [PMID: 24921212 DOI: 10.1097/MCG.0000000000000116]

28 Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, et. al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol. 2006 Nov 2;6:33. [PMID: 17081293 DOI: 10.1186/1471-230X-6-33]

29 Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, et. al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology. 2007 Apr;45(4):846-54. [PMID: 17393509 DOI: 10.1002/hep.21496]

30 Leoni S, Tovoli F, Napoli L, Serio I, Ferri S, Bolondi L. Current guidelines for the management of non-alcoholic fatty liver disease: A systematic review with comparative analysis. World J Gastroenterol. 2018;24(30):3361-3373. [PMID: 30122876 DOI: 10.3748/wjg.v24.i30.3361]

31 Paik JM, Golabi P, Biswas R, Alqahtani S, Venkatesan C, Younossi ZM. Nonalcoholic Fatty Liver Disease and Alcoholic Liver Disease are Major Drivers of Liver Mortality in the United States. Hepatol Commun. 2020 Apr 4;4(6):890-903. [PMID: 32490324 DOI: 10.1002/hep4.1510]

32 Paik JM, Henry L, De Avila L, Younossi E, Racila A, Younossi ZM. Mortality Related to Nonalcoholic Fatty Liver Disease Is Increasing in the United States. Hepatol Commun. 2019 Aug 14;3(11):1459-1471. [PMID: 31701070 DOI: 10.1002/hep4.1419]

33 Edenvik P, Davidsdottir L, Oksanen A, Isaksson B, Hultcrantz R, Stål P. Application of hepatocellular carcinoma surveillance in a European setting. What can we learn from clinical practice? Liver Int. 2015 Jul;35(7):1862-71. [PMID: 25524812 DOI: 10.1111/liv.12764]

34 Besutti G, Valenti L, Ligabue G, Bassi MC, Pattacini P, et. al. Accuracy of imaging methods for steatohepatitis diagnosis in non-alcoholic fatty liver disease patients: A systematic review. Liver Int. 2019 Aug;39(8):1521-1534. [PMID: 30972903 DOI: 10.1111/liv.14118].

35 Srivastava A, Gailer R, Tanwar S, et al. Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease. J Hepatol. 2019;71(2):371-378. [PMID: 30965069 DOI: 10.1016/j.jhep.2019.03.033]

36 Blond E, Disse E, Cuerq C, et al. EASL-EASD-EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease in severely obese people: do they lead to over-referral?. Diabetologia. 2017;60(7):1218-1222. [PMID: 28352941 DOI: 10.1007/s00125-017-4264-9]

37 Blank V, Petroff D, Beer S, et al. Current NAFLD guidelines for risk stratification in diabetic patients have poor diagnostic discrimination. Sci Rep. 2020;10(1):18345. Published 2020 Oct 27. [PMID: 33110165 DOI: 10.1038/s41598-020-75227-x]

38 Sberna AL, Bouillet B, Rouland A, et al. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD) and European Association for the Study of Obesity (EASO) clinical practice recommendations for the management of non-alcoholic fatty liver disease: evaluation of their application in people with Type 2 diabetes. Diabet Med. 2018;35(3):368-375. [PMID: 29247558 DOI:10.1111/dme.13565]

5.1.2. Accuracy of Imaging Methods for Steatohepatitis Diagnosis in Non-alcoholic Fatty Liver Disease Patients: A Systematic Review*

Summary: To assess the role of NAFLD diagnostic tools, we mapped the available methods through a scoping review, and we summarized the evidences on their accuracy through a systematic review.

Background & Aims: Non-invasive tests to diagnose non-alcoholic steatohepatitis (NASH) are urgently needed. This systematic review aims to evaluate imaging accuracy in diagnosing NASH among non-alcoholic fatty liver disease (NAFLD) patients, using liver biopsy as reference.

Methods: Eligible studies were systematic reviews and cross-sectional/cohort studies of NAFLD patients comparing imaging with histology, considering accuracy and/or associations. MEDLINE, Scopus, EMBASE, and Cochrane Library databases were searched up to April 2018. Studies were screened on title/abstract, then assessed for eligibility on full-text. Data were extracted using a pre-designed form. Risk of bias was assessed using Quality Assessment of Diagnostic Accuracy Studies-2 tool.

Results: Of the 641 studies screened, 58 were included in scoping review, 30 of which (with accuracy results) in data synthesis. Imaging techniques included: elastography (transient elastography-TE, acoustic radiation force impulse-ARFI, magnetic resonance elastography-MRE), ultrasound (US), magnetic resonance (MR), computed tomography and scintigraphy. Histological NASH definition was heterogeneous. In 28/30 studies, no prespecified threshold was used (high risk of bias). AUROCs were up to 0.82 for TE, 0.90 for ARFI, 0.93 for MRE and 0.82 for US scores. MR techniques with higher accuracy were spectroscopy (AUROC=1 for alanine), susceptibility-weighted imaging (AUROC=0.91), multiparametric MR (AUROC=0.80), optical analysis (AUROC=0.83), gadoxetic acid-enhanced MR (AUROCs=0.85) and superparamagnetic iron oxide-enhanced MR (AUROC=0.87). Results derived mostly from single studies without independent prospective validation.

Conclusions: There is currently insufficient evidence to support the use of imaging to diagnose NASH. More studies are needed on US and MR elastography and non-elastographic techniques, to date the most promising methods.

*Published in *Liver International*: Besutti G, Valenti L, Ligabue G, Bassi MC, Pattacini P, Guaraldi G, Giorgi Rossi P. Accuracy of imaging methods for steatohepatitis diagnosis in non-alcoholic fatty liver disease patients: A systematic review. Liver Int. 2019 Aug;39(8):1521-1534. doi: 10.1111/liv.14118. Epub 2019 May 8. PMID: 30972903.

Supplementary material in Appendix A

INTRODUCTION

The estimated overall global prevalence of non-alcoholic fatty liver disease (NAFLD) is around 25% and projected at 33.5% in 2030 [1]. While simple steatosis without evidence of inflammation and hepatocellular injury (non-alcoholic fatty liver) is generally a benign condition, non-alcoholic steatohepatitis (NASH) can progress to fibrosis, cirrhosis, liver failure and hepatocellular carcinoma.

Since only histological analysis can accurately evaluate NAFLD patterns, liver biopsy is the gold standard for assessment, and it should be considered in patients who are at increased risk of having steatohepatitis and/or fibrosis [2]. Major drawbacks are its invasive nature, risk of complications, sampling errors and inter and intra-observer variability [3].

Currently, there are no approved therapies for NASH. However, several drugs are now in phase 2 and 3 trials, and results are expected in 1-2 years [4]. If medical treatments become available, screening for steatohepatitis and fibrosis will be recommended in high-risk patients. The lack of non-invasive tools to identify patients who may benefit from a therapeutic intervention is a central issue. Should liver biopsy be avoided or reserved for a more limited number of undetermined or high-risk patients, the benefit-harm balance of NASH screening and therapies would undergo a major change.

Non-invasive imaging modalities such as magnetic resonance imaging (MRI) or spectroscopy (MRS) with calculation of proton density fat fraction (PDFF) accurately measure hepatic fat [5]. On the other hand, since fibrosis is the most important histological feature associated with long-term mortality in patients with NAFLD [6], research on non-invasive tests, either serum biomarkers and imaging-based techniques, have focused on this outcome [7,8].

However, the diagnosis of NASH provides important prognostic information indicating an increased risk of fibrosis progression, prompting a closer follow-up, and its resolution represents the main outcome for clinical trials [9]. Several marker panels have been proposed to differentiate between simple steatosis and NASH, with inconsistent results [10]. Some imaging methods, mostly ultrasound (US) or MR techniques, have shown promising potential in NASH diagnosis.

The objective of this systematic review is to evaluate the diagnostic accuracy of non-invasive imaging techniques in diagnosing NASH with or without fibrosis in patients with or at high risk of NAFLD, using liver biopsy as the reference standard.

MATERIALS AND METHODS

This review was conducted in two phases: 1) a scoping review aimed at mapping all the imaging tests proposed in the literature for NASH diagnosis; 2) data synthesis for those tests for which accuracy studies were available.

Study eligibility

Eligible studies were systematic reviews of studies comparing imaging and histology in the diagnosis of NASH and cross-sectional (prospective or retrospective) and cohort studies comparing one or more imaging techniques with the reference standard (liver histology). Complete protocol has been registered in the PROSPERO database (ID CRD42018089989).

Only studies that recruited patients with an available direct NAFLD assessment (biopsy- or imaging-proven) or patients at high risk of NAFLD based on metabolic factors met the inclusion criteria.

Only studies considering the following outcomes were included: diagnostic accuracy in terms of sensitivity and specificity or area under the receiving operating characteristic curve (AUROC) (main outcome), associations between index test and reference standard and reproducibility (secondary outcomes).

Since the evaluation of the presence and resolution of NASH is currently the main goal of histological assessment of liver damage in patients with NAFLD, studies focusing only on the assessment of fibrosis or steatosis, without a specific aim at differentiation between simple steatosis and NASH, were not included.

Studies reported only as abstracts or published in languages other than English were excluded.

Study search and selection

A systematic search was conducted in MEDLINE, The Cochrane Library, EMBASE and Scopus, adapting the search algorithm to the requirement of each database. No limit was applied in terms of publication date. References of included studies were reviewed to identify any additional relevant study. The last search was conducted in April 2018. The search algorithm designed for MEDLINE is reported in Supplementary Methods section.

One reviewer (GB) screened the search results based on title/abstract; a second reviewer (PGR) screened a computer-generated random sample of 20% of the references to identify potential sources of disagreement, which were resolved by consensus. Then, one reviewer (GB) examined eligibility based on the full text of the relevant articles. When unclear, inclusion was decided by group consensus. Reasons for exclusion are reported in Supplementary Table 1.

Data extraction and synthesis

One reviewer (GB) extracted data on study design, country, objective, population (number and characteristics of included patients), technical information on imaging techniques, histological classification system, outcomes, prevalence of steatohepatitis and results. These data were collected in a pre-designed data extraction sheet. A cross-check of the extracted data for accuracy was conducted by another reviewer (PGR). The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [11] was used by two reviewers (GB, PGR) to assess the risk of bias by consensus.

Summary statistics were used to describe the studies, subjects and outcomes. Data pooling would be considered only for sensitivity and specificity, and in case of sufficient homogeneity of outcomes, diagnostic techniques and procedures. Furthermore, data reporting would be necessary to allow the use of consistent positivity thresholds when needed. Otherwise, only narrative synthesis would be done. The quality of the evidence was rated with the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines [12]. Test-related consequences were considered only for those techniques with contrast media or radiotracer administration, or radiation exposure. Resource consumption in terms of human and technological resources, operator-dependence, and stage of development according to the presence of harmonized procedures and defined/agreed positivity thresholds were also taken into consideration.

RESULTS

Characteristics of the included studies

Study selection according to the PRISMA flow diagram [13] is reported in Figure 1. Sixty-one studies met eligibility criteria for scoping review; of these, 30 reported accuracy results.

Included studies were carried out from 1999 to 2018, principally in Europe, the United States and Japan. No systematic review specifically addressing imaging test for NASH diagnosis was found. Fourteen studies were retrospective, 46 were prospective, and one was described as mixed retrospective/prospective.

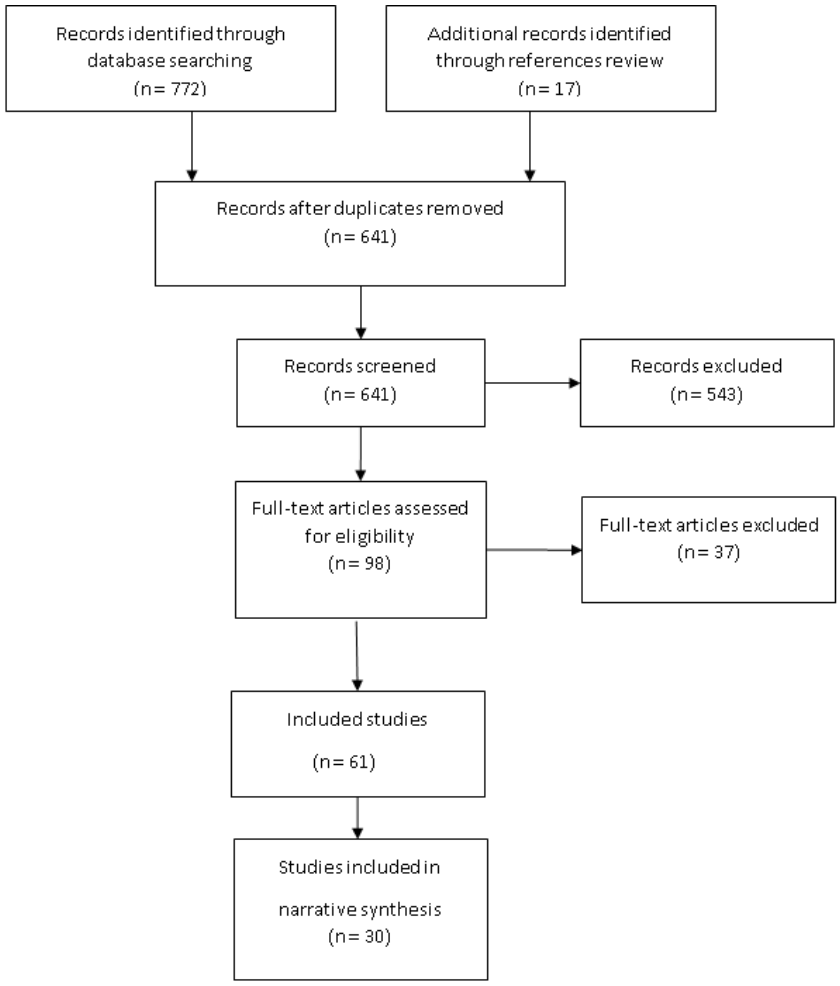


Figure 1: PRISMA 2009 Flowchart of searched, screened and included studies.

Population

All studies included patients with proven NAFLD or at high risk of NAFLD and NASH. The number of patients ranged from 8 to 513, with a total of 4693 patients included, though the number of tested patients for each technique was much smaller. Eighteen studies included a control group of healthy subjects, tested with index test but not with liver biopsy. These patients were not considered for accuracy measures in this review. Five studies also considered a subgroup of patients affected with a chronic liver disease other than NAFLD, none of these included an accuracy analysis. Two studies were specifically conducted on children or young adults, two on patients with type 2 diabetes and 4 on morbidly obese bariatric surgery patients. Most patients in the remaining studies were also overweight or obese, with mean body mass index ranging from 25 to 38.

Index Tests

Most studies compared a single imaging technique with histology; seventeen evaluated and compared more than one technique (Figure 2). Figure 3 classifies the index tests in a matrix of the types of imaging techniques and the targeted physical feature.

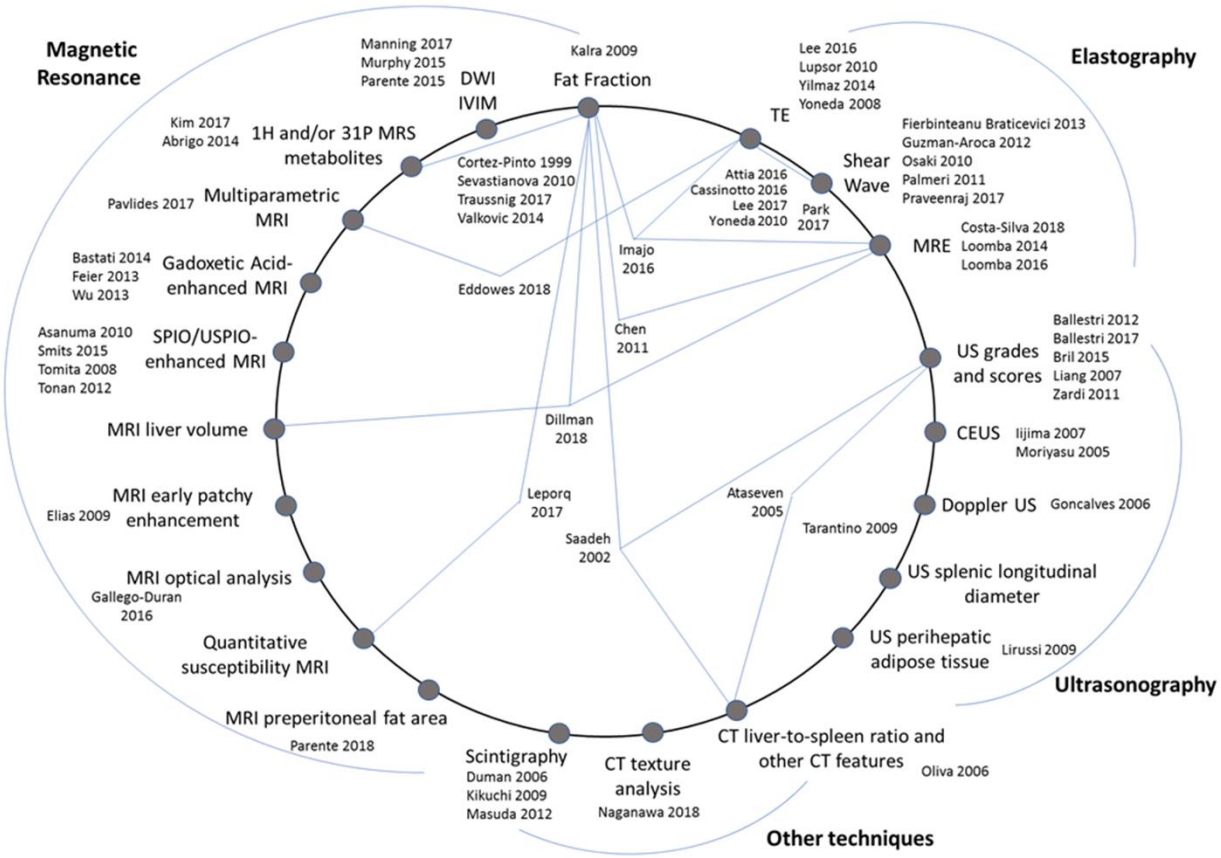


Figure 2: Diagram depicting all evaluated imaging techniques, subdivided into 4 categories (elastography, ultrasonography, magnetic resonance, other). Studies assessing a single technique are reported outside the circle, while studies reported inside the circle compared two or more techniques, linked to each study through lines. TE, transient elastography; ARFI, acoustic radiation force impulse; MRE, magnetic resonance elastography; CEUS, contrast-enhanced ultrasound; MRS, magnetic resonance spectroscopy; IVIM, intravoxel incoherent motion; DWI, diffusion-weighted imaging. Studies without accuracy results are cited in Supplementary References.

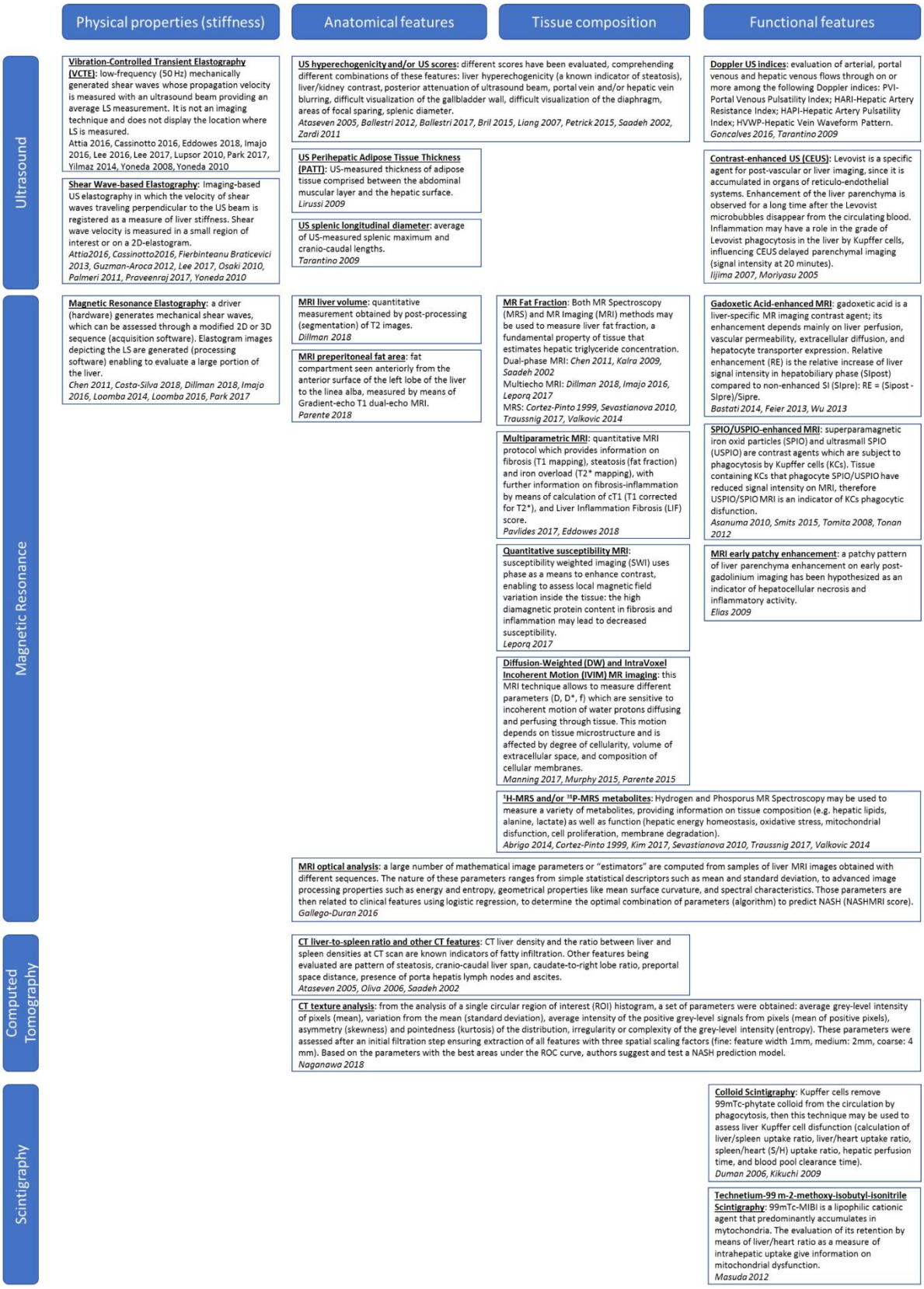


Figure 3: Classification of index tests based on the kind of feature studied (physical properties - liver stiffness, anatomical features, tissue composition or functional features). For each index test a brief explanation is reported together with the studies addressing each modality.

Histopathological analysis

Liver histology was mostly obtained through US-guided percutaneous biopsy; in 8 studies it was obtained from intra-operative biopsies or surgical specimens.

Heterogeneous histopathological definitions of NASH were used [14-17]. The accepted definition of NASH as the contemporary presence of steatosis, lobular inflammation and ballooning independently of fibrosis was generally followed, but in 6 studies fibrosis was included in the definition of NASH or classified with NASH [18-23]. Even among studies which referred to the most used classification by the Clinical Research Network [17], cases defined as borderline or with NAFLD Activity Score 3-4 were either classified with simple steatosis or with NASH. NASH prevalence ranged from 32% to 90%.

Outcomes

The main outcome (diagnostic accuracy for NASH diagnosis, i.e. differentiation between simple steatosis and NASH) was considered in 30 studies. In 4 of these, accuracy was measured in terms of AUROC, without identification of a cut-off value, while in the other 26 optimal cut-off values were reported with respective sensitivities and specificities. The remaining 31 studies reported only associations between index test and histopathological assessment (Supplementary Table 2). Reproducibility was only evaluated in a minority of the included studies (n=9).

Risk of bias analysis

Results are reported in Supplementary Figure 2 and Supplementary Tables 3 and 4. Apart from two studies which included an estimation cohort and a validation cohort [24-25], all studies were judged at high risk of bias introduced by the index test because no prespecified thresholds were used. Patient selection introduced a high risk of bias in nearly 50% of the accuracy studies.

Synthesis of accuracy results

Because of the large heterogeneity in imaging techniques and technical parameters, positivity thresholds, and NASH histopathologic definition, data pooling was not possible. In this narrative synthesis (Table 1-4), only the 30 studies reporting accuracy are considered. A more detailed description of accuracy results and a synthesis of secondary outcomes are reported in Supplementary Results section and Supplementary Tables 5-8. The level of the certainty of the evidence, according to GRADE criteria, is reported for each technique in supplementary Table 9, and results are summarized in Table 5.

Elastographic techniques

Study	Study design and index test	Population and NASH prevalence	NASH definition	Accuracy simple steatosis vs NASH
Transient Elastography (TE) – liver stiffness (LS) Coefficient attenuation parameter (CAP)				
Eddowes 2018[26]	Prospective LS	50 patients; 38(76%) with NASH, 47 with reliable TE	Steatosis, lobular inflammation and ballooning	AUROC=0.82(0.70-0.94) AUROC for NAS≥5=0.74(0.59-0.89)
Imajo 2016[18]	Prospective LS; CAP	142 patients; 108(76%) with NASH, 127 with reliable TE	Steatosis, inflammation, ballooning, and pericellular/perisinusoidal fibrosis	AUROC=0.80(0.73–0.88) † AUROC for NAS≥5=0.65(0.54-0.77) †
Lee 2016[27]	Retrospective LS; CAP	183 patients 94(51.4%) with NASH	Steatosis, inflammation and ballooning; NAS≥5	LS>7 kPa: AUROC=0.751(0.677–0.824); sensitivity=86.2%, specificity=58.4% CAP>250 dB/m: AUROC=0.743(0.669–0.816), sensitivity=96%, specificity=49% Score based on LS, CAP and ALT: AUROC=0.812(0.724–0.880)
Park 2017[5]	Prospective LS	104 patients 76(76%) with NASH ‡	NAS≥2	AUROC=0.35(0.22-0.49) Cut-off>5.6 KPa: sensitivity=61.1%, specificity=59.1%, PPV=83%, NPV=31.7%
Acoustic Radiation Force Impulse (ARFI) – shear wave velocity (SWV)				
Fierbinteanu Braticevici 2013[28]	Prospective SWV	64 patients 43(67%) with NASH	Brunt 1999/Kleiner 2005 criteria. Patients divided into simple steatosis and NASH, *borderline patients excluded*.	AUROC=0.87 Cut-off>1.10 m/s: sensitivity=77%, specificity=72%, PPV=85%, NPV=60%
Guzman-Aroca 2012[19]	Prospective SWV	32 bariatric patients 24(75%) with NASH/fibrosis (18 with inflammation and 6 with fibrosis)	Matteoni 1998 criteria. Patients categorized as simple steatosis, inflammation and fibrosis. *Comparisons between SS and NASH/fibrosis.*	NASH and/or fibrosis vs simple steatosis: AUROC=0.9 Cut-off 1.3 m/s: sensitivity=85%, specificity=83%, PPV=89%, NPV=77%
Magnetic Resonance Elastography (MRE) – liver stiffness (LS)				
Chen 2011[23] (2D MRE)	Retrospective LS	58 patients 36(72%) with NASH/fibrosis (7 inflammation and 29 fibrosis)	Brunt 1999. Patients categorized as simple steatosis, inflammation without fibrosis, and NAFLD with fibrosis, the latter two classified as NASH.	AUROC=0.93 Cut-off>2.74 KPa: sensitivity=94%, specificity=73%, PPV=85%, NPV=89% Cut-off>2.90 KPa: sensitivity=83%,

				specificity=82%, PPV=88%, NPV=75%
Costa-Silva 2018[29] (2D MRE)	Prospective LS	49 patients 25(51%) with NASH	NAS≥5	AUROC=0.79 Cut-off 3.24 Kpa: sensitivity=72%, specificity=88%, PPV=86%, NPV=72%. in fibrosis=0 patients (n=21): AUROC=0.78 Cut-off 3.22 kPa: sensitivity=69%, specificity=87%
Imajo 2016[18] (2D MRE)	Prospective LS	142 patients; 108(76%) with NASH	Steatosis, inflammation, ballooning and pericellular/perisinusoidal fibrosis	AUROC=0.81 § AUROC for NAS≥5=0.77 §
Loomba 2014[30] (2D MRE)	Prospective LS	117 patients 106(91%) with NASH	Kleiner 2005. Borderline with definite NASH.	AUROC=0.73 Cut-off 3.26 Kpa: sensitivity=42%; specificity=92%; PPV=95%; NPV=32%
Loomba 2016[31] (2D and 3D MRE)	Prospective LS	100 patients 87(87%) with NASH	Kleiner 2005. Borderline with definite NASH.	2D MRE (60 Hz): AUROC=0.75; optimal cut-off=2.92 Kpa; 3D MRE (60 Hz): AUROC=0.76; optimal cut-off=2.42 Kpa; 3D MRE (40 Hz): AUROC=0.74; optimal cut-off=1.93 KPa
Park 2017[5] (2D MRE)	Prospective LS	104 patients 76(76%) with NASH ‡	NAS≥2	AUROC=0.70 Cut-off>2.53 KPa: sensitivity=63.9%, specificity=68.2%, PPV=86.8%, NPV=36.6%

Table 1: Summary of included studies with one or more elastographic techniques as index test and diagnostic accuracy as outcome. AUROC: area under the receiver operating characteristic curve. NAS, NAFLD Activity Score. † For the combination of liver stiffness and CAP; ‡ Histological data reported for 100/104 patients; § For the combination of MRE and Proton Density Fat Fraction.

US non-elastographic techniques

Study	Study design and index test	Population and NASH prevalence	NASH definition	Accuracy simple steatosis vs NASH
US B-mode parameters and scores				
Ballestri 2012[32]	Prospective US-fatty liver indicator (US-FLI) (2-8): liver/kidney contrast (2–3), US posterior attenuation (0–1), vessel blurring (0–1), difficult visualization of gallbladder wall (0–1) or diaphragm (0–1), focal sparing (0–1)	53 patients; 35(66%) with NASH	Steatosis, lobular inflammation and ballooning; severe NASH for NAS≥ 5	AUROC=0.76 for NASH; 0.80 for severe NASH. US-FLI<4 ruled out severe NASH with NPV=94%; specificity=46%.
Liang 2007[20]	Prospective US fatty score (FS) (0–8): parenchymal echogenicity, far gain attenuation, gallbladder wall blurring, portal vein wall blurring, and hepatic vein blurring. Modified FS (MFS) (0–2): 0 for FS<7 and the sum of parenchymal echogenicity + gallbladder wall blurring <3; score 1 for FS≥7 or the latter ≥3; score 2 for FS≥7 and the latter ≥3	101 obese bariatric patients; 72(71%) with NASH	Fibrosis (≥grade 1) or acinar zone 3 hepatocellular injury with ballooning (≥grade 2)	FS: AUC=0.79; cut-off 7; sensitivity=81%; specificity=66%; accuracy=76%; PPV=85%; NPV=58% MFS: AUC=0.82; cut-off 3; sensitivity=72%; specificity=86%; accuracy=76%; PPV=93%; NPV=56%
Lirussi 2009[33]	Prospective US PATT (perihepatic adipose tissue thickness)	65 patients (33 with liver biopsy); 27(82%) with NASH	Brunt 1999. Borderline with definite NASH	Cut-off 11.8 mm: sensitivity=100%, Specificity=50%, AUROC=75%. To predict necro-inflammatory activity grading: sensitivity=80%, specificity=50%, AUROC=60%
Petrick 2015[22]	Prospective US-Fatty liver (mild, moderate, or severe according to the fall in echo amplitude, extent of liver/kidney discrepancy and of echo loss from portal vein)	513 bariatric patients 146(28%) with steatohepatitis; 164(32%) with NASH.	Brunt 1999. Steatohepatitis defined as lobular inflammation; NASH defined as steatohepatitis, fibrosis or cirrhosis	For steatohepatitis: US fatty liver (mild+): sensitivity=89%; specificity=45%; PPV=39%; NPV=91%; Accuracy=58%
Tarantino 2009[34]	Prospective Spleen longitudinal diameter	83 patients; 43(52%) with NASH	Kleiner 2005. Lobular inflammation 0-3,	AUROC=0.920

			no further specified NASH definition	Cut-off 116 mm: sensitivity=88%, specificity=95%
Zardi 2011[35]	Retrospective US score (0-6): echo amplitude attenuation (0–2), focal fat sparing (0–1), splenic diameter (0–3).	94 patients; 74(79%) with NASH	Steatosis, lobular inflammation and ballooning.	Cut-off≥5: sensitivity=74%, specificity=66%; only echo attenuation and focal fat sparing (cut-off=1): sensitivity=92%, specificity=75%.
Contrast-Enhanced Ultrasound (CEUS)				
Iijima 2007[36]	Prospective CEUS Signal intensities 5 and 20 minutes after Levovist administration	66 patients (liver biopsy in 31 patients: 21 with NASH; *in the remaining 35 NASH was clinically excluded*) + 10 healthy volunteers †	Brunt 1999. NASH for presence of parenchymatitis independently of fibrosis	Signal intensity 5 minutes Cut-off=137.8: sensitivity=100%, specificity=95%, accuracy=80%. Signal intensity 20 minutes Cut-off=43.6: sensitivity, specificity and accuracy=100%.

Table 2: Summary of included studies with one or more US non-elastography techniques as index test and diagnostic accuracy as outcome. AUROC: area under the receiving operating characteristic curve. NAS: NAFLD Activity Score. † not clear whether included in analysis.

MR non-elastographic techniques

Study	Study design and index test	Population and NASH prevalence	Definition of NASH	Accuracy simple steatosis vs NASH
^{1}H-MRS and/or ^{31}P-MRS metabolites				
Abrigo 2014[21] (*^{31}P-MRS)*	Prospective Nucleotide Triphosphate (α peak)/Triphosphate (αNTP/TP)	132 patients 95(72%) with NASH	Matteoni 1998. NASH for type 3 and 4 (fat accumulation and ballooning ± Mallory hyaline or fibrosis)	α-NTP/TP: AUROC=0.71 Cut-off≤10.57%: sensitivity=28%; specificity=91%; PPV=78%; NPV=43% Cut-off≤16.36%: sensitivity=91%; specificity=16%; PPV=65%; NPV=50%.
Kim 2017[37] (long echo time ^{1}H-MRS)	Prospective Alanine (Ala), lactate+trygliceride (Lac+TG)	26 patients; 11(42%) with NASH	NAS≥5	Ala: AUROC=1.00 Cut-off>16.04%: sensitivity=100%, specificity=100% Lac+TG: AUROC=0.78 Cut-off>360.8%: sensitivity=82%, specificity=67%
Multiparametric MRI (Liver MultiScan)- corrected T1 (cT1), Liver Inflammation and Fibrosis (LIF) score				
Eddowes 2018[26]	Prospective T1 corrected for T2* (cT1)	50 patients 38(76%) with NASH	Lobular inflammation and ballooning	AUROC for NASH vs SS=0.69 AUROC for NAS≥5 vs <5=0.74
Pavlides 2017[38]	Prospective LIF score (0-4) based on cT1 cut-offs.	71 patients 46(65%) with NASH	Steatosis, ballooning, lobular inflammation	AUROC=0.80 Cut-off 1.4: sensitivity=91%, specificity=52%
Diffusion weighted (DW) MRI and Intravoxel Incoherent Motion DW MRI (D, D, f)*				
Parente 2015[40]	Prospective Pure molecular-based (D), perfusion-related (D*), and vascular (f) Fractions	59 T2DM patients; 22(37%) with NASH	Steatosis, lobular inflammation and ballooning	-D: AUROC=0.742; cut-off 0.760: sensitivity=69% specificity=66%; -D*: AUROC=0.678; cut-off 41.45: sensitivity=68% specificity=71%; -f: AUROC=0.607; cut-off 34.23: sensitivity=49% specificity=70%.
Quantitative susceptibility MRI				
Leporq 2017[39]	Retrospective Susceptibility (ppm)	32 patients; 20(62.5%) with NASH	Steatosis, ballooning, lobular inflammation	AUROC=0.91
MRI optical analysis				
Gallego-Duran 2016[24]	Prospective NASHMRI score obtained from	126 patients (estimation cohort n=39 and	Kleiner 2005. Ballooning and inflammation.	NASHMRI score: -estimation cohort:

	most predicting estimators	validation cohort n=87); 65(51%) with NASH		AUROC=0.88. Best cut-off>0.5: sensitivity=87%, specificity=74%, PPV=80%, NPV=82% -validation cohort: AUROC=0.83. Cut-off>0.5: sensitivity=87%, specificity=60%, PPV=71% and NPV=81%.
Gadoxetic acid-enhanced MRI				
Bastati 2014[43]	Retrospective Relative Enhancement in hepatobiliary phase	81 patients; 35(43%) with NASH	NASH for activity≥2 and steatosis≥1 with any fibrosis	AUROC=0.85 Cut-off≤1.24: sensitivity=97%; specificity=63%
SPIO/USPIO-enhanced MRI				
Smits 2015[44] (USPIO)	Prospective Difference (Δ) in R2* between contrast-enhanced and baseline	24 patients (6 simple steatosis patients not biopsy-proven) 13(54%) with NASH	NAS≥5 when steatosis, inflammation and ballooning present	AUROC=0.87 Cut-off<45.5 sec^{-1}: sensitivity=77%; specificity=91%. Cut-off<58.3 sec^{-1}: sensitivity=85%; specificity=73%.
Tomita 2008[45] (SPIO)	Prospective Relative decrease in T2 (%T2) and time constant (T)	19 patients; 10(53%) with NASH	NAS≥5	T: AUROC=0.79 Cut-off=42.8: specificity=67%, sensitivity=100%, PPV=77%, NPV=100%. %T2: AUROC=0.83 Cut-off=32.5: specificity=73%, sensitivity=88%, PPV=70%, NPV=89%.
MRI Liver Volume				
Dillman 2018[41]	Retrospective Liver Volume	69 children and young adults ≤21 years old; 37(54%) with NASH	NAS≥5	AUC=0.741
MRI preperitoneal fat area				
Parente 2018[42]	Prospective Preperitoneal fat area (cm^2)	66 T2DM patients; 23(35%) with NASH	Steatosis, ballooning and lobular Inflammation	Cut-off=5: sensitivity=93%; specificity=55%

Table 3: Summary of included studies with one or more MR non-elastographic techniques as index test and diagnostic accuracy as outcome. AUROC: area under the receiving operating characteristic curve. T2DM: Type 2 Diabetes Mellitus; NAS: NAFLD Activity Score.

Other techniques

Study	Study design and index test definition	Population and NASH prevalence	Definition of NASH	Accuracy simple steatosis vs NASH
Computed Tomography (CT)				
Naganawa 2018[25]	Retrospective Non-Contrast-Enhanced CT texture features; logistic models for NASH from the most predictive features	88 patients (learning dataset=53 patients and validation dataset=35 patients). Prevalence of NASH not reported.	NAS≥3	Patients without high suspicion of fibrosis: NASH model based on mean0 and skewness2, with cut-off=0.45: AUROC=0.93 and 0.94 in learning and validation datasets; accuracy=94%, specificity=92%, sensitivity=100%, PPV=100%, NPV=80%. Patients with high suspicion of fibrosis: NASH model based on mean0 and kurtosis4, with cut-off=0.81: AUROC=0.81 and 0.60 in learning and validation datasets, accuracy=42%, specificity=31%, sensitivity=100%, PPV=100%, NPV=21%.
Liver Scintigraphy				
Kikuchi 2009[46] (Tc99m-phytate colloid scintigraphy)	Prospective Liver-to-spleen uptake ratio	37 patients; 29(78%) with definite NASH.	Kleiner 2005. Definite NASH for NAS≥5 (no patient with borderline NASH)	AUC=0.82 Cut-off value=2.93: specificity=75%, sensitivity=100%, PPV=94%, NPV=100%

Table 4: Summary of included studies with techniques other than elastography, US and MR as index test and diagnostic accuracy as outcome. AUROC: area under the receiving operating characteristic curve. NAS: NAFLD Activity Score.

	Potential harms		Resource consumption			Operator-dependence	Accuracy			Stage of development	
	Contrast media or Radiotracer	Radiation dose	Human	Technology: hardware	Technology: software		Certainty of the evidence[†]	Sensitivity[‡]	Specificity[‡]	Harmonized procedures[§]	Defined and agreed positivity threshold[§]
Transient Elastography	no	no	low	low	low	yes	very low to low	varies	low	Yes	no
US shear wave-based elastography	no	no	low	low	low	yes	low	fair	fair	Yes	no/yes
MR Elastography	no	no	high	very high	high	no	very low	varies	fair	Yes	no/yes
US non-elastographic scores and parameters	no	no	low	low	low	yes	very low	varies	varies	No	no
Contrast-enhanced US	yes	no	low	low	low	yes	very low	good	good	No	no
MR Spectroscopy	no	no	very high	high	high	no	very low	varies	varies	No	no
Multiparametric-MRI	no	no	high	high	high	no	low to very low	good	low	Yes	no
IVIM-DW-MRI	no	no	high	high	high	no	very low	varies	fair	Yes	no
Susceptibility-weighted MRI	no	no	high	high	high	no	very low	good	fair	No	no
MRI optical analysis	no	no	very high	high	very high	no	low to moderate	fair	fair	No	no
MRI morphological parameters	no	no	high	high	high	no	very low	fair	varies	No	no
Contrast-enhanced MR	yes	no	high	high	high	no	very low	fair	fair	No	no
CT texture analysis	no	yes	very high	high	very high	no	very low	good	varies	No	no
Scintigraphy	yes	yes	high	high	high	no	very low	good	fair	No	no

Table 5: Advantages and disadvantages of the techniques under evaluation in terms of potential harms, resource consumption, operator-dependence, accuracy, and stage of development. † level of evidence was classified according to GRADE criteria. ‡ According to table 1 to 4, low was assigned if the results with different procedures were substantially <=60%, fair if >60% to 90%, good >90%; otherwise we reported varies. § Procedures were considered harmonized and positivity thresholds defined and agreed on when more than one study reported on the same techniques with similar procedures and positivity thresholds (no/yes was reported when cut-off values were similar among studies, even if data pooling was not possible due to other sources of heterogeneity).

Among elastographic techniques (Table 1), the accuracy of TE was evaluated in four studies with different histopathologic definitions of NASH, showing AUROCs ranging from 0.65 (0.54-0.77) to 0.75 (0.68-0.82) for definite NASH, with sensitivity/specificity up to 86%/58% for NAS≥5 and 89%/90% for high-risk patients (NASH or fibrosis>1) [18, 26-27]. ARFI was evaluated in two studies, both with high risk of bias, resulting in sensitivities of 77%-85% and specificities of 72%-83%, using similar cut-off values [19, 28]. MRE was evaluated for NASH diagnosis in six studies, again with different NASH definitions, resulting in AUROCs ranging from 0.70 to 0.79 in studies not including fibrosis in NASH definition [5,18,29-31], with sensitivity and specificity of 72% and 87% for NAS≥5 and similar results in a subset of patients without fibrosis [29].
US non-elastographic techniques (Table 2) include several parameters and scores that took into consideration features related to the severity of steatosis, spleen diameter or visceral adiposity, all evaluated in one single study [20,22,32-35], resulting in AUROCs ranging from 0.76 of US-fatty liver indicator (US-FLI) for NAS≥2 to 0.92 of splenic diameter. With a cut-off of 4, US-FLI presented 100% sensitivity and 46% specificity for the diagnosis of severe NASH [32]. The accuracy of contrast-enhanced US for NASH diagnosis was evaluated in one single study limited by partial verification, with sensitivity and specificity up to 100% [36].
Among MR non-elastographic techniques (Table 3), the ^{31}P-MRS-derived ratio between nucleotide triphosphates (α-peak) and triphosphates (αNTP/TP), reflecting cellular energetic failure [21], and the concentration of specific metabolites (e.g. alanine, lactate, triglycerides) assessed by ^{1}H-MRS [37], showed AUROCs ranging from 0.71 for αNTP/TP and 1.00 for alanine, the latter evaluated in a small sample of 26 patients for NAS ≥5. Multiparametric MRI (mpMRI) demonstrated AUROCs of 0.69, 0.74 and 0.80, respectively, in the differentiation between NASH and SS when considering corrected T1 (cT1) as index test, in the differentiation between NAS<5 and ≥5 for the same index test, and in the diagnosis of NASH by using Liver Inflammation and Fibrosis (LIF) score [26,38]. An optimal cut-off for LIF has recently been identified (1.4), with sensitivity 91% and specificity

52%. For cT1 as well, an optimal cut-off (875 ms) has been suggested, but to distinguish between low- and high-risk (NASH or fibrosis>1) patients, with sensitivity/specificity of 97%/50%. Other MRI approaches include quantitative susceptibility imaging [39], intravoxel incoherent motion (IVIM) diffusion-weighted MRI [40], and morphological evaluation such as liver volume [41] and preperitoneal fat area [42], all evaluated in one single study, with AUROCs ranging from 0.61/0.68/0.74 for different IVIM parameters to 0.91 for susceptibility, the last one tested in a small sample of 32 patients. Moreover, a score based on MRI optical analysis estimators produced an AUROC of 0.83 with sensitivity/specificity of 87%/60% [24]. Concerning contrast media-based approaches, gadoxetic acid enhancement in hepatobiliary phase showed sensitivity/specificity of 97%/63% in a retrospective study of 81 patients [43], while superparamagnetic iron oxide (SPIO) and ultrasmall SPIO (USPIO)-enhanced MRI-derived ΔR2* demonstrated sensitivity/specificity up to 91%/73% for USPIO in a study of 25 patients for NAS≥5 [44-45].

Among other techniques (Table 4), CT texture features and TC99m-phytate colloid scintigraphy were assessed in small series (n=35 and 37 patients), resulting in AUROCs up to 0.94 and 0.82, respectively [25,46].

The presence of direct consequence of the test on the health, the qualitative analysis of resource consumption, operator-dependence, and the state of the art of the techniques are reported in Table 5.

DISCUSSION

We found more than 40 different tests proposed for non-invasive diagnosis of NASH. Tests were based on at least four different principles, including quantification of liver stiffness, anatomical features, tissue composition and functional features, combined with four imaging modalities: ultrasound, MR, CT and scintigraphy. Several authors proposed scores based on combinations of different characteristics usually collected through the same imaging approach. This landscape produced an enormous quantity of possible tests, each one proposed by one or few groups of researchers but lacking robust and independent validation. Although the first study retrieved was from 1999, indicating almost 20 years of research in the field, the picture remains that of an early stage of development of the putative technologies. Indeed, when more than one study was present, procedures and positivity thresholds were not uniform, and pooling of results was not possible. Another sign of this early phase of development is that when positivity thresholds were defined, they were usually established *a posteriori*, without confirmatory follow-up studies.

The scarce clinical utility for making a precise diagnosis of NASH in the absence of a clear practical consequence (e.g. access to treatment) most likely limited the research on non-invasive tests at an academic level. On the other hand, recent guidelines recommend having a histological

diagnosis of NASH [2,9]. Indeed, resolution of NASH is presently considered a major endpoint in clinical trials, which will hopefully soon lead to the approval of the first NASH therapies [9], providing a strong rationale for the non-invasive assessment of this condition.

This new perspective demands that research on non-invasive tests for diagnosis of NASH enter a new phase, starting from those tests which have emerged as promising thanks to their initial accuracy, are based on feasible techniques and have no or minimal direct harms of testing.

Even if a feasibility analysis of the different techniques is beyond the scope of this review, some issues are self-evident: work load and costs are higher for MR than for US, and techniques which require contrast media administration or complex post-processing, for example MRS, have additional costs. As for direct harms, they may include radiation dose (CT, scintigraphy) and contrast media administration (gadoxetic acid and SPIO/USPIO). Other techniques are substantially free of direct harms.

Based on accuracy data, the most promising tests among techniques which are relatively feasible and harmless are US and MR imaging, including both elastography (shear wave-based elastography, MRE) and non-elastographic techniques (some US scores, multiparametric MRI, susceptibility-weighted imaging), which can possibly be combined. Their combination with circulating biomarkers may also provide an added value in terms of accuracy, and research is also very active in this field [47]. A clinically applicable diagnostic algorithm will probably comprise scoring system and circulating biomarkers to be used to select high-risk patients who could benefit from a combination of imaging tests [48].

Some of these techniques may have intrinsic limitations for NASH diagnosis. Elastographic techniques have been validated to assess fibrosis. Even if liver stiffness increase may also be due to inflammation, there is the possibility that these techniques have an acceptable accuracy in diagnosing NASH as a consequence of the strong association between the presence of NASH and fibrosis. Hence, they could have intrinsic limit in sensitivity, not identifying NASH without fibrosis. However, Costa-Silva et al. observed a similar accuracy of MRE for NASH diagnosis in patients with and without fibrosis [29].

Techniques aimed at quantifying fat accumulation have failed to reach a mature stage of validation in NASH diagnosis. Steatosis is a necessary condition for both NASH and NAFLD, but assessment of hepatic fat amount may not be sufficient to identify patients with inflammation. Similarly, US scores mostly evaluating liver hyperechogenicity [22] present high referral rates and low positive predictive values to obtain high sensitivity. Preperitoneal fat area and perihepatic adipose tissue thickness, evaluated by means of MRI and US [33,42], likewise showed high sensitivities and

relatively low specificities at the proposed thresholds. Indeed, these are not direct measures of inflammation but rather indicators of visceral adiposity.

Some limitations of this review must be acknowledged. First, the search algorithm included only some techniques specifically reported in the string. Second, the choice not to pool data from the few studies that analyzed the same technique, but with different procedures, thresholds and populations, was somewhat arbitrary.

In conclusion, several imaging techniques have been tested for accuracy in NASH diagnosis. US and MR imaging, including both elastography and non-elastographic techniques, have shown promising accuracy and have no direct harms. Their combination with circulating biomarkers may provide efficient algorithms, thereby contributing to increasing diagnostic accuracy. However, the studies were conducted in limited series of patients, with different clinical features and selection criteria, using various NASH definitions and lacking independent validation. The picture of this early stage of development underlines the need for large collaborative multicenter studies with prospective design and clear definitions of outcomes, which would allow a direct comparison of the most promising imaging and biomarker approaches for NASH diagnosis.

REFERENCES

1. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018;67(1):123–33. doi:10.1002/hep.29466
2. Chalasani N, Younossi Z, Lavine JE, et al. The Diagnosis and Management of Nonalcoholic Fatty Liver Disease : Practice Guidance From the American Association for the Study of Liver Diseases. Hepatology. 2018;67(1):328–57. doi: 10.1002/hep.29367
3. Sumida Y, Nakajima A, Itoh Y. Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. World J Gastroenterol. 2014;20(2):475–85. doi: 10.3748/wjg.v20.i2.475
4. Issa D, Patel V, Sanyal A. Future therapy for non-alcoholic fatty liver disease. Liver Int. 2018;38(Suppl 1):56–63. doi: 10.1111/liv.13676
5. Park C, Nguyen P, Hernandez C, et al. Magnetic Resonance Elastography vs Transient Elastography in Detection of Fibrosis and Noninvasive Measurement of Steatosis in Patients with Biopsy-proven Nonalcoholic Fatty Liver Disease. Gastroenterology. 2017;152(3):598–607. doi: 10.1053/j.gastro.2016.10.026
6. Dulai P, Singh S, Patel J, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. Hepatology. 2017;65(5):1557–65.

doi: 10.1002/hep.29085

7. Crossan C, Tsochatzis E, Longworth L, et al. Cost-effectiveness of non-invasive methods for assessment and monitoring of liver fibrosis and cirr hosis in patients with chronic liver disease: systematic review and economic evaluation. Heal Technol Assess. 2015;19(9):1–409. doi: 10.3310/hta19090
8. Xiao G, Zhu S, Xiao X, Yan L, Yang J, Wu G. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: A meta-analysis. Hepatology. 2017;66(5):1486–501. doi: 10.1002/hep.29302
9. European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease.. J Hepatol. 2016;64(6):1388–402. doi: 10.1016/j.jhep.2015.11.004
10. Bedossa P, Patel K. Biopsy and Noninvasive Methods to Assess Progression of Nonalcoholic Fatty Liver Disease. Gastroenterology. 2016;150(8):1881–1822. doi: 10.1053/j.gastro.2016.03.008.
11. Whiting P, Rutjes A, Westwood M,et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36. doi: 10.7326/0003-4819-155-8-201110180-00009
12. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011 Apr;64(4):380-2. doi: 10.1016/j.jclinepi.2010.09.011.
13. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Med. 2009;6(7):e1000097. doi: 10.1371/journal.pmed.1000097
14. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. Gastroenterology. 1999;116:1413-1419.
15. Brunt EM1, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic Steatohepatitis : A Proposal for Grading and Staging the Histological Lesions. Am J Gastroenterol. 1999;94(9):2467–74.
16. Bedossa P, Poitou C, Veyrie N, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. Hepatology. 2012;56:1751-1759. doi: 10.1002/hep.25889
17. Kleiner DE, Brunt EM, Van Natta M, et al. Design and Validation of a Histological Scoring

System for Nonalcoholic Fatty Liver Disease. Hepatology. 2005;41(6):1313–21.

18. Imajo K, Kessoku T, Honda Y, et al. Magnetic Resonance Imaging More Accurately Classifies Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease Than Transient Elastography. Gastroenterology. 2016;150:626–37. doi: 10.1053/j.gastro.2015.11.048
19. Guzman-Aroca F, Frutos-Bernal M, Bas A, et al. Detection of non-alcoholic steatohepatitis in patients with morbid obesity before bariatric surgery: preliminary evaluation with acoustic radiation force impulse imaging. Eur Radiol. 2012;22(11):2525–32. doi: 10.1007/s00330-012-2505-3
20. Liang RJ, Wang HH, Lee WJ, Liew PL, Lin JT, Wu MS. Diagnostic Value of Ultrasonographic Examination for Nonalcoholic Steatohepatitis in Morbidly Obese Patients Undergoing Laparoscopic Bariatric Surgery. Obes Surg. 2007;17:45–56.
21. Abrigo JM, Shen J, Wong VW, et al. Non-alcoholic fatty liver disease : Spectral patterns observed from an in vivo phosphorus magnetic resonance spectroscopy study. J Hepatol. 2014;60:809–15. doi: 10.1016/j.jhep.2013.11.018
22. Petrick A, Benotti P, Wood G, et al. Utility of Ultrasound, Transaminases, and Visual Inspection to Assess Nonalcoholic Fatty Liver Disease in Bariatric Surgery Patients. Obes Surg. 2015;25(12):2368–75. doi: 10.1007/s11695-015-1707-6
23. Chen J, Talwalkar JA, Yin M, Glaser KJ, Sanderson SO, Ehman RL.. Early detection of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease by using MR elastography. Radiology. 2011;259(3):749–56. doi: 10.1148/radiol.11101942
24. Gallego-Durán R, Cerro-Salido P, Gomez-Gonzalez E, et al. Imaging biomarkers for steatohepatitis and fibrosis detection in non-alcoholic fatty liver disease. Sci Rep. 2016;6:31421. doi: 10.1038/srep31421
25. Naganawa S, Enooku K, Tateishi R, et al. Imaging prediction of nonalcoholic steatohepatitis using computed tomography texture analysis. Eur Radiol. 2018;28(7):3050–8. doi: 10.1007/s00330-017-5270-5
26. Eddowes P, McDonald N, Davies N, et al. Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. Aliment Pharmacol Ther. 2018;47:631–44. doi: 10.1111/apt.14469
27. Lee HW, Park SY, Kim SU, et al. Discrimination of Nonalcoholic Steatohepatitis Using Transient Elastography in Patients with Nonalcoholic Fatty Liver Disease. PLoS One. 2016;11(6):e0157358. doi: 10.1371/journal.pone.0157358
28. Fierbinteanu Braticevici C, Sporea I, Panaitescu E, Tribus L. Value Of Acoustic Radiation

Force Impulse Imaging Elastography For Non-Invasive Evaluation Of Patients With Nonalcoholic Fatty Liver Disease. Ultrasound Med Biol. 2013;39(11):1942–50. doi: 10.1016/j.ultrasmedbio.2013.04.019

29. Costa-Silva L, Ferolla SM, Lima AS, Vidigal PVT, Ferrari TCA. MR elastography is effective for the non-invasive evaluation of fibrosis and necroinflammatory activity in patients with nonalcoholic fatty liver disease. Eur J Radiol. 2018;98:82–9. doi: 10.1016/j.ejrad.2017.11.003
30. Loomba R, Wolfson T, Ang B, et al. Magnetic Resonance Elastography Predicts Advanced Fibrosis in Patients With Nonalcoholic Fatty Liver Disease: A Prospective Study. Hepatology. 2014;60:1920–1928. doi: 10.1002/hep.27362
31. Loomba R, Cui J, Wolfson T, et al. Novel 3D magnetic resonance elastography for the noninvasive diagnosis of advanced fibrosis in NAFLD: A prospective study. Am J Gastroenterol. 2016;111(7):986–94. doi: 10.1038/ajg.2016.65
32. Ballestri S, Lonardo A, Romagnoli D, et al. Ultrasonographic fatty liver indicator , a novel score which rules out NASH and is correlated with metabolic parameters in NAFLD. Liver Int. 2012;32(8):1242–52. doi: 10.1111/j.1478-3231.2012.02804.x
33. Lirussi F, Vitturi N, Azzalini L, et al. Perihepatic Adipose Tissue Thickness: a New Non-Invasive Marker of NAFLD? J Gastrointestin Liver Dis. 2009;18(1):61–6.
34. Tarantino G, Conca P, Pasanisi F, et al. Could inflammatory markers help diagnose nonalcoholic steatohepatitis? Eur J Gastroenterol Hepatol. 2009;21(5):504–11. doi: 10.1097/MEG.0b013e3283229b40
35. Zardi EM, De Sio I, Ghittoni G, et al. Which Clinical and Sonographic Parameters May Be Useful to Discriminate NASH from Steatosis? J Clin Gastroenterol. 2011;45:59–63. doi: 10.1097/MCG.0b013e3181dc25e3.
36. Iijima H, Moriyasu F, Tsuchiya K, et al. Decrease in accumulation of ultrasound contrast microbubbles in non-alcoholic steatohepatitis. Hepatol Res. 2007;37:722–30.
37. Kim T, Jun HY, Kim K, et al. Hepatic Alanine Differentiates Nonalcoholic Steatohepatitis From Simple Steatosis in Humans and Mice: A Proton MR Spectroscopy Study With Long Echo Time. J Magn Reson Imaging. 2017;46:1298–310. doi: 10.1002/jmri.25673.
38. Pavlides M, Banerjee R, Tunnicliffe EM, et al. Multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease severity. Liver Int. 2017;37:1065–73. doi: 10.1111/liv.13284
39. Leporq B, Lambert SA, Ronot M, Vilgrain V, Van Beers BE. Simultaneous MR quantification of hepatic fat content, fatty acid composition, transverse relaxation time and

magnetic susceptibility for the diagnosis of non-alcoholic steatohepatitis. NMR Biomed. 2017; 30(10). doi: 10.1002/nbm.3766

40. Parente DB, Paiva FF, Neto JA, et al. Intravoxel Incoherent Motion Diffusion Weighted MR Imaging at 3.0 T : Assessment of Steatohepatitis and Fibrosis Compared with Liver Biopsy in Type 2 Diabetic Patients. PLoS One. 2015;10(5):e0125653. doi: 10.1371/journal.pone.0125653
41. Dillman JR, Trout AT, Costello EN, et al. Quantitative Liver MRI-Biopsy Correlation in Pediatric and Young Adult with Nonalchohlic Fatty Liver Disease: Can One Be Used to Predict the Other? AJR. 2018;210:166–74. doi: 10.2214/AJR.17.18446
42. Parente D, Oliveira Neto J, Brasil P, et al. Preperitoneal fat as a non-invasive marker of increased risk of severe non-alcoholic fatty liver disease in patients with type 2 diabetes. J Gastroenterol Hepatol. 2018;33:511–7. doi: 10.1111/jgh.13903
43. Bastati N, Feier D, Wibmer A, et al. Noninvasive Differentiation of Simple Steatosis and Steatohepatitis by Using Gadoxetic Acid–enhanced MR Imaging in Patients with Disease: A Proof-of-Concept Study. Radiology. 2014;271(3):739–47. doi: 10.1148/radiol.14131890
44. Smits L, Coolen B, Panno M, et al. Noninvasive Differentiation between Hepatic Steatosis and Steatohepatitis with MR Imaging Enhanced with USPIOs in Patients with Nonalcoholic Fatty Liver Disease: A Proof-of-Concept Study. Radiology. 2016;278(3):782–91. doi: 10.1148/radiol.2015150952
45. Tomita K, Tanimoto A, Irie R, et al. Evaluating the Severity of Nonalcoholic Steatohepatitis With Superparamagnetic Iron Oxide-Enhanced Magnetic Resonance Imaging. J Magn Reson Imaging. 2008;28:1444–50. doi: 10.1002/jmri.21596
46. Kikuchi M, Tomita K, Nakahara T, et al. Utility of quantitative Tc-phytate scintigraphy to diagnose early-stage non-alcoholic steatohepatitis. Scand J Gastroenterol. 2009;44:229–36. doi: 10.1080/00365520802433249.
47. Wong VW, Adams LA, de Lédinghen V, Wong GL, Sookoian S. Noninvasive biomarkers in NAFLD and NASH - current progress and future promise. Nat Rev Gastroenterol Hepatol. 2018;15(8):461–78. doi: 10.1038/s41575-018-0014-9.
48. Yoneda M, Imajo K, Nakajima A. Non-Invasive Diagnosis of Nonalcoholic Fatty Liver Disease. Am J Gastroenterol. 2018 May 1:1409-1411. doi: 10.1038/s41395-018-0170-0.

5.1.3. Application of guidelines for the management of nonalcoholic fatty liver disease in three prospective cohorts of HIV-monoinfected patients*

Summary: Since the actual impact of the application of NAFLD guidelines in high-risk patients in clinical practice is debated, we evaluated the application of NAFLD guidelines in HIV-positive patients, who are at increased risk for advanced NAFLD.

Objectives: Current guidelines recommend use of a diagnostic algorithm to assess disease severity in cases of suspected nonalcoholic fatty liver disease (NAFLD). We applied this algorithm to HIV-monoinfected patients.

Methods: We analysed three prospective screening programmes for NAFLD carried out in the following cohorts: the Liver Disease in HIV (LIVEHIV) cohort in Montreal, the Modena HIV Metabolic Clinic (MHMC) cohort and the Liver Pathologies in HIV in Palermo (LHivPa) cohort. In the LIVEHIV and LHivPa cohorts, NAFLD was diagnosed if the controlled attenuation parameter (CAP) was ≥ 248 dB/m; in the MHMC cohort, it was diagnosed if the liver/spleen Hounsfield unit (HU) ratio on abdominal computerized tomography scan was < 1.1. Medium/high-risk fibrosis category was defined as fibrosis-4 (FIB-4) ≥ 1.30. Patients requiring specialist referral to hepatology were defined as either having NAFLD and being in the medium/high-risk fibrosis category or having elevated alanine aminotransferase (ALT).

Results: A total of 1534 HIV-infected adults without significant alcohol intake or viral hepatitis coinfection were included in the study. Of these, 313 (20.4%) patients had the metabolic comorbidities (obesity and/or diabetes) required for entry in the diagnostic algorithm. Among these patients, 123 (39.3%) required specialist referral to hepatology, according to guidelines. A total of 1062 patients with extended metabolic comorbidities (any among obesity, diabetes, hypertension and dyslipidaemia) represented most of the cases of NAFLD (79%), elevated ALT (75.9%) and medium/high-risk fibrosis category (75.4%). When the algorithm was extended to these patients, it was found that 341 (32.1%) would require specialist referral to hepatology.

Conclusions: According to current guidelines, one in five HIV-monoinfected patients should undergo detailed assessment for NAFLD and disease severity. Moreover, one in ten should be referred to hepatology. Expansion of the algorithm to patients with any metabolic comorbidities may be considered.

***** Published in *HIV Medicine*: Sebastiani G, Cocciolillo S, Mazzola G, Malagoli A, Falutz J, Cervo A, Petta S, Pembroke T, Ghali P, Besutti G, Franconi I, Milic J, Cascio A, Guaraldi G. Application

of guidelines for the management of nonalcoholic fatty liver disease in three prospective cohorts of HIV-monoinfected patients. HIV Med. 2020 Feb;21(2):96-108. doi: 10.1111/hiv.12799.

Supplementary Materials in Appendix B

INTRODUCTION

In Western countries, people infected with the Human Immunodeficiency Virus (HIV) live longer thanks to the widespread use of antiretroviral treatment (ART) [1, 2]. In this newly aging HIV-infected population, liver cirrhosis has become a leading cause of death [3]. Besides coinfections with hepatitis C (HCV) and B (HBV) viruses, non-alcoholic fatty liver disease (NAFLD) is an emerging concern for people aging with HIV infection [3, 4]. NAFLD is the most frequent hepatic disease in Western countries. NAFLD is a liver fat accumulation exceeding 5% of hepatocytes in the absence of other causes of liver disease. Non-alcoholic steatohepatitis (NASH), that is the evolutive counterpart of NAFLD, is a progressive disease characterized by liver fibrosis leading to cirrhosis and related complications [5]. NASH has become the second indication for liver transplantation in North America and is projected to become the leading indication in the next 10-20 years [6, 7]. Furthermore, NASH constitutes the fastest rising cause of hepatocellular carcinoma (HCC), the second leading cause of cancer-related death in the world [8]. HIV-infected patients are at higher risk of NAFLD than the general population as a result of multiple cofactors, including lifelong use of ART, HIV itself, host factors and extremely prevalent dysmetabolic conditions [9, 10]. Reported prevalence of NAFLD ranges from to 13 to 65% in HIV mono-infected patients [4, 11-14]. Moreover, NASH and significant liver fibrosis are at least twice frequent in HIV mono-infected patients than the general population [15-19].

Recent guidelines from the European Association for the Study of the Liver (EASL) recommend a diagnostic algorithm in at-risk populations for NAFLD [20]. This stepwise algorithm contemplates screening for progressive liver disease, defined as presence of elevated alanine aminotransferase (ALT) or presence of NAFLD with significant liver fibrosis, by non-invasive diagnostic methods, such as serum fibrosis biomarkers and ultrasound. A similar pathway has been recommended by the European AIDS Clinical Society (EACS) [21]. Thus far, these guidelines have not been applied to a population at high risk of NAFLD as HIV mono-infected patients.

The aim of this study was to apply the EASL and EACS guidelines to screen for NAFLD and progressive liver disease in three large cohorts of well-characterized HIV mono-infected patients. We

also aimed to identify factors associated with risk of progressive liver disease requiring specialized hepatological care, according to those guidelines.

METHODS

Study design and population

We conducted a retrospective analysis of the LIVEr in HIV (LIVEHIV), Modena HIV Metabolic Clinic (MHMC) and Liver pathologies in HIV in Palermo (LHivPa) Cohorts, which are three prospectively maintained cohorts of HIV-infected individuals [15, 22, 23]. The LIVEHIV Cohort is a prospective routine screening program for NAFLD and liver fibrosis established in September 2013 at McGill University Health Centre (MUHC). Patients undergo screening for NAFLD and liver fibrosis by transient elastography (TE) with controlled attenuation parameter (CAP). The MHMC Cohort was initiated in 2004 to assess longitudinal metabolic changes among people living with HIV. Patients undergo annual comprehensive assessments in multiple domains, including metabolic and endocrinological variables, computerized tomography (CT) scan, bone mineral density, organ function, and social factors. The LHivPa Cohort was initiated in 2011 at the Infectious Diseases Outpatient Clinic of the "Paolo Giaccone" University Hospital in Palermo. Metabolic assessment through physical and biochemical parameters is conducted at least annually. Since 2017, patients undergo screening for NAFLD and liver fibrosis by TE with CAP. We included all consecutive patients with HIV infection (as documented by positive enzyme-linked immunosorbent assay [ELISA] with Western blot confirmation) aged ≥18 years with availability of clinical and biochemical parameters included in the EASL and EACS algorithm. Exclusion criteria were: (i) positivity for HCV antibody or hepatitis B surface antigen; (ii) evidence of other liver disease (autoimmune hepatitis, primary sclerosing cholangitis, primary biliary cholangitis, hemochromatosis and Wilson's disease); (iii) significant alcohol intake, defined by an Alcohol Use Disorders Identification Test (AUDIT-C) questionnaire scores ≥4 for men and ≥3 for women[24]; (iv) history of HCC; (v) liver transplantation; (vi) only for the LIVEHIV and LHivPa Cohorts, contraindications to TE examination (pregnancy, pacemaker insertion) and failure of TE examination or unreliable measurement. All participants provided informed written consent. The Research Ethics Board of the Research Institute of MUHC (study code 14-182-BMD), MHMC (study code 254/12) and the Ethics Committee of the "Paolo Giaccone" University Hospital (study code v.1.05.1.18) approved the study, which was conducted according to the Declaration of Helsinki.

Outcome measures

The purpose of this study was to apply EASL and EACS guidelines on NAFLD to the LIVEHIV, MHMC and LHivPa Cohorts in order characterize HIV mono-infected patients at risk of progressive liver disease. The decisional algorithm from those guidelines proposes that patients with metabolic risk factors first be categorized based on the presence/absence of NAFLD, then subsequently categorized based on the presence of elevated ALT and, finally, on the presence of significant liver fibrosis [20, 21]. We considered all HIV mono-infected to be at risk of NAFLD based on the data stemmed from our cohorts [14, 23] and on a recent meta-analysis reporting a prevalence of significant liver fibrosis of 22% and of NAFLD of 35% [11]. Moreover, HIV-infected individuals on ART have higher prevalence of metabolic comorbidities than the general population [9]. A meta-analysis of 10 studies reported a pooled odds ratio of 3.85 (95% confidence interval [CI], 2.93-5.07) for diabetes in ART-exposed patients [25].

The primary study outcome was prevalence and associated factors of risk of progressive liver disease, defined as ALT > upper limit of normal (ULN; 45 IU/L in the LIVEHIV Cohort; 40 IU/L in the MHMC Cohort; 41 IU/L in males and 31 IU/L in females in the LHivPa Cohort) or NAFLD with significant liver fibrosis, as per guidelines. This would prompt referral to hepatology for specialist review. NAFLD was defined as CAP ≥248 decibel per meter (dB/m) [14, 26] in the LIVEHIV and LHivPa Cohorts and as liver/spleen Hounsfield Unit (HU) ratio <1.1 on CT scan in the MHMC Cohort [27, 28]. In all cohorts, significant liver fibrosis was defined as fibrosis-4 score (FIB-4) value >2.67 [29].

Clinical and biological parameters

Clinical and biochemical data were collected within 3 months from the TE examination or the CT scan. ART drugs were classified as: nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse-transcriptase inhibitors (NNRTIs), protease inhibitors (PIs) and integrase inhibitors. HIV viral load was assessed by Roche Cobas Amplicor assay, Roche Diagnostics, Hoffmann-La Roche Ltd, lower limit of detection 50 copies/mL. The simple biomarker FIB-4 was also calculated as follows: (age [years] × AST)/(platelet count [109/L] × (ALT)1/2) [30].

Transient elastography with CAP in the LIVEHIV and LHivPa Cohorts

TE examinations were performed on a 4-hour fasting patient by two experienced operators (>500 examinations before the study) [31]. The standard M probe was used in all patients. The XL probe was used in case of failure with M probe and if body mass index (BMI) >30 Kg/m2. Examinations with no successful measurements after at least 10 attempts were deemed failures. The following

criteria were applied to define the result of TE as reliable: at least 10 validated measures and an interquartile range (IQR) <30% of the median. Patients with known risk factors for false positive measurement were excluded [31]. A cut-off value ≥10 kPa was used to define advanced liver disease requiring hepatologic surveillance for end-stage liver complications [32, 33].

Computed Tomography scan in the MHMC Cohort

CT examinations were performed with a 64-multislice CT (LightSpeed VTC; General Electric Medical System). Hepatic and splenic attenuation values were measured by noncontrast CT using circular region of interest cursors in the two organs. Measurements were manually obtained in regions of uniform parenchyma attenuation, with care taken to avoid vessels and other areas that might give spuriously increased or decreased measurements. Measurements from each point of the liver were averaged. The L:S ratio was calculated as follows: L:S ratio p average attenuation value of liver (4 points)/attenuation value of spleen [23, 27, 28].

Statistical analysis

We compared characteristics of participants by outcome status using Student's t-test or Kruskal-Wallis test for continuous variables and Pearson's χ^2 or Fisher's exact test for categorical variables. Factors associated with risk of progressive liver disease requiring specialist referral were determined using unadjusted and adjusted logistic regression models. We reported results as adjusted odds ratios (aOR) with 95% CI. Due to the cross-sectional nature of any such measured associations, we made no causal claims from these analyses. All adjusted regression models included covariates that were determined a priori to be clinically important, based on previous literature. Final models were adjusted for age, male sex, black non-Hispanic ethnicity, BMI, diabetes, hypertension, active injection drug use (IDU), duration of HIV infection, CD4 cell count, undetectable HIV viral load, exposure to PIs, total cholesterol. To establish which of the models had the best goodness-of-fit measure, the corrected Akaike information criteria (AIC) and the Bayesian information criteria (BIC) were compared among the models. A lower AIC and/or BIC indicated a better fit. All tests were two-tailed and with a significance level of α=0.05. Statistical analyses were performed using STATA 13.1 (STATA Corp. LP, College Station, Texas, USA).

RESULTS

After applying the inclusion and exclusion criteria, 1614 HIV mono-infected individuals were included (Figure 1). Table 1 reports the main demographic, clinical, biochemical and virological

characteristics of the study population by Cohort (LIVEHIV, MHMC, LHivPa). Patients in the LHivPa Cohort were more likely to have NAFLD, while those in the MHMC Cohort were more likely to have significant liver fibrosis by FIB-4.

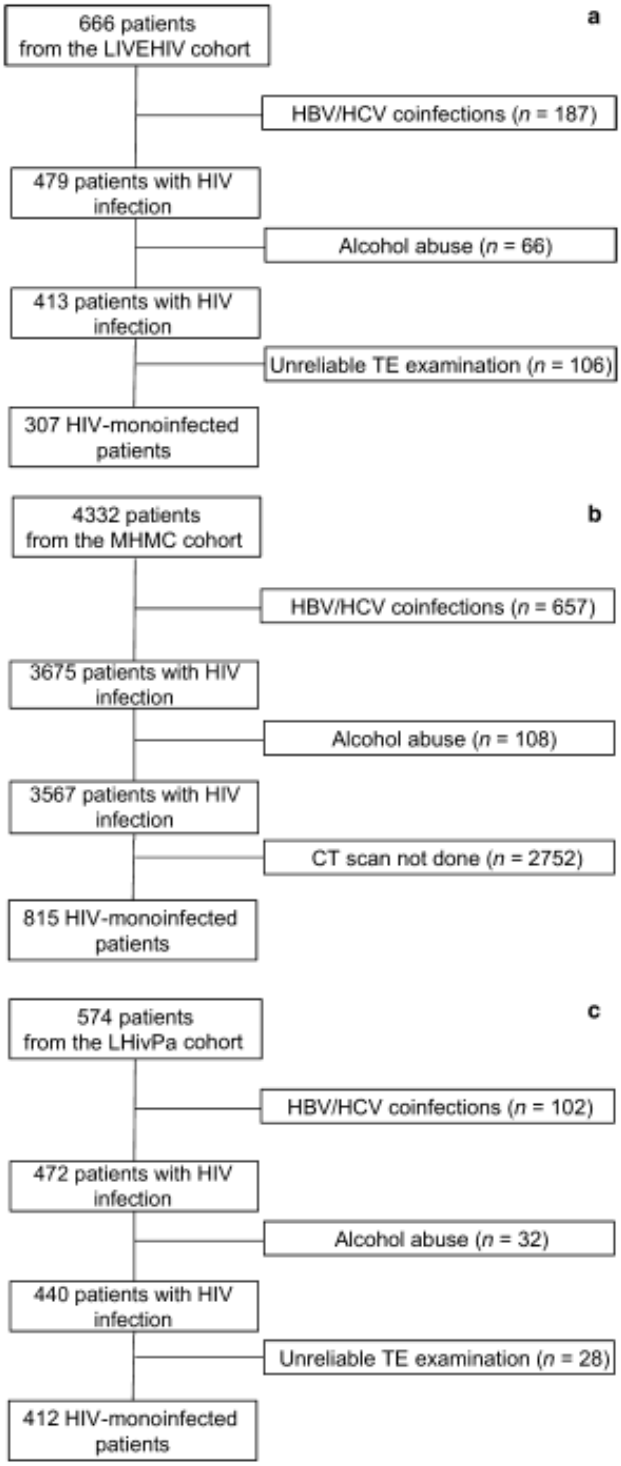


Figure 1. Flow chart displaying the selection of study participants in: (a) the LIVEHIV Cohort; (b) the MHCH Cohort; (c) the LHivPa Cohort. Liver stiffness measures by Fibroscan were considered reliable if the ratio of the IQR over the median of the 10 measures was no more than 30%.

Table 1. Characteristics of study population (n=1614).

	Whole study population (n=1614)	**LIVEHIV Cohort** (n=386)	**MHMC Cohort** (n=816)	**LHivPa Cohort** (n=412)
Age (years)	49 (43-55)	50 (42-57)*	49 (45-54)*	48 (40-55)*
Male gender (%)	1168 (72.4)	288 (74.6)	582 (71.3)	298 (72.3)
Ethnicity (%)				
White/Caucasian	1357 (84.1)	159 (41.2)**	808 (99.0)**	390 (94.7)**
Black non-Hispanic	186 (11.5)	156 (40.4)**	10 (0.1)**	20 (4.9)**
BMI (Kg/m^2)	24.1 (21.9-26.6)	25.4 (23.2-28.6)**	23.6 (21.3-25.8)**	24.2 (22.0-26.6)**
Obesity (BMI > 30) (%)	203 (12.6)	117 (30.3)**	43 (5.3)**	43 (10.4)**
Hypertension (%)	546 (33.8)	67 (17.4)**	385 (47.2)**	94 (22.8)**
Diabetes (%)	229 (14.2)	84 (21.8)*	122 (15.0)*	23 (5.6)**
Active tobacco smoker (%)	558 (34.6)	42 (10.9)**	300 (36.8)**	216 (52.4)**
MSM (%)	577 (35.8)	133 (32.2)**	262 (32.1)**	182 (44.2)**
Active IDU (%)	254 (15.7)	11 (2.9)**	226 (27.7)**	17 (4.1)**
HIV duration (years)	16.0 (8.0-21.8)	12 (7-19)**	18.8 (13.7-23.2)**	10 (4-18)**
CD4 cell count (cells/µL)	639 (453-805)	710 (480-797) **	593 (437-774) **	666 (467-874) **
Undetectable HIV viral load (≤50 copies) (%)	1273 (78.9)	194 (50.3)**	759 (93.0)**	320 (77.7)**
Current ART regimen (%)				
NRTIs	1374 (85.1)	343 (88.9)*	684 (83.8)*	347 (84.2)*
NNRTIs	712 (44.1)	222 (57.5)**	308 (37.7)**	182 (44.2)**
PIs	857 (53.1)	235 (60.9)**	463 (56.7)**	159 (38.6)**
Integrase inhibitors	579 (35.9)	128 (33.2)**	210 (25.7)**	241 (58.5)**

Platelets (10^9/L)	214 (174-257)	202 (166-244)**	206 (162-247)**	240 (201-286)**
Total cholesterol (mmol/L)	4.8 (4.1-5.5)	4.6 (3.9-5.4)**	4.9 (4.2-5.7)**	4.6 (3.9-5.3)**
LDL cholesterol (mmol/L)	2.8 (2.3-3.4)	2.7 (2.2-3.3)**	3.0 (2.4-3.6)**	2.7 (2.1-3.2)**
HDL cholesterol (mmol/L)	1.2 (1.0-1.4)	1.2 (1.0-1.4)	1.2 (1.0-1.4)	1.2 (1.0-1.4)
Triglycerides (mmol/L)	1.4 (1.0-2.2)	1.3 (0.9-2.2)**	1.6 (1.1-2.4)**	1.2 (0.8-1.8)**
ALT (IU/L)	25 (18-38)	25 (19-35)**	28 (20-44)**	21 (15-29)**
AST (IU/L)	23 (19-30)	24 (20-30)**	25 (20-34)**	19 (16-23)**
NAFLD (%)	541 (33.5)	130 (33.7)**	233 (28.6)**	178 (43.2)**
FIB-4 >2.67 (%)	107 (6.6)	19 (4.9)**	80 (9.8)**	8 (1.9)**
Risk of progressive liver disease (%)	398 (24.7)	73 (18.9)**	233 (28.6)**	92 (22.3)**

Legend: Continuous variables are expressed as median (IQR) and categorical variables as number (%). *p <0.05; **p <0.001. The p-values refer to Kruskal-Wallis test or χ^2 test among the three centers.

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; FIB-4, fibrosis-4 score; HDL, high-density lipoprotein cholesterol; HIV, human immunodeficiency virus; IDU, injection drug use; IU, international units; LDL, low-density lipoprotein cholesterol; LHivPa, Liver pathologies in HIV in Palermo; LIVEHIV, LIVEr disease in HIV; MHMC, Modena HIV Metabolic Clinic; MSM, men having sex with men; NAFLD, non-alcoholic fatty liver disease; NNRTIs, non-nucleoside reverse transcriptase inhibitors; NRTIs, nucleoside reverse transcriptase inhibitors; PIs, Protease Inhibitors.

Application of the EASL/EACS algorithm for the management of NAFLD

Figure 2a depicts the EASL/EACS decisional algorithm recommended for patients at risk of NAFLD when applied to our pooled cohort of HIV mono-infected patients. NAFLD was present in 541 patients, accounting for a prevalence of 33.5%. Of these, 163 (30.1%) had abnormal ALT and 22 (4.1%) had normal ALT with significant liver fibrosis, respectively. Taken together, 29.0% of patients diagnosed with NAFLD were at risk of progressive liver disease. Among patients without NAFLD, 213 (19.9%) had elevated ALT and were considered at risk of progressive liver disease. Overall, 398 patients were at risk of progressive liver disease, requiring specialist referral to hepatology as per EASL/EACS guidelines, accounting for a prevalence of 24.7%. Supplemental Figure S1 reports prevalence of NAFLD, elevated ALT, FIB-4>2.67 and risk of progressive liver disease according to patients' characteristics.

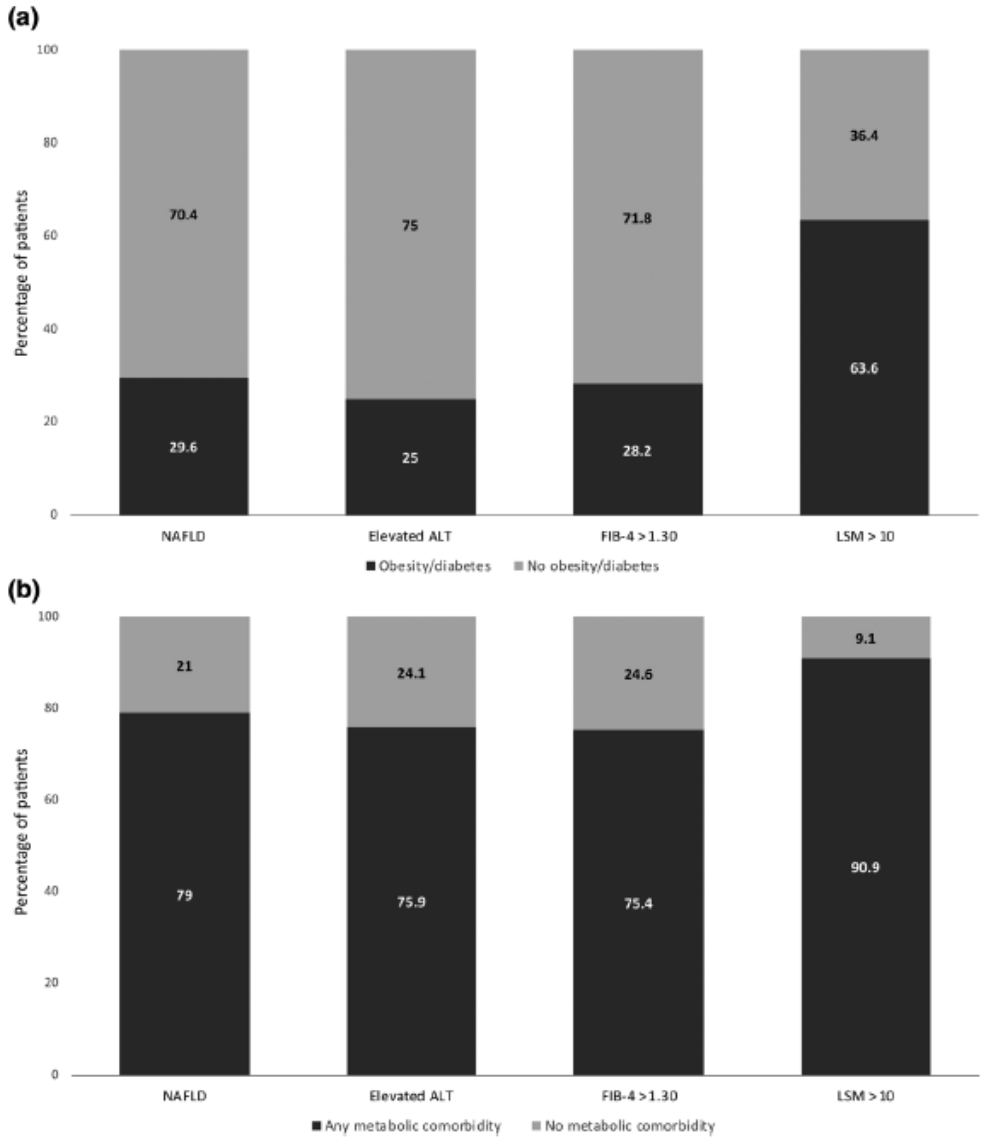


Figure 2. **a)** EASL/EACS algorithm applied to 1614 HIV mono-infected patients from the LIVEHIV, MHMC and LHivPa Cohorts and relative decisional tree. **b)** Flow chart of linkage to hepatology-specialized care of HIV mono-infected patients at risk of progressive liver disease in the LIVEHIV Cohort.

Factors associated with risk of progressive liver disease

Table 2 reports the characteristics of patients with and without risk of progressive liver disease and the related univariable analysis. In the LIVEHIV and LHivPa Cohorts, TE examination was suggestive of advanced liver disease (≥10 kPa) in 45 out of 165 (27.3%) patients at risk of progressive liver disease as compared to only 51 out of 633 (8.1%) patients not at risk (p<0.001). After adjustments, risk of progressive liver disease requiring specialist referral was independently predicted by diabetes (aOR 1.48, 95% CI 1.01-2.19; p=0.049), hypertension (aOR 1.43, 95% CI 1.03-1.98; p=0.033) and active IDU (aOR 3.04, 95% CI 2.06-4.49; p<0.001) while black ethnicity (aOR 0.35,

95% CI 0.19-0.66; p=0.001), undetectable HIV viral load (aOR 0.54, 95% CI 0.35-0.82; p=0.004) and cholesterol (aOR 0.72, 95% CI 0.62-0.83; p<0.001) were protective (Table 3). This model had lower AIC and BIC values than others, hence providing support for its use. The prevalence of risk of progressive liver disease, according to current guidelines, was 23.2% in those with neither obesity nor diabetes (Supplemental Table S1).

Table 2. Univariable analyses by outcome status, i.e. risk of progressive liver disease.

	Risk of progressive liver disease (n=398)	**No Risk of progression liver disease** (n=1216)
Age (years)	50 (45-53)	49 (43-55)
Male (%)	290 (72.9)	878 (72.2)
Ethnicity (%)		
Caucasian	346 (86.9)*	1011 (83.1)*
Black non Hispanic	28 (7.0)*	158 (13.0)*
BMI (Kg/m^2)	23.8 (21.5-26.6)	24.2 (22.0-26.6)
Obesity (BMI>30)(%)	56 (14.1)	147 (12.1)
Hypertension (%)	156 (39.2)*	390 (32.1)*
Diabetes (%)	75 (18.8)*	154 (12.7)*
Active tobacco smoker (%)	144 (36.2)	414 (34.0)
MSM (%)	106 (26.6)**	471 (38.7)**
Active IDU (%)	119 (29.9)**	135 (11.1)**
HIV duration (years)	18.7 (11.0-23.9)**	15.0 (7.4-21.0)**
CD4 cell count (cells/µL)	604 (427-784)	654 (461-810)
Undetectable HIV viral load (≤50 copies) (%)	307 (77.1)	966 (79.4)
Current ART regimen (%)		
NRTIs	348 (87.4)	1026 (84.4)
NNRTIs	159 (39.9)*	553 (45.5)*
PIs	231 (58.0)*	626 (51.5)*
Integrase inhibitors	128 (32.2)	451 (37.1)
Platelets (10^9/L)	187 (134-235)**	221 (185-262)**
Total cholesterol (mmol/L)	4.6 (3.9-5.3)**	4.8 (4.1-5.6)**
LDL cholesterol (mmol/L)	2.7 (2.0-3.2)**	2.9 (2.3-3.5)**
HDL cholesterol (mmol/L)	1.2 (1.0-1.4)	1.2 (1.0-1.4)

Triglycerides (mmol/L)	1.5 (1.0-2.3)	1.4 (1.0-2.1)
ALT (IU/L)	54 (43-71)**	22 (16-29)**
AST (IU/L)	39 (30-55)**	21 (18-25)**

Legend: Risk of progressive liver disease was defined as elevated ALT or NAFLD with significant liver fibrosis. Continuous variables are expressed as median (IQR) and categorical variables as number (%). *p <0.05; **p <0.001. The p-values refer to t test or χ^2 test between patients with and without risk of progressive liver disease.

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; HDL, high-density lipoprotein cholesterol; HIV, human immunodeficiency virus; IDU, injection drug use; IU, international units; LIVEHIV, LDL, low-density lipoprotein cholesterol; LIVEr disease in HIV; MHMC, Modena HIV Metabolic Clinic; MSM, men having sex with men; NAFLD, non-alcoholic fatty liver disease; NNRTIs, non-nucleoside reverse transcriptase inhibitors; NRTIs, nucleoside reverse transcriptase inhibitors; PIs, Protease Inhibitors.

Table 3. Multivariable analysis of factors associated with risk of progressive liver disease.

Variable	**OR (95% CI)**	**aOR (95% CI)**
Age (per 10 years)	1.04 (0.93-1.17)	0.87 (0.72-1.06)
Male sex (yes vs no)	1.03 (0.80-1.33)	1.29 (0.90-1.84)
Black ethnicity (yes vs no)	0.51 (0.33-0.78)*	0.35 (0.19-0.66)*
BMI (per Kg/m^2)	0.99 (0.97-1.02)	1.01 (0.98-1.05)
Diabetes (yes vs no)	1.60 (1.18-2.17)*	1.48 (1.01-2.19)*
Hypertension (yes vs no)	1.37 (1.08-1.73)*	1.43 (1.03-1.98)*
Active IDU (yes vs no)	3.42 (2.58-4.52)**	3.04 (2.06-4.49)**
Duration HIV infection (per 10 years)	1.57 (1.27-1.69)**	1.08 (0.85-1.38)
Undetectable HIV viral load (yes vs no)	0.87 (0.64-1.19)	0.54 (0.35-0.82)*
CD4 count (per 100 cells)	0.99 (0.99-1.00)	0.99 (0.99-1.00)
PIs exposure (yes vs no)	1.29 (1.03-1.62)*	1.28 (0.94-1.75)
Cholesterol (per mmol/L)	0.80 (0.72-0.89)**	0.72 (0.62-0.83)**

Legend: Odds ratios (OR) and 95% confidence interval are presented for each variable in the unadjusted and adjusted analysis. *p <0.05; **p <0.001. Abbreviations: aOR, adjusted odds ratio; BMI, body mass index;

HIV, human immunodeficiency virus; IDU, injection drug use; MSM, men having sex with men; PIs, protease inhibitors.

Linkage to hepatology specialized care of HIV mono-infected patients at risk of progressive liver disease

In the LIVEHIV Cohort, all the 73 patients at risk of progressive liver disease, according to EASL/EACS guidelines, were referred to the hepatologist (GS) (Figure 2b). Among these, 43 (58.9%) patients have been retained to care, with hepatologic follow-up every 6-12 months. All retained patients have also been referred for further specialized care (endocrinology and nutritional consultation). Twenty-eight patients have been enrolled in clinical trials, including vitamin E treatment or ART switch. Eight patients have also undergone a liver biopsy. The mean liver biopsy length was 18±5 mm. NAFLD was present in all patients who underwent a liver biopsy. Significant liver fibrosis (stages F2-4) and cirrhosis (F4) were present in 4 (50%) and 1 (12.5%) cases, respectively[34]. Among patients who underwent a liver biopsy, one patient was listed for liver transplantation and is awaiting a graft.

DISCUSSION

This study, based on a large pooled cohort of well characterized HIV mono-infected patients without significant alcohol intake recruited from three prospective cohorts, aimed to apply recent NAFLD guidelines in order to determine prevalence and factors associated with risk of progressive liver disease. We showed that a significant proportion of HIV mono-infected patients are at risk of progressive liver disease requiring referral to specialized care in hepatology. The main associated factors were diabetes, hypertension and active IDU, while black ethnicity, undetectable HIV viral load and higher total cholesterol were protective.

Liver disease is now the second leading cause of non-AIDS-related death among people living with HIV[2]. While co-infection with HCV is believed to have driven this trend, increasing rates of NAFLD may also have contributed. The prevalence of NAFLD in patients with HIV varies considerably among studies, ranging from 13% to 65%, likely reflecting the different tools used to diagnose NAFLD, the population sampled and the HIV treatment regimens [11]. Recent evidences show that NAFLD is not only more frequent in the setting of HIV infection, but also more severe. NASH, the evolutive counterpart of NAFLD, has been reported in up to 65% of HIV mono-infected patients with chronically elevated ALT and in 10% of those attending a routine screening program [18, 35, 36]. A

cross-sectional case-control study showed that HIV-associated NAFLD has more severe liver disease and higher rates of NASH compared to age/sex-matched HIV-negative controls [16]. Furthermore, NAFLD has high rates of progression to liver fibrosis and cirrhosis in HIV mono-infected patients [15]. The 6.6% prevalence of significant liver fibrosis we found is in line with other cohorts [11]. However, NAFLD lacks specific signs or symptoms and remain silent until end-stage complications arise. Neither can be readily diagnosed using routine blood tests, such as liver transaminases. In the general population, up to 79% of patients with NAFLD have normal ALT level [37]. Similarly, we and others found that 67% to 83% of HIV mono-infected patients with NAFLD have normal ALT [14, 38]. On the other hand, several studies showed that elevated ALT is highly suggestive of presence of NASH or liver fibrosis [35, 36, 39]. Overall, there is a need for a diagnostic algorithm for the assessment and monitoring of HIV-infected patients with existing or being at risk of NAFLD and of progressive liver disease requiring a dedicated referral for specialized care in hepatology.

The EASL/EACS guidelines recently recommended a stepwise diagnostic algorithm to screen for NAFLD and risk of progressive liver disease in patients with metabolic risk factors by non-invasive diagnostic methods. Our population was screened for NAFLD by CAP in the LIVEHIV and LHivPa Cohorts and by CT-scan in the MHMC Cohort. CAP is a TE software able to quantify fat in the liver, with a reported sensitivity of 89–91% [40, 41]. CAP has been validated and applied in HIV mono-infected patients [14, 42]. Unenhanced CT scan is increasingly used in HIV-infected patients to detect moderate to severe hepatic steatosis based on attenuation difference between liver and spleen, with sensitivity and specificity at 82% and 100% respectively [27, 28]. In our pooled cohort of 1614 patients, we found that 33.5% had NAFLD diagnosed by CAP or CT scan, a figure that is consistent with previous literature [11]. Despite the use of different diagnostic tools and the diversity of patient populations included in the pooled cohort, NAFLD was a frequent occurrence in consecutive HIV mono-infected patients.

When applying the EASL/EACS algorithm to our pooled cohort of 1614 HIV mono-infected patients, we found that 24.7% were at risk of progressive liver disease requiring referral for hepatology consultation, defined as presence of elevated ALT or NAFLD with significant liver fibrosis. This is a striking figure considering that we applied the guidelines to all consecutive patients from the three cohorts, and not only to those with metabolic risk factors. Importantly, 23.2% of patients without obesity or diabetes were at risk of progressive liver disease, according to the guidelines. Our findings suggest that there is a significant proportion of HIV mono-infected patients at risk of progressive liver disease beyond the main metabolic risk factors for NAFLD, likely due to immunologic components intrinsic to the HIV infection and to chronic use of ART. For example, HIV viral protein R induces

fatty liver in mice, thus suggesting an HIV-specific pathogenesis of NAFLD [43]. HIV also functions to decrease the number of Kupffer cells in the liver, and in doing so significantly impairs the ability of the liver to clear products of microbial translocation from the portal blood [44]. This combined immune-metabolic pathogenesis of progressive liver disease in HIV mono-infected patients is further supported by our finding that undetectable HIV viral load, a proxy for HIV infection control, was independently associated with risk of progressive liver disease.

Our multivariable analysis identified diabetes, hypertension and active IDU as independent factors associated with risk of progressive liver disease. NAFLD is known to be strongly associated with central obesity and insulin resistance states, such as type 2 diabetes and metabolic syndrome [41, 45]. Insulin resistance is frequent in HIV mono-infected patients due to aging and to specific ART [9, 46]. Increasing age, overweight, race, cumulative exposure to ART and dyslipidemia have been associated with the risk of diabetes in HIV mono-patients [46]. The Multicenter AIDS Cohort Study showed that the relative risk of incident diabetes in HIV patients on chronic ART use was four times higher than in HIV negative controls (10% vs 3% over 4 years) [47]. In the general population, hypertension is not only a major risk factor for NAFLD, but also for liver fibrosis progression, which is doubled in hypertensive NAFLD patients [48]. Hypertension is particularly frequent in people aging with HIV [49]. Active IDU was also independently associated with risk of progressive liver disease in our pooled cohort. We hypothesize that this finding could be related to a worse HIV control due to lower adherence to ART, which has been reported in the literature [50, 51]. Indeed, active IDU patients had lower CD4 cell count and longer HIV duration (data not shown). Black ethnicity was a protective factor for risk of progressive liver disease, as previously shown in both the general population and in HIV-infected patients [52-54]. Interestingly, low total cholesterol was also associated with risk of progressive liver disease. Evidence from the literature shows that cholesterol and its fractions progressively decrease with increasing severity of liver disease [55]. Indeed, patients with advanced liver disease have a failed lipid synthesis despite severity of insulin resistance [56]. Consistent with this hypothesis, we found that 27.3% of patients with risk of progressive liver disease already had advanced liver disease diagnosed by TE in the LIVEHIV and LHivPa Cohorts. In the LIVEHIV Cohort, all 73 patients found to be at risk of progressive liver disease were referred to the hepatologist (GS). Thus far, 58.9% of these patients have been retained to care for surveillance and specialized interventions, including diagnostic liver biopsy and referral for liver transplant.

Our study has several strengths. It is the first study applying the algorithm proposed by the EASL Clinical Practice Guidelines for NAFLD and the EACS guidelines in a large population of HIV mono-infected patients recruited from three prospective and diverse routine screening programs for liver

disease. Importantly, we applied the guidelines to all consecutive patients including those without metabolic risk factors, showing that the general HIV mono-infected population is at risk of progressive liver disease. Second, we adopted two easily accessible and validated non-invasive diagnostic tools for NAFLD. Third, a subsequent follow-up with appropriate linkage to a hepatologic specialized care was provided in the LIVEHIV cohort. Fourth, despite the demographic, serological and clinical differences among the three cohorts and the different diagnostic tool employed for NAFLD, we obtained similar results in terms of prevalence of liver disease by applying the guidelines.

We acknowledge several limitations of this study. First, it was conducted at tertiary-care referral centers, as such the prevalence of the outcome could be higher than primary/secondary care settings. Second, liver biopsy, the current gold-standard for NAFLD and liver fibrosis diagnosis, was unavailable for most of our patients. However, screening studies employing liver biopsy are unfeasible, ethically questionable and not recommended for routine clinical use. This is further supported by the fact that the EASL and EACS guidelines recommend non-invasive tests as screening tools. Fourth, the cross-sectional study design limits our ability to speculate on the dynamics of the disease over time and did not allow us to explore associations of specific factors, such as individual ART drugs, with the risk of progressive liver disease.

In conclusion, according to current EASL and EACS guidelines, one in four HIV mono-infected patients is at risk of progressive liver disease, which requires referral to specialized care in hepatology, even in absence of obesity or diabetes. The identified predictors, namely diabetes, hypertension, active IDU, as well as non-invasive diagnostic tests, could help target screening and interventional studies in the busy setting of HIV clinics, as well as help select patients who would benefit the most from hepatologic referral.

REFERENCES

1. Fang CT, Chang YY, Hsu HM, Twu SJ, Chen KT, Lin CC, et al. Life expectancy of patients with newly-diagnosed HIV infection in the era of highly active antiretroviral therapy. *Qjm* 2007; 100(2):97-105.

2. Smith CJ, Ryom L, Weber R, Morlat P, Pradier C, Reiss P, et al. Trends in underlying causes of death in people with HIV from 1999 to 2011 (D:A:D): a multicohort collaboration. *Lancet* 2014; 384(9939):241-248.

3. Puoti M, Moioli MC, Travi G, Rossotti R. The burden of liver disease in human immunodeficiency virus-infected patients. *Semin Liver Dis* 2012; 32(2):103-113.
4. Lemoine M, Serfaty L, Capeau J. From nonalcoholic fatty liver to nonalcoholic steatohepatitis and cirrhosis in HIV-infected patients: diagnosis and management. *Curr Opin Infect Dis* 2012; 25(1):10-16.
5. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* 2017.
6. Charlton MR, Burns JM, Pedersen RA, Watt KD, Heimbach JK, Dierkhising RA. Frequency and outcomes of liver transplantation for nonalcoholic steatohepatitis in the United States. *Gastroenterology* 2011; 141(4):1249-1253.
7. Wong RJ, Aguilar M, Cheung R, Perumpail RB, Harrison SA, Younossi ZM, et al. Nonalcoholic Steatohepatitis is the Second Leading Etiology of Liver Disease Among Adults Awaiting Liver Transplantation in the U.S. *Gastroenterology* 2014.
8. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; 61(2):69-90.
9. Guaraldi G, Lonardo A, Maia L, Palella FJ, Jr. Metabolic concerns in aging HIV-infected persons: from serum lipid phenotype to fatty liver. *AIDS* 2017; 31 Suppl 2:S147-S156.
10. Rockstroh JK, Mohr R, Behrens G, Spengler U. Liver fibrosis in HIV: which role does HIV itself, long-term drug toxicities and metabolic changes play? *Curr Opin HIV AIDS* 2014; 9(4):365-370.
11. Maurice JB, Patel A, Scott AJ, Patel K, Thursz M, Lemoine M. Prevalence and risk factors of nonalcoholic fatty liver disease in HIV-monoinfection. *AIDS* 2017; 31(11):1621-1632.
12. Sterling RK, Smith PG, Brunt EM. Hepatic steatosis in human immunodeficiency virus: a prospective study in patients without viral hepatitis, diabetes, or alcohol abuse. *J Clin Gastroenterol* 2013; 47(2):182-187.
13. Crum-Cianflone N, Dilay A, Collins G, Asher D, Campin R, Medina S, et al. Nonalcoholic fatty liver disease among HIV-infected persons. *J Acquir Immune Defic Syndr* 2009; 50(5):464-473.
14. Vuille-Lessard E, Lebouche B, Lennox L, Routy JP, Costiniuk CT, Pexos C, et al. Nonalcoholic fatty liver disease diagnosed by transient elastography with controlled attenuation parameter in unselected HIV monoinfected patients. *AIDS* 2016; 30(17):2635-2643.
15. Pembroke T, Deschenes M, Lebouche B, Benmassaoud A, Sewitch M, Ghali P, et al. Hepatic steatosis progresses faster in HIV mono-infected than HIV/HCV co-infected patients and is associated with liver fibrosis. *J Hepatol* 2017; 67(4):801-808.

16. Vodkin I, Valasek MA, Bettencourt R, Cachay E, Loomba R. Clinical, biochemical and histological differences between HIV-associated NAFLD and primary NAFLD: a case-control study. *Aliment Pharmacol Ther* 2015; 41(4):368-378.

17. Lemoine M, Lacombe K, Bastard JP, Sebire M, Fonquernie L, Valin N, et al. Metabolic syndrome and obesity are the cornerstones of liver fibrosis in HIV-monoinfected patients: results of the METAFIB study. *AIDS* 2017.

18. Benmassaoud A, Ghali P, Cox J, Wong P, Szabo J, Deschenes M, et al. Screening for nonalcoholic steatohepatitis by using cytokeratin 18 and transient elastography in HIV mono-infection. *PLoS One* 2018; 13(1):e0191985.

19. Price JC, Seaberg EC, Badri S, Witt MD, D'Acunto K, Thio CL. HIV monoinfection is associated with increased aspartate aminotransferase-to-platelet ratio index, a surrogate marker for hepatic fibrosis. *J Infect Dis* 2012; 205(6):1005-1013.

20. European Association for the Study of the L, European Association for the Study of D, European Association for the Study of O. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* 2016; 64(6):1388-1402.

21. Ryom L, Boesecke C, Bracchi M, Ambrosioni J, Pozniak A, Arribas J, et al. Highlights of the 2017 European AIDS Clinical Society (EACS) Guidelines for the treatment of adult HIV-positive persons version 9.0. *HIV Med* 2018; 19(5):309-315.

22. Guaraldi G, Orlando G, Zona S, Menozzi M, Carli F, Garlassi E, et al. Premature age-related comorbidities among HIV-infected persons compared with the general population. *Clin Infect Dis* 2011; 53(11):1120-1126.

23. Guaraldi G, Squillace N, Stentarelli C, Orlando G, D'Amico R, Ligabue G, et al. Nonalcoholic fatty liver disease in HIV-infected patients referred to a metabolic clinic: prevalence, characteristics, and predictors. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2008; 47(2):250-257.

24. Reinert DF, Allen JP. The Alcohol Use Disorders Identification Test (AUDIT): a review of recent research. *Alcohol Clin Exp Res* 2002; 26(2):272-279.

25. Nduka CU, Stranges S, Kimani PK, Sarki AM, Uthman OA. Is there sufficient evidence for a causal association between antiretroviral therapy and diabetes in HIV-infected patients? A meta-analysis. *Diabetes/metabolism research and reviews* 2017; 33(6).

26. Karlas T, Petroff D, Sasso M, Fan JG, Mi YQ, de Ledinghen V, et al. Individual patient data meta-analysis of controlled attenuation parameter (CAP) technology for assessing steatosis. *J Hepatol* 2017; 66(5):1022-1030.

27. Zeb I, Li D, Nasir K, Katz R, Larijani VN, Budoff MJ. Computed tomography scans in the evaluation of fatty liver disease in a population based study: the multi-ethnic study of atherosclerosis. *Acad Radiol* 2012; 19(7):811-818.
28. Lee DH. Imaging evaluation of non-alcoholic fatty liver disease: focused on quantification. *Clin Mol Hepatol* 2017; 23(4):290-301.
29. Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2009; 7(10):1104-1112.
30. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* 2006; 43(6):1317-1325.
31. Castera L. Noninvasive methods to assess liver disease in patients with hepatitis B or C. *Gastroenterology* 2012; 142(6):1293-1302 e1294.
32. Maurice JB, Brodkin E, Arnold F, Navaratnam A, Paine H, Khawar S, et al. Validation of the Baveno VI criteria to identify low risk cirrhotic patients not requiring endoscopic surveillance for varices. *J Hepatol* 2016; 65(5):899-905.
33. Margini C, Murgia G, Stirnimann G, De Gottardi A, Semmo N, Casu S, et al. Prognostic Significance of Controlled Attenuation Parameter in Patients With Compensated Advanced Chronic Liver Disease. *Hepatol Commun* 2018; 2(8):929-940.
34. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. *Am J Gastroenterol* 1999; 94(9):2467-2474.
35. Ingiliz P, Valantin MA, Duvivier C, Medja F, Dominguez S, Charlotte F, et al. Liver damage underlying unexplained transaminase elevation in human immunodeficiency virus-1 mono-infected patients on antiretroviral therapy. *Hepatology* 2009; 49(2):436-442.
36. Morse CG, McLaughlin M, Matthews L, Proschan M, Thomas F, Gharib AM, et al. Nonalcoholic Steatohepatitis and Hepatic Fibrosis in HIV-1-Monoinfected Adults With Elevated Aminotransferase Levels on Antiretroviral Therapy. *Clin Infect Dis* 2015; 60(10):1569-1578.
37. Browning JD, Szczepaniak LS, Dobbins R, Nuremberg P, Horton JD, Cohen JC, et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* 2004; 40(6):1387-1395.
38. Lombardi R, Sambatakou H, Mariolis I, Cokkinos D, Papatheodoridis GV, Tsochatzis EA. Prevalence and predictors of liver steatosis and fibrosis in unselected patients with HIV mono-infection. *Dig Liver Dis* 2016; 48(12):1471-1477.

39. Lombardi R, Lever R, Smith C, Marshall N, Rodger A, Bhagani S, et al. Liver test abnormalities in patients with HIV mono-infection: assessment with simple noninvasive fibrosis markers. *Ann Gastroenterol* 2017; 30(3):349-356.
40. Sasso M, Beaugrand M, de Ledinghen V, Douvin C, Marcellin P, Poupon R, et al. Controlled attenuation parameter (CAP): a novel VCTE guided ultrasonic attenuation measurement for the evaluation of hepatic steatosis: preliminary study and validation in a cohort of patients with chronic liver disease from various causes. *Ultrasound in medicine & biology* 2010; 36(11):1825-1835.
41. Dowman JK, Tomlinson JW, Newsome PN. Systematic review: the diagnosis and staging of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis. *Alimentary pharmacology & therapeutics* 2011; 33(5):525-540.
42. Macias J, Gonzalez J, Tural C, Ortega-Gonzalez E, Pulido F, Rubio R, et al. Prevalence and factors associated with liver steatosis as measured by transient elastography with controlled attenuation parameter in HIV-infected patients. *AIDS* 2014; 28(9):1279-1287.
43. Agarwal N, Iyer D, Gabbi C, Saha P, Patel SG, Mo Q, et al. HIV-1 viral protein R (Vpr) induces fatty liver in mice via LXRalpha and PPARalpha dysregulation: implications for HIV-specific pathogenesis of NAFLD. *Sci Rep* 2017; 7(1):13362.
44. Kaspar MB, Sterling RK. Mechanisms of liver disease in patients infected with HIV. *BMJ Open Gastroenterol* 2017; 4(1):e000166.
45. Ballestri S, Zona S, Targher G, Romagnoli D, Baldelli E, Nascimbeni F, et al. Nonalcoholic fatty liver disease is associated with an almost twofold increased risk of incident type 2 diabetes and metabolic syndrome. Evidence from a systematic review and meta-analysis. *J Gastroenterol Hepatol* 2016; 31(5):936-944.
46. Samaras K. The burden of diabetes and hyperlipidemia in treated HIV infection and approaches for cardiometabolic care. *Curr HIV/AIDS Rep* 2012; 9(3):206-217.
47. Brown TT, Cole SR, Li X, Kingsley LA, Palella FJ, Riddler SA, et al. Antiretroviral therapy and the prevalence and incidence of diabetes mellitus in the multicenter AIDS cohort study. *Arch Intern Med* 2005; 165(10):1179-1184.
48. Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis progression in nonalcoholic fatty liver vs nonalcoholic steatohepatitis: a systematic review and meta-analysis of paired-biopsy studies. *Clin Gastroenterol Hepatol* 2015; 13(4):643-654 e641-649; quiz e639-640.
49. d'Arminio Monforte A, Diaz-Cuervo H, De Luca A, Maggiolo F, Cingolani A, Bonora S, et al. Evolution of major non-HIV-related comorbidities in HIV-infected patients in the Italian Cohort of Individuals, Naive for Antiretrovirals (ICONA) Foundation Study cohort in the period 2004-2014. *HIV Med* 2018.

50. Shannon K, Kerr T, Lai C, Ishida T, Wood E, Montaner JS, et al. Nonadherence to antiretroviral therapy among a community with endemic rates of injection drug use. *J Int Assoc Physicians AIDS Care (Chic)* 2005; 4(3):66-72.
51. Azar P, Wood E, Nguyen P, Luma M, Montaner J, Kerr T, et al. Drug use patterns associated with risk of non-adherence to antiretroviral therapy among HIV-positive illicit drug users in a Canadian setting: a longitudinal analysis. *BMC Infect Dis* 2015; 15:193.
52. Rafiq N, Stepanova M, Lam B, Nader F, Srishord M, Younossi ZM. Predictors of chronic liver disease in individuals with human immunodeficiency virus infection. *Ann Hepatol* 2013; 13(1):60-64.
53. Sajja KC, Mohan DP, Rockey DC. Age and ethnicity in cirrhosis. *J Investig Med* 2014; 62(7):920-926.
54. Lazo M, Hernaez R, Eberhardt MS, Bonekamp S, Kamel I, Guallar E, et al. Prevalence of nonalcoholic fatty liver disease in the United States: the Third National Health and Nutrition Examination Survey, 1988-1994. *Am J Epidemiol* 2013; 178(1):38-45.
55. Chrostek L, Supronowicz L, Panasiuk A, Cylwik B, Gruszewska E, Flisiak R. The effect of the severity of liver cirrhosis on the level of lipids and lipoproteins. *Clin Exp Med* 2014; 14(4):417-421.
56. Siddiqui MS, Fuchs M, Idowu MO, Luketic VA, Boyett S, Sargeant C, et al. Severity of nonalcoholic fatty liver disease and progression to cirrhosis are associated with atherogenic lipoprotein profile. *Clin Gastroenterol Hepatol* 2015; 13(5):1000-1008 e1003.

5.1.4. Feasibility and efficiency of the application of international guidelines for NAFLD assessment in type 2 diabetes patients: a prospective study. *

Summary: Since the actual impact of the application of NAFLD guidelines in high-risk patients in clinical practice is debated, we evaluated the application of NAFLD guidelines in Type 2 diabetes patients, who are at increased risk for advanced NAFLD.

Objective: We aimed to evaluate the feasibility and efficiency of a guidelines-compliant NAFLD assessment algorithm in newly diagnosed type-2 diabetes (T2D) patients.

Methods: Consecutive newly diagnosed <75 y/o T2D patients (January 2019-January 2020) without coexistent liver disease or excessive alcohol consumption were enrolled. To referrals, patients were stratified based on liver enzymes, fatty liver index, ultrasound, fibrosis scores and liver stiffness measurement. Referral rates and positive predictive values (PPVs) for histological non-alcoholic steatohepatitis (NASH) and significant fibrosis were evaluated.

Results: Of 171 enrolled patients (age 59±10.2 years, 42.1% females), 115 (67.3%) were referred to a hepatologist due to abnormal liver enzymes (n=60) or steatosis plus indeterminate (n=37) or high (n=18) NAFLD fibrosis score. Liver biopsy was proposed to 30 (17.5%), but only 14 patients accepted. Of these, 12 had NASH, only one with significant fibrosis. The PPV of hepatological referral was 12/76 (15.8%) for NASH and 1/76 (1.3%) for NASH with significant fibrosis. The PPV of liver biopsy referral was 12/14 (85.7%) for NASH and 1/14 (7.1%) for NASH with significant fibrosis.

Conclusions: By applying a guidelines-compliant algorithm, a high proportion of T2D patients was referred for hepatological evaluation and liver biopsy. Further studies are necessary to refine non-invasive algorithms.

**Manuscript undergoing final revisions before being submitted*. Besutti G, Monelli F, Venturelli F, Bonilauri L, Manicardi E, Manicardi V, Massari M, Riva N, Schianchi S, Froio E, Tagliavini E, Bonelli E, Ligabue G, Pattacini P, Giorgi Rossi P. Feasibility and efficiency of the application of international guidelines for NAFLD assessment in type 2 diabetes patients: a prospective study.

INTRODUCTION

The estimated overall global prevalence of non-alcoholic fatty liver disease (NAFLD) is increasing and projected at 33.5% in 2030 [1]. Patients with metabolic risk factors such as obesity, metabolic syndrome, and type 2 diabetes (T2D) are at increased risk of developing not only fatty liver, but also inflammation and hepatocellular injury (Non-Alcoholic Steatohepatitis, NASH), and fibrosis, which may progress to cirrhosis and liver failure, being also major risk factors for hepatocellular carcinoma (HCC) [2]. While accurate elastography techniques exist for the assessment of liver fibrosis [3], non-invasive tests for NASH are in a premature stage of development [4], hence the gold standard for NAFLD assessment for the diagnosis of NASH and the staging of liver fibrosis remains liver biopsy [2].

Several guidelines exist on NAFLD assessment [5-7]. A screening strategy for NAFLD in the general population is generally not recommended, but most guidelines recommend a screening or case-finding approach among patients at high risk for NASH or fibrosis, i.e. patients with metabolic risk factors, and in particular in those with type 2 diabetes (T2D) [8-10]. Indeed, recent studies confirmed that patients with T2D are the group characterized by the highest prevalence of severe liver damage and fibrosis [11]. However, the implementation of these screening strategies in clinical practice is strongly restrained by uncertainties surrounding diagnostic tests and treatment options [2]. The absence of approved therapeutic options particularly weakens the rationale of algorithms that have been proposed to identify patients with NASH, but without severe liver fibrosis. Indeed, for individuals with severe fibrosis the referral to liver disease specialists is warranted in order to implement HCC surveillance and manage liver-related complications in those with cirrhosis [11].

On the other hand, the EASL-EASD-EASO NAFLD recommendations by major European Societies [8] recommended referral of patients with less severe liver disease with the goal of intensifying lifestyle management, include them in clinical trials and ultimately prevent the progression to severe liver disease. However, in retrospective studies, this approach resulted in a referral to the hepatologist of 33-85% of 179 T2DM [12], and 68-75% of 385 severely obese individuals [13]. More recently, a prospective study applying European and German guidelines to a cohort of 204 T2DM patients found similar results, with 60-77% patients referred to the hepatologist [14].

The results of these studies have underlined the weaknesses of existing guidelines in the selection of patients for liver disease specialist referral. As a consequence, in general practice the proposed algorithms are rarely applied, while in specialized centers these algorithms are frequently implemented combining different biomarkers to reduce referral rates [15-16].

Within this context, the aim of this study was to evaluate the feasibility and efficiency of the application of a screening algorithm to detect patients with NASH and NASH with significant fibrosis

in newly diagnosed T2D patients. The tested algorithm has been derived from the relevant recommendations reported by the major guidelines [8].

METHODS

Study design and Ethics

This was a prospective observational monocentric study approved by the local Ethics committee (Comitato Etico Area Vasta Emilia Nord) with protocol number 2018/0085325. All patients provided a written informed consent to participate to the study.

Study Population

All consecutive T2D patients at their first evaluation by a diabetes specialist at our institution from January 2019 to January 2020 were eligible. Inclusion criteria were age ≥18 years, diagnosis of T2D according to national guideline recommendations [17], and consent to participate to the study. The choice to include only patients at their first diabetologist visit, i.e. mostly newly diagnosed, was justified because, once eventually the program would be scaled up, newly diagnosed patients would be the majority of those needing to be screened.

We estimated a sample size of about 300 indeterminate-to-high risk patients based on the number of yearly first visits in diabetes clinics extracted by the population-based diabetes registry [18]. This sample size would allow a precision of +/-5% of the estimate of positive predictive value of referral. Due to changes happened in health care organization, partly related to the initial results of the present study and partly to the COVID-19 health emergency, we report here the results relative to the enrollment between January 2019 and January 2020.

NAFLD assessment algorithm

Laboratory and anthropometric data were collected at the moment of the first diabetologist evaluation. When laboratory data necessary for risk stratification were not available, patients were invited to complete their blood tests as soon as possible.

Figure 1 shows how the EASL-EASD-EASO clinical practice 2016 guidelines algorithm [8] has been contextualized in our clinical practice. Few differences were introduced in order to minimize the impact on workload and the number of different appointments for the patient, according to the available technological resources. Patients aged >75 years, or with secondary causes of steatosis (moderate-to-severe alcohol intake defined as ≥20 g/die in women an ≥30 g/die in men, steatogenic drugs) or known chronic liver diseases were not stratified for NAFLD and referred to a second assessment in diabetes primary care clinics including, if appropriate, nutritional and psychological

support, liver disease specialist assessment, and antiviral therapy. The remaining patients were stratified according to liver enzymes and non-invasive markers of steatosis and fibrosis.

To estimate the likelihood of fatty liver disease, the fatty liver index (FLI), was calculated by using body mass index (BMI), waist circumference, gamma-glutamyl transpeptidase (GGT) and triglycerides [19]. The preferred tool for steatosis diagnosis is liver ultrasound (US), however, for population-based screening studies, serum biomarkers such as FLI are accepted as alternatives [7]. To minimize the impact on feasibility, we screened patients by using FLI and used US to confirm the presence of steatosis in patients with high FLI. Age, BMI, fasting glucose level, (AST), alanine aminotransferase (ALT), platelet count and albumin were used to calculate the NAFLD fibrosis score (NFS) as a marker of fibrosis [20].

Patients with abnormal liver enzymes and patients with US-confirmed steatosis and indeterminate (from 1.455 to 0.675) or high NFS (>0.675) were considered at risk for advanced NAFLD and were referred to the hepatologist, who performed liver stiffness measurement (LSM) by means of vibration-controlled transient elastography (VCTE) in all patients with normal liver enzymes. After further excluding other liver diseases or contraindications, patients with abnormal liver enzymes or LSM >7 kPa were referred to liver biopsy.

Patient risk was also reclassified using age adjusted NFS cut-off values (NFS>0.12 for indeterminate risk in patients aged>65 years) [21]. Moreover, based on age, AST, platelet count, and ALT, the FIB-4 score was calculated as an alternative surrogate biomarker of fibrosis [22], and patients were reclassified into low, indeterminate and high risk according to FIB-4 cut offs 1.3 and 2.96 [15]. Finally, age adjusted cut-offs (FIB-4 >2 for indeterminate risk in patients aged >65 years) were also applied [21].

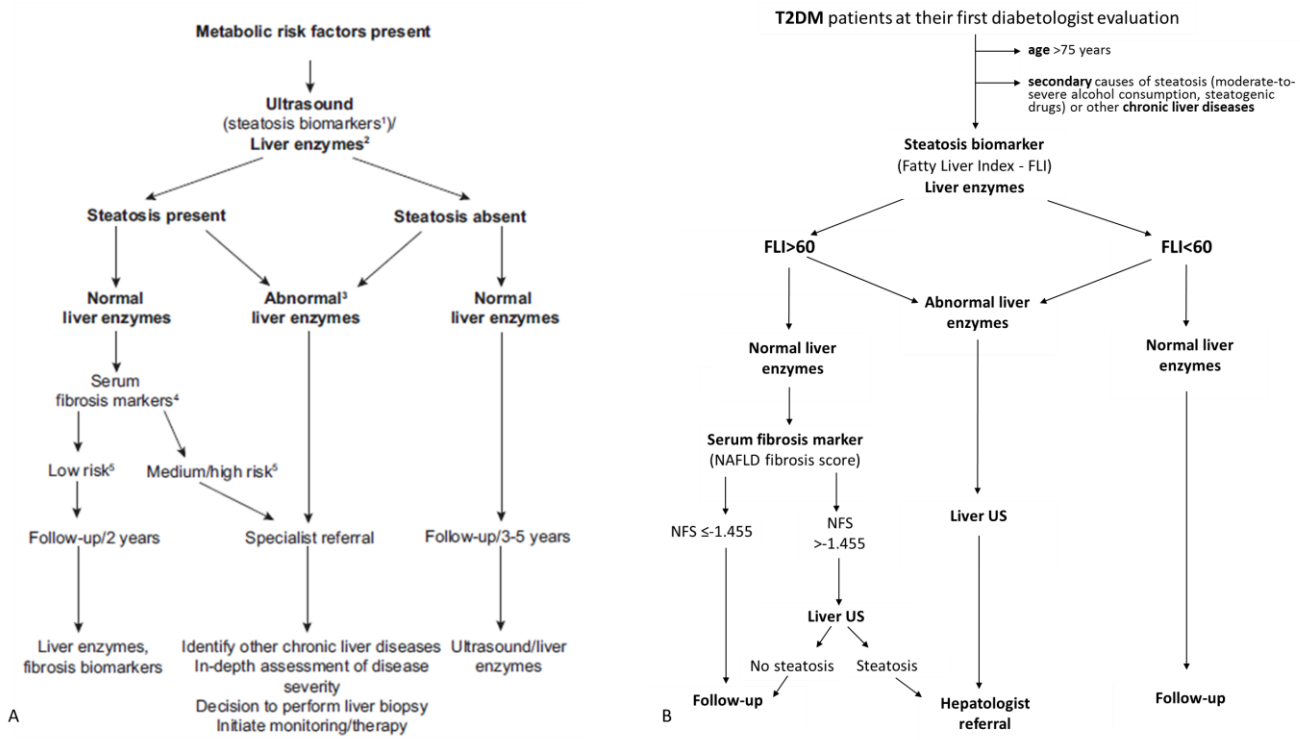


Figure 1: Algorithm proposed by main European scientific societies [7] (A) and its operational application in the study (B).

Liver US

Patients referred to liver US underwent a conventional abdominal ultrasound after a fasting period of at least six hours. All examinations were performed by a single radiologist (GB), using the same ultrasound device (Philips Affiniti 70G equipped with convex Philips C5-1 probe and the ElastoPQ technique). The radiologist evaluated the presence of liver steatosis, based on the visual comparison of echogenicity between liver and right kidney parenchyma.

Patients' risk in those with US-confirmed steatosis was reclassified according to US-fatty liver index (US-FLI) [23] and point shear wave elastography (PSWE) liver stiffness [24]. For PSWE, at least 10 valid measurement were acquired for each patient in the right liver lobe through intercostal spaces. In US-based reclassification, patients were considered at high risk for NASH and/or significant fibrosis if US-FLI was ≥4 [21] or PSWE liver stiffness was ≥5.7 KPa [24].

Liver biopsy and histopathological examination

After excluding contraindications (ascites, platelet count <50.000/mmc, INR>1.5, prothrombin time >50% and serum bilirubin >3 mg/dl), patients referred to liver biopsy by the hepatologist and who accepted the procedure, underwent US-guided liver biopsy with 16-gauge needles. Histology was assessed by two experienced liver histopathologists (EF and ET) blinded to fibrosis scores and US

results. Samples with <11 portal tracts or <15 mm in length were excluded from the analysis. Data were collected on presence and severity of steatosis (0-3), hepatocellular ballooning (0-2), lobular inflammation (0-2), and fibrosis (0-4). Samples were classified in not NAFLD, NAFL not NASH, and NASH, and the SAF (Steatosis, Activity, Fibrosis) score was calculated, according to the classification proposed by the Fatty Liver Inhibition of Progression (FLIP) consortium [25]. Finally, histologically severe disease was defined as the presence of activity ≥2 and/or fibrosis stage ≥F2 [26]. Significant fibrosis was considered for fibrosis stage ≥F2.

Outcome measures and statistical analyses

To evaluate the feasibility of applying the proposed algorithm to the study population, different process indicators were reported, including patient distribution in different risk categories, referral rates for hepatologist assessment and liver biopsy in all at-risk patients and in subgroups with different reasons for referral. We also evaluated patient adherence to different referrals. To evaluate the efficiency of the proposed algorithm, we estimated the positive predictive value (PPV) of specialist referral and liver biopsy referral in the prediction of the presence of NASH and clinically significant fibrosis (≥2) at histological examination. We also assessed the possible impact of using different or additional positivity/referral criteria. Referral rates and PPVs were re-calculated using: 1) FIB-4 instead of NFS, and age adjusted cut-offs, in patients with normal liver enzymes, 2) US-FLI and PSWE to further stratify risk for NASH and significant fibrosis in patients with US-confirmed steatosis.

RESULTS

Study population

From January 2019 to January 2020, 272 consecutive T2DM patients (age 60±13.5 years, 40.4% females) at their first diabetologist appointment were considered eligible for inclusion in the study (Table 1). For 40 of these patients, laboratory data used for risk evaluation were not complete at the time of the first diabetologist appointment nor were they collected in subsequent months. Other 30 patients were over age 75 years, and 31 had coexistent causes of fatty liver or other chronic liver diseases (23 reported moderate-to-high alcohol consumption, 1 was treated with steatogenic drugs and 7 had other known chronic liver diseases) (Figure 2).

Risk stratification and referral rates

The remaining 171 patients were stratified as follows: low-risk patients according to liver enzymes, FLI and NFS (n=35, 20.5%), indeterminate-risk patients with FLI≥60 and NFS between -1.455 and

0.675 (n=49, 28.7%), high-risk patients for FLI≥60 and NFS>0.675 (n=27, 15.8%), and high-risk patients for abnormal liver enzymes (n=60, 35.1%) (Figure 2).

As a consequence, the application of the algorithm to these 171 patients generated 115 (67.3%) referrals for hepatological assessment (60 due to abnormal liver enzymes and 55 due to US-confirmed steatosis plus indeterminate-to-high NFS), and 30 (17.5%) referrals for liver biopsy (19 due to abnormal liver enzymes and 11 due to US-confirmed steatosis plus indeterminate-to-high NFS). Referral rates calculated on the whole population (n=272) were 50% for liver US, 38.6% for hepatologist assessment and 11% for liver biopsy (Table 1).

	Eligible patients (n=272)	**Patients stratified for NAFLD risk (n=171)**	**Low-risk patients (n=35)**	**Patients with increased liver enzymes (n=60)**	**Patients with FLI≥60 and indeterminate NFS (n=49)**	**Patients with FLI≥60 and high NFS (n=27)**
Age (years)	60 ± 13.5	59 ± 10.2	59 ± 7.8	56 ± 9.9	63 ± 3.5	64 ±7.2
Female sex (%)	110 (40.4%)	72 (42.1%)	13 (37.1%)	26 (43.3%)	21 (42.9%)	12 (44.4%)
Body Mass Index (kg/m^2)	31 (27.5;35.7)	32 (27.9;35.8)	26.3 (25.1;28.4)	32.2 (28.9;37.1)	31.7 (29.4;33.6)	37.6 (35.1;40.2)
Waist circumference (cm)	105 (98;115)	106 (99;115)	98 (90.5;104)	107 (100;118)	105 (102;110)	118 (111.5;130)
Glycated haemoglobin (mml/mol)	57 (49; 73)	57 (48;72)	50 (44;63)	60 (51;85)	53 (49;64)	58 (47;70)
AST (U/l)	24 (20;33)	24 (20;33)	20 (17;23.5)	41 (29.5;53)	22 (19;25)	24 (21;26)
ALT (U/l)	30 (22;44)	31 (23;45)	23 (17.5;28.5)	57 (44.5;78)	26 (21;33)	27 (23.5;34.5)
GGT (U/l)	38 (25;63)	40 (25;64.5)	26 (18;36.5)	89 (73;159)	32 (22;44)	35 (25;49)
Total cholesterol (mg/dl)	192 (168;218)	193 (173;218)	195 (174;216)	194 (181; 218)	196 (175;224)	175 (153;208)
LDL cholesterol (mg/dl)	124 (102;147)	126 (109;151)	123 (107;145)	131 (109;162)	127 (110;153)	105 (81;129)
HDL cholesterol (mg/dl)	44 (38;49)	43 (36;49)	46 (42;51)	42 (36;50)	43 (35;48)	43 (36;49)
Triglycerides (mg/dl)	142 (107;200)	147 (110.5;213)	117 (85;152.5)	169.5 (135;250)	161 (126;217)	140 (93.5;163.5)
Albumin (g/dl)	4.1 (3.9;4.3)	4 (3.9;4.2)	4.1 (3.9;4.3)	4.1 (3.9;4.4)	4.1 (3.9;4.2)	3.9 (3.7;4)
Platelet count ($x10^9/l$)	232 (194;276)	230 (194;274)	258 (200;291)	240 (206;302)	244 (214;269)	181 (169;202)
Hepatologist referral (%)	115 (42.3%)	115 (67.3%)	0 (0%)	60 (100%)	37 (75.5%)	18 (66.7%)
Liver biopsy referral (%)	30 (11.0%)	30 (17.5%)	0 (0%)	19 (31.7%)	9 (18.4%)	2 (7.4%)

Table 1: Clinical characteristics of the patients eligible for the study, patients included and stratified for advanced NAFLD, and subgroups of patients with different risk for NASH/fibrosis. Low-risk patients had both normal liver enzymes and FLI and/or NFS; patients with abnormal liver enzymes were considered at high risk irrespective of scores; patients with FLI≥60 and normal liver enzymes were considered at high or indeterminate risk when NFS was >0.675 or between -1.455 and 0.675. FLI, Fatty Liver Index; NFS, NAFLD Fibrosis Score; US, ultrasound.

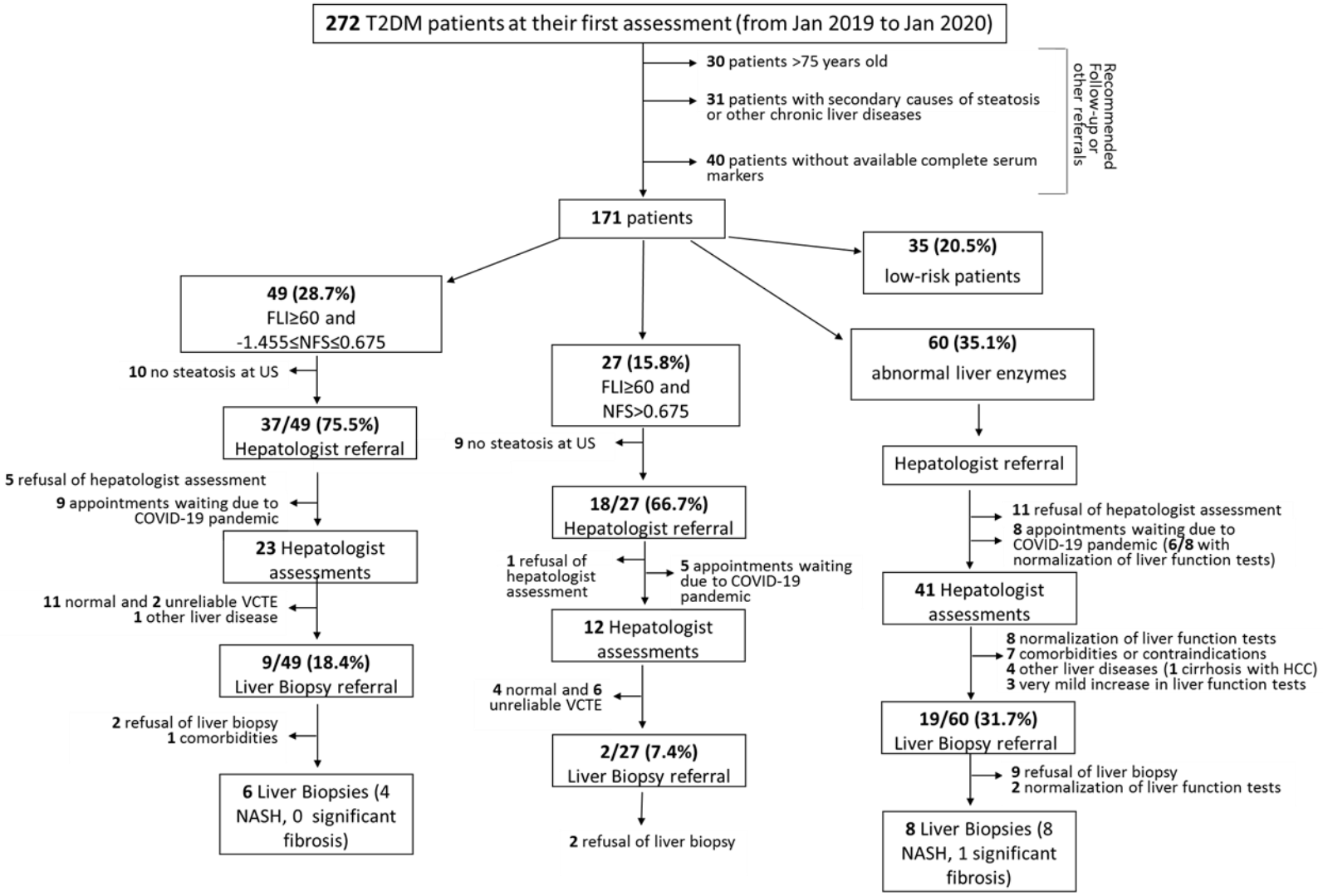


Figure 2: Flow chart representing the application of the proposed algorithm to a cohort of T2DM patients at their first diabetologist assessment. T2DM, type 2 diabetes mellitus; FLI, Fatty Liver Index; NFS, NAFLD Fibrosis Score; US, ultrasound; VCTE, vibration controlled transient elastography; HCC, hepatocellular carcinoma.

Liver disease assessments

Among the 115 patients referred to the liver specialist, 17 refused or ignored the appointment twice, 22 appointments were postponed due to the COVID-19 pandemic, and 76 patients underwent the hepatological assessment. One of the referred patients for abnormal liver enzymes was diagnosed with cirrhosis and HCC, even if he had no visible steatosis at US.

Among the 76 patients evaluated by the hepatologist, 30 were referred for liver biopsy. Of note, 8 out of 60 patients with abnormal liver enzymes were not referred for liver biopsy due to normalization of liver enzymes during the time gap between diabetologist and hepatologist assessments, and other 6 patients whose appointment was postponed had normalization of liver enzymes while waiting. The primary reason for the lack of referral for liver biopsy in patients with abnormal scores was a normal (n=16) or unreliable (n=8) LSM. The 8 patients with unreliable LSM were all severely obese and were referred for follow-up after dietary counselling.

Adherence to referrals

Ninety-eight out of 115 patients referred to the hepatologist (85.2%) accepted. Among the patients who refused, 8 already knew to have fatty liver from previous US examinations and underestimated the clinical significance of steatosis. Of the 30 patients referred to liver biopsy, 13 refused (43.3%) preferring follow-up.

Liver biopsy results

Of the 17 patients who accepted to undergo a liver biopsy, 2 had normalization of liver enzymes and 1 experienced a severe comorbidity while waiting for liver biopsy. The remaining 14 patients (8 initially referred to the hepatologist for abnormal liver enzymes and 6 for steatosis plus indeterminate NFS) underwent a liver biopsy. Among patients referred to liver biopsy for abnormal liver enzymes, all had NASH, only one with significant fibrosis. Among patients referred for US-confirmed steatosis plus indeterminate NFS, 2 had no histological steatosis (<5%) thus were classified as not NAFLD, and the other 4 patients had NASH, all without significant fibrosis. Activity, SAF score and disease severity were slightly higher in patients referred for abnormal liver enzymes (Table 2).

		Patients referred to liver biopsy (n=14)	Patients referred for abnormal LFTs (n=8)	Patients referred for abnormal scores (n=6)
Histological steatosis percentage, median (range)		37.5% (<5%; 75%)	55% (5%; 75%)	27.5% (<5%; 40%)
Steatosis, n (%)	0	2 (14.3%)	-	2 (33.3%)
	1	5 (35.7%)	2 (25%)	3 (50%)
	2	4 (28.6%)	3 (37.5%)	1 (16.7%)
	3	3 (21.4%)	3 (37.5%)	-
Lobular Inflammation, n (%)	0	-	-	-
	1	10 (71.4%)	5 (62.5%)	5 (83.3%)
	2	4 (28.6%)	3 (37.5%)	1 (16.7%)
Hepatocellular ballooning, n (%)	0	-	-	-
	1	12 (85.7%)	6 (75%)	6 (100%)
	2	2 (14.3%)	2 (25%)	-
Activity, n (%)	0	-	-	-
	1	-	-	-
	2	9 (64.3%)	4 (50%)	5 (83.3%)
	3	4 (28.6%)	3 (37.5%)	1 (16.7%)
	4	1 (7.1%)	1 (12.5%)	-
Fibrosis, n (%)	0	3 (21.4%)	2 (25%)	1 (16.7%)
	1	10 (71.4%)	5 (62.5%)	5 (83.3%)
	2	-	-	-
	3	1 (7.1%)	1 (12.5%)	-
	4	-	-	-
SAF score (median, range)		5 (2; 10)	5 (4; 10)	4 (2; 5)
NAFLD classification, n (%)	Not NAFLD	2 (14.3%)	-	2 (33.3%)
	NAFL not NASH	-	-	-
	NASH	12 (85.7%)	8 (100%)	4 (66.7%)
Disease severity, n (%) *	Not severe	7 (58.3%)	4 (50%)	3 (75%)
	Severe	5 (41.7%)	4 (50%)	1 (25%)
Liver biopsy referral PPV for NASH		12/14 (85.7%)	8/8 (100%)	4/6 (66.7%)
Liver biopsy referral PPV for significant fibrosis		1/14 (7.1%)	1/8 (12.5%)	-

Table 2: Histological characteristics of the 14 patients who underwent liver biopsy. SAF, Steatosis Activity Fibrosis; NAFLD, Non-Alcoholic Fatty Liver Disease; NASH, Non-Alcoholic Steatohepatitis; PPV, positive predictive value. *reported for NASH patients (excluding the 2 patients classified as not NAFLD).

PPV of hepatologist assessment and liver biopsy referrals

The PPVs of liver specialist referral were 12/76 (15.8%) for NASH and 1/76 (1.3%) for NASH with significant fibrosis. PPVs in patients referred for abnormal liver enzymes were 8/41 for NASH and

1/41 for NASH with significant fibrosis, while in patients referred for steatosis plus indeterminate NFS they were 4/35 and 0/35.

The PPVs of liver biopsy referral were 12/14 (85.7%) for NASH and 1/14 (7.1%) for NASH with significant fibrosis, 8/8 and 1/8 in patients referred for abnormal liver enzymes and 4/6 and 0/6 in patients referred for steatosis plus indeterminate NFS.

Reclassification with FIB-4 and age-adjusted NFS

By using FIB-4 (instead of NFS) in combination with FLI in patients with normal liver enzymes (Supplementary Table 1, Figure 3), none of the 35 low-risk patients was reclassified as indeterminate or high risk. Instead, 41 patients out of 76 (53.9%) originally considered indeterminate-to-high risk based on NFS were reclassified as low-risk when using FIB-4 (5/27 originally classified as high risk and 36/49 originally classified as indeterminate risk patients). Twenty-two of the 41 reclassified patients were assessed by the hepatologist and 8 of them had increased LSM and were referred for liver biopsy. The 5 patients who accepted liver biopsy resulted in one not NAFLD and 4 NASH without significant fibrosis. The use of FIB-4 instead of NFS would have avoided 5/6 liver biopsies referred for abnormal fibrosis score, and 4 NASH cases without significant liver fibrosis would have been missed. By applying age adjusted FIB-4 cut-offs [21], other 11 patients would have been re-classified as low-risk, and all of the 6 biopsies would have been avoided, including the remaining patient who received liver biopsy resulting in not NAFLD (Supplementary Table 1).

By applying age adjusted NFS cut-offs [21], 23 patients originally classified as indeterminate-risk were re-classified as low-risk. This would have avoided 21 hepatologist referrals and 6 liver biopsy referrals. Four of the 6 liver biopsies referred for abnormal fibrosis scores would have been avoided (one not NAFLD and 3 NASH without significant fibrosis).

Reclassification with US-FLI and PSWE

In patients whit US-confirmed steatosis, US-FLI and PSWE were evaluated. By using combined cut-off values of US-FLI≥4 [23] or PSWE liver stiffness >5.7 KPa [24], 12 out of 60 patients who were referred to the hepatologist for increased liver enzymes were reclassified as US-based low risk. Of these 12, only one was referred for liver biopsy, but refused. Similarly, 16 out of 55 patients originally referred for abnormal NFS were reclassified as US-based low-risk. Of these 16, 3 were referred for liver biopsy resulting in one not NAFLD and two NASH without significant fibrosis (Supplementary Table 2, Figure 3). Overall, the use of US-based risk stratification for NASH and significant fibrosis [23-24] in patients with abnormal liver enzymes would have reduced the rate of referrals to the

hepatologist without missing cases, thus increasing PPV, while in patients with normal liver enzymes it would have also lead to miss two NASH cases.

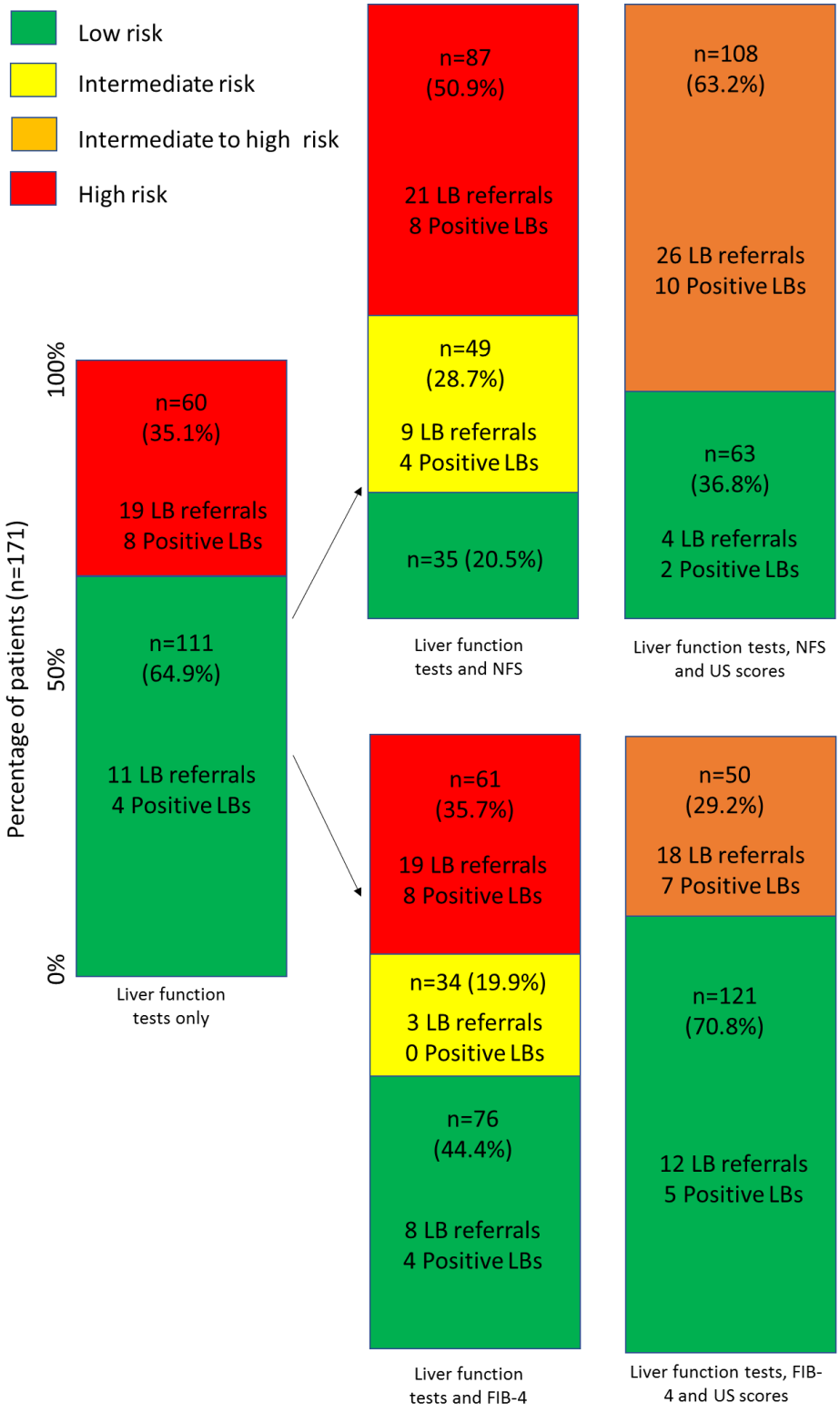


Figure 3: Risk classification hypothesis using alternative fibrosis biomarkers besides liver function tests: intermediate (from -1.455 to 0.675) and high (>0.675) NFS, or intermediate (from 1.3 to 2.96)

and high (>2.96) FIB-4, and adding US scores as an adjunctive criterion. LB= liver biopsy. Positive liver biopsies are those resulting in NASH and/or significant fibrosis.

DISCUSSION

In this study, we sought to implement currently held European guidelines for liver disease specialist referral of patients with dysmetabolism [8], to real-life practice in the evaluation of patients with a new T2D diagnosis. A main finding was that, after stratification by liver enzymes, steatosis and fibrosis indices, the application of the adapted guidelines-compliant algorithm was burdened by major limitations. The first one is represented by the high proportion of patients needing referral, as 67% patients were referred to liver disease specialist assessment and 17.5% for liver biopsy. Secondly, the adherence of patients to referrals was low, especially for liver biopsy, with more than 40% refusals. The PPVs of hepatologist and liver biopsy referrals were 15.8% and 85.7% for NASH, and most importantly 1.3% and 7.1% for clinically significant fibrosis, the main prognostic determinant for liver-related events. Trying to apply more specific criteria for referring patients to the hepatologist, including FIB-4 and US-determined indices, decreased referral rates, but also the detection of cases with NASH without significant liver fibrosis. As surveillance for liver disease complications is recommended only for patients with severe fibrosis, while the population of patients targeted by investigational NASH therapies is currently represented by those with concurrent presence of NASH and clinically significant fibrosis, application of these more stringent criteria may therefore represent an alternative strategy to improve the cost-effectiveness of the fatty liver referral pathway.

Our findings are in line with epidemiological data on the prevalence of fatty liver disease in individuals with T2D [27] and with the results of recent studies [12-13] reporting that, according to current guidelines, 33-85% of T2D and 68-75% severely obese patients would be referred to a liver clinic. Similarly, a prospective study on a cohort of 204 T2D patients resulted in a referral rate to the hepatologist of 60-77% and a 7% referral rate to biopsy [14]. Despite we only considered newly diagnosed patients, who tend to have less severe liver damage [11, 28], the referral was similar. Furthermore, we did not consider older patients and those with coexistent triggers of liver disease in order to minimize the impact on feasibility. However, it should be noted that the exclusion of patients with secondary causes of steatosis, even if coherent with the definition of NAFLD, may need to be reconsidered. In fact, the recent proposal to change the focus from NAFLD to metabolic dysfunction associated fatty liver disease (MAFLD) [29] may result in the inclusion of patients reporting high alcohol intake or the use of steatogenic drugs, who are known to be at high risk of liver disease progression. This new definition is especially relevant for T2D patients, and by applying it 24 patients

would have not been excluded from initial stratification, probably leading to an increase in referral rates, but for a high-risk subset.

Since main guidelines do not specify which serum biomarkers of fibrosis to use, we choose to use NFS because it was already applied in the standard care at our institution. However, since impaired fasting glucose/diabetes are included in its formula, NFS overestimates the presence of severe liver fibrosis in T2D patients [30]. Thus, it is not surprising that the use of FIB4 instead of NFS produced more low-risk patients, leading to lower referral rates, as already described [11, 14, 31]. We found that, if we had used FIB-4 instead of NFS, all NASH cases in patients with normal liver enzymes would have been missed. Similarly, lower referral rates along with missing of NASH cases would have been obtained by applying age-adjusted NFS or FIB-4 cut-offs [21]. It should be noted that both of these scores are aimed at finding patients with severe fibrosis, and not NASH. Even if fibrosis is the most important histological feature associated with morbidity and mortality in NAFLD patients [32], European guidelines still recommend a case-finding approach also of patients with NASH, since they may progress to significant fibrosis in upcoming years [8]. Our results along with the recent literature [11] suggest that this is not currently feasible in real-life settings, where instead it should remain mandatory to identify patients with severe fibrosis, and probably those with both NASH and significant fibrosis who will be amenable of treatment when the first NASH drugs will become available.

As such, a sequential two-step combination of a largely available fibrosis score followed by elastography is gaining momentum as the preferred referral strategy in individuals with dysmetabolism and T2D [11, 15-16]. Indeed, the use of imaging biomarkers along with serum biomarkers, may improve the specificity of the screening algorithm. Blank et al. demonstrated a decrease in referral rate when using VCTE [14]. In other studies which included liver biopsy results, VCTE and controlled attenuation parameter (CAP) [28] or magnetic resonance imaging (MRI) and MRE [33] combined were used to screen for steatosis and increased liver stiffness in T2D patients. However, MRE is not largely available and too costly to be used as a screening tool, and also VCTE is usually available only in liver specialized clinics, being usually performed by the hepatologist, after referral. On the contrary, the increasing availability of US-based elastography techniques [34] allows the screening for increased liver stiffness along with steatosis in a first step. In the present study, we applied US-based indices including US-FLI and LSM by PSWE to refine referrals. This strategy resulted in lower referrals without loss of diagnosis of significant fibrosis. Some NASH diagnoses would have been missed, yet again indicating the incapacity of available biomarkers to select NASH patients.

Interestingly, among patients who delayed the hepatologist visit or the biopsy, improvements of the hepatic function was common. This could be due simply to a regression to the mean phenomenon or to the effectiveness of dietary and lifestyle changes occurring just after diabetes diagnosis. Either way, the repetition of liver enzymes before the hepatologist referral should be encouraged. This point is underlined only in some of the existing recommendations, including the Italian and WGO guidelines [9-10], which stress the need to consider liver biopsy in case of the lack of normalization of liver enzymes. It should be noted that this last point is probably even more meaningful in our study, since patients at their first diabetologist assessment were more prone to undergo lifestyle changes that can result in liver enzymes normalization.

Another hint that can be derived from our findings involves the need to focus on the better way to propose and discuss with patients the pros and cons of diagnostic procedures, in order to be sure that the low compliance to referrals is the consequence of an informed choice and not of unclear messages on the clinical significance of fatty liver, which is frequently underestimated [35].

The main strengths of this study are its population-based nature and its pragmatic approach. These characteristics allowed us to estimate the real impact on workload of the potential application of European NAFLD screening guidelines, including actual patients' compliance, and proportion of non-eligible patients. The pragmatic approach was also the cause of some of the weaknesses of the study, mainly leading to a large proportion of patients lacking full evaluation. Moreover, to minimize the impact on feasibility, we used FLI as a steatosis biomarker and then confirmed steatosis with US in patients stratified according to fibrosis biomarkers. Only patients with US-confirmed steatosis (unless they had abnormal liver enzymes or US signs of cirrhosis) were referred to the hepatologist. This may create confusion, since usually steatosis assessment is the first step and does not follow stratification of fibrosis risk. It can be argued that patients with indeterminate-to-high risk according to fibrosis biomarkers should have been referred to the hepatologist regardless of the presence of absence of steatosis. In fact, by excluding from hepatologist assessment patients with high risk of fibrosis and absence of steatosis, we may miss cases of burnt-out steatohepatitis [36].

In conclusion, we found that by applying a screening algorithm for NASH and NASH with significant fibrosis to T2D patients, which was adapted from the EASL-EASD-EASO 2016 guidelines to the local health care setting, 67% of patients were referred to a liver disease specialist and 17.5% to liver biopsy. Patients' adherence was 82% for hepatological evaluation and 40% for biopsy, with referral PPVs for NASH with significant fibrosis being only 1.7% and 7.1%, respectively. Refinement in the selection criteria for patients at higher risk of liver events or candidate to treatment, including the combined evaluation of liver stiffness, will be required to allow a wider implementation of these measures to improve liver outcomes in T2D patients.

References

1. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018 Jan;67(1):123-133. doi: 10.1002/hep.29466.

2. Chalasani N, Younossi Z, Lavine JE, et al. The Diagnosis and Management of Nonalcoholic Fatty Liver Disease : Practice Guidance From the American Association for the Study of Liver Diseases. Hepatology. 2018;67(1):328–57. doi: 10.1002/hep.29367.

3. Xiao G, Zhu S, Xiao X, Yan L, Yang J, Wu G. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: A meta-analysis. Hepatology. 2017 Nov;66(5):1486-1501. doi: 10.1002/hep.29302.

4. Besutti G, Valenti L, Ligabue G, Bassi MC, Pattacini P, Guaraldi G, Giorgi Rossi P. Accuracy of imaging methods for steatohepatitis diagnosis in non-alcoholic fatty liver disease patients: A systematic review. Liver Int. 2019 Aug;39(8):1521-1534. doi: 10.1111/liv.14118.

5. Zhu JZ, Hollis-Hansen K, Wan XY, Fei SJ, Pang XL, Meng FD, Yu CH, Li YM. Clinical guidelines of non-alcoholic fatty liver disease: A systematic review. World J Gastroenterol. 2016 Sep 28;22(36):8226-33. doi: 10.3748/wjg.v22.i36.8226.

6. Leoni S, Tovoli F, Napoli L, Serio I, Ferri S, Bolondi L. Current guidelines for the management of non-alcoholic fatty liver disease: A systematic review with comparative analysis. World J Gastroenterol. 2018 Aug 14;24(30):3361-3373. doi: 10.3748/wjg.v24.i30.3361.

7. Monelli F, Venturelli F, Bonilauri L, Manicardi E, Manicardi V, Rossi PG, Massari M, Ligabue G, Riva N, Schianchi S, Bonelli E, Pattacini P, Bassi MC, Besutti G. Systematic review of existing guidelines for NAFLD assessment. Hepatoma Res 2021;7:25. http://dx.doi.org/10.20517/2394-5079.2021.03.

8. European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO

Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. J Hepatol. 2016 Jun;64(6):1388-402. doi: 10.1016/j.jhep.2015.11.004.

9. Loria P, Adinolfi LE, Bellentani S, Bugianesi E, Grieco A, et. al. NAFLD Expert Committee of the Associazione Italiana per lo studio del Fegato. Practice guidelines for the diagnosis and management of nonalcoholic fatty liver disease. A decalogue from the Italian Association for the Study of the Liver (AISF) Expert Committee. Dig Liver Dis. 2010 Apr;42(4):272-82.

10. LaBrecque DR, Abbas Z, Anania F, Ferenci P, Khan AG, et. al. World Gastroenterology Organisation. World Gastroenterology Organisation global guidelines: Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. J Clin Gastroenterol. 2014 Jul;48(6):467-73.

11. Boursier J, Tsochatzis EA. Case-finding strategies in non-alcoholic fatty liver disease. JHEP Rep. 2020 Dec 5;3(2):100219. doi: 10.1016/j.jhepr.2020.100219.

12. Sberna AL, Bouillet B, Rouland A, Brindisi MC, Nguyen A, Mouillot T, et al. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD) and European Association for the Study of Obesity (EASO) clinical practice recommendations for the management of non-alcoholic fatty liver disease: evaluation of their application in people with Type 2 diabetes. Diabet Med. 2018 Mar;35(3):368-375. doi: 10.1111/dme.13565.

13. Blond E, Disse E, Cuerq C, Drai J, Valette PJ, Laville M, et al. EASL-EASD-EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease in severely obese people: do they lead to over-referral? Diabetologia. 2017 Jul;60(7):1218-1222. doi: 10.1007/s00125-017-4264-9.

14. Blank V, Petroff D, Beer S, Böhlig A, Heni M, Berg T, et al. Current NAFLD guidelines for risk stratification in diabetic patients have poor diagnostic discrimination. Sci Rep. 2020 Oct 27;10(1):18345. doi: 10.1038/s41598-020-75227-x.

15. Srivastava A, Gailer R, Tanwar S, Trembling P, Parkes J, Rodger A, et al. Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease. J Hepatol. 2019 Aug;71(2):371-378. doi: 10.1016/j.jhep.2019.03.033.

16. Boursier J, Guillaume M, Leroy V, Irlès M, Roux M, Lannes A, et al. New sequential combinations of non-invasive fibrosis tests provide an accurate diagnosis of advanced fibrosis in NAFLD. J Hepatol. 2019 Aug;71(2):389-396. doi: 10.1016/j.jhep.2019.04.020.

17. Standard Italiani per la cura del diabete mellito 2018 – Associazione Medici Diabetologi e Società Italiana di diabetologia – accessible at https://www.siditalia.it/pdf/Standard%20di%20Cura%20AMD%20%20SID%202018_protetto.pdf.

18. Ballotari P, Chiatamone Ranieri S, Vicentini M, Caroli S, Gardini A, Rodolfi R, et al. Building a population-based diabetes register: an Italian experience. Diabetes Res Clin Pract. 2014 Jan;103(1):79-87. doi: 10.1016/j.diabres.2013.11.020.

19. Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, et. al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol. 2006 Nov 2;6:33.

20. Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, et. al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology. 2007 Apr;45(4):846-54. Doi: 10.1002/hep.21496.

21. McPherson S, Hardy T, Dufour JF, Petta S, Romero-Gomez M, Allison M, et al. Age as a Confounding Factor for the Accurate Non-Invasive Diagnosis of Advanced NAFLD Fibrosis. Am J Gastroenterol. 2017 May;112(5):740-751. doi: 10.1038/ajg.2016.453.

22. Sterling, R. K. et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology 43, 1317–1325 (2006).

23. Ballestri S, Lonardo A, Romagnoli D, et al. Ultrasonographic fatty liver indicator , a novel score which rules out NASH and is correlated with metabolic parameters in NAFLD. Liver Int. 2012;32(8):1242–52. doi: 10.1111/j.1478-3231.2012.02804.x.

24. Ferraioli G, Tinelli C, Lissandrin R, Zicchetti M, Dal Bello B, Filice G, et al. Point shear wave elastography method for assessing liver stiffness. World J Gastroenterol. 2014 Apr 28;20(16):4787-96. doi: 10.3748/wjg.v20.i16.4787

25. Bedossa P. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. *Hepatology*. 2014;60:565–575.

26. Nascimbeni F, Bedossa P, Fedchuk L, Pais R, Charlotte F, Lebray P, Poynard T, Ratziu V; LIDO (Liver Injury in Diabetes and Obesity) Study Group. Clinical validation of the FLIP algorithm and the SAF score in patients with non-alcoholic fatty liver disease. J Hepatol. 2020 May;72(5):828-838. doi: 10.1016/j.jhep.2019.12.008.

27. Younossi ZM, Golabi P, de Avila L, Paik JM, Srishord M, Fukui N, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: A systematic review and meta-analysis. J Hepatol. 2019 Oct;71(4):793-801. doi: 10.1016/j.jhep.2019.06.021.

28. Kwok R, Choi KC, Wong GL, et al. Screening diabetic patients for non-alcoholic fatty liver disease with controlled attenuation parameter and liver stiffness measurements: a prospective cohort study. Gut 2016;65:1359-1368.

29. Fouad Y, Waked I, Bollipo S, Gomaa A, Ajlouni Y, Attia D. What's in a name? Renaming 'NAFLD' to 'MAFLD'. Liver Int. 2020 Jun;40(6):1254-1261. doi: 10.1111/liv.14478. Epub 2020 Apr 28.

30. Anstee QM, Lawitz EJ, Alkhouri N, et al. Noninvasive tests accurately identify advanced fibrosis due to NASH: baseline data from the stellar trials. Hepatology 2019;70:1521–30.

31. Ciardullo S, Muraca E, Perra S, Bianconi E, Zerbini F, Oltolini A, et al. Screening for non-alcoholic fatty liver disease in type 2 diabetes using non-invasive scores and association with diabetic complications. BMJ Open Diabetes Res Care. 2020 Feb;8(1):e000904. doi: 10.1136/bmjdrc-2019-000904.

32. Dulai P, Singh S, Patel J, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. Hepatology. 2017;65(5):1557–65. doi: 10.1002/hep.29085.

33. Doycheva I, Cui J, Nguyen P, Costa EA, Hooker J, et. al. Non-invasive screening of diabetics in primary care for NAFLD and advanced fibrosis by MRI and MRE. Aliment Pharmacol Ther. 2016 Jan;43(1):83-95.

34. Barr RG. Shear wave liver elastography. Abdom Radiol (NY). 2018 Apr;43(4):800-807. [PMID: 29116341 DOI: 10.1007/s00261-017-1375-1].

35. Edenvik P, Davidsdottir L, Oksanen A, Isaksson B, Hultcrantz R, Stål P. Application of hepatocellular carcinoma surveillance in a European setting. What can we learn from clinical practice? Liver Int. 2015 Jul;35(7):1862-71. doi: 10.1111/liv.12764.

36. Powell EE, Cooksley WG, Hanson R, Searle J, Halliday JW, Powell LW. The natural history of nonalcoholic steatohepatitis: a follow-up study of forty-two patients for up to 21 years. Hepatology. 1990 Jan;11(1):74-80. doi: 10.1002/hep.1840110114.

5.1.5. Accuracy of Imaging Techniques in Diagnosing Steatohepatitis and Fibrosis in NAFLD Patients (ImagingNAFLD): study protocol*

Summary: To assess the role of NAFLD diagnostic tools we designed a prospective study to evaluate diagnostic accuracy of US and MR techniques for the diagnosis of NASH and fibrosis in patients with NAFLD, which is currently ongoing.

Non-alcoholic fatty liver disease (NAFLD) is a highly prevalent condition, and when fatty liver is associated with inflammation and hepatocellular injury (steatohepatitis), it can lead to fibrosis, cirrhosis, liver failure and hepatocellular carcinoma. Liver biopsy is the gold standard for NAFLD assessment but has several drawbacks. Several drugs for NASH are now in phase 2-3 trials, and if medical treatments become available, non-invasive tools to identify patients who may benefit from a therapeutic intervention will be strongly needed. Some imaging methods have shown promising potential in fibrosis and NASH diagnosis. This study aims to evaluate the diagnostic accuracy of non-invasive imaging methods, including ultrasound (US) and Magnetic Resonance (MR) techniques, in diagnosing NASH and fibrosis in patients with or at high risk of NAFLD, using liver biopsy as the reference standard. Consecutive patients with a clinical indication for liver biopsy assessment of NAFLD are enrolled in this non-inferiority study. They undergo both a liver US and a multiparametric unenhanced liver MR examination. As reference standard, histological diagnosis of fibrosis and steatohepatitis made according to the fatty liver inhibition of progression (FLIP) algorithm is used. Sensitivity and specificity of imaging parameters alone or in different combinations will be calculated with the aim of finding one or more tests with at least 90% sensitivity/specificity compared to liver biopsy.

The enrollment began in March 2019 but, due to the COVID-19 pandemic, it was interrupted in March 2020 and resumed in November 2020, and it is now proceeding slowly. On March 1, 2021, 22 patients have been enrolled. First interim analyses are planned when results will be available for 50 patients. These analyses will guide the decision about continuing recruitment up to the anticipated sample of 100 patients.

*Study protocol registered in Clinicaltrial.gov, with Identifier NCT04785937

Detailed Description

The estimated overall global prevalence of non-alcoholic fatty liver disease (NAFLD) is around 25% and projected at 33.5% in 2030 [1]. While simple steatosis without evidence of inflammation and hepatocellular injury (non-alcoholic fatty liver) is generally a benign condition, non-alcoholic steatohepatitis (NASH) can progress to fibrosis, cirrhosis, liver failure and hepatocellular carcinoma. Since only histological analysis can accurately evaluate NAFLD patterns, liver biopsy is the gold standard for assessment, and it should be considered in patients who are at increased risk of having steatohepatitis and/or fibrosis [2,3]. Major drawbacks are its invasive nature, risk of complications, sampling errors and inter and intra-observer variability [4]. Currently, there are no approved therapies for NASH. However, several drugs are now in phase 2 and 3 trials, and results are expected in 1-2 years [5]. If medical treatments become available, screening for steatohepatitis and fibrosis will be recommended in high-risk patients. The lack of non-invasive tools to identify patients who may benefit from a therapeutic intervention is a central issue. Should liver biopsy be avoided or reserved for a more limited number of undetermined or high-risk patients, the benefit-harm balance of NASH screening and therapies would undergo a major change. Some imaging methods, mostly ultrasound (US) or Magnetic Resonance (MR) techniques, have shown promising potential in fibrosis and NASH diagnosis [6,7].

The objective of this study is to evaluate the diagnostic accuracy of non-invasive imaging techniques including US and MR methods, in diagnosing NASH and fibrosis in patients with or at high risk of NAFLD, using liver biopsy as the reference standard.

Consecutive patients with a clinical indication for liver biopsy assessment of NAFLD are enrolled in this non-inferiority study. They undergo both a liver ultrasound (US), including shear wave elastography (SWE) with liver stiffness measurement [6] and US- fatty liver index (US-FLI) [8], and a multiparametric unenhanced liver magnetic resonance examination [9,10] including MR spectroscopy (MRS), Proton Density Fat Fraction (PDFF) and T2* measurement with Multiecho technique, T1 mapping with Inversion Recovery method, and Intravoxel Incoherent Motion diffusion weighted imaging (IVIM-DWI), measuring different parameters. As reference standard, histological diagnosis of fibrosis and steatohepatitis made according to the fatty liver inhibition of progression (FLIP) algorithm [11] is used. Sensitivity and specificity of imaging parameters alone or in different combinations will be calculated, with the aim of finding one or more tests with at least 90% sensitivity/specificity compared to liver biopsy.

Study Design

This single group interventional diagnostic study is estimated to enroll 100 participants (interim analyses are planned after 50 participants). Actual Study Start Date: January 1, 2019. Estimated Study Completion Date: June 1, 2022.

Arms and Interventions

Experimental Arm: Imaging and Biopsy. All patients undergo both liver biopsy and liver imaging (US and multiparametric MR) to assess the diagnostic performance of imaging compared to histopathological examination in the diagnosis of NASH and fibrosis.

Intervention/treatment: Diagnostic test – Ultrasound and Magnetic Resonance (MR). Liver ultrasound (US), including shear wave elastography (SWE) with liver stiffness measurement and US- fatty liver indicator (US-FLI), and a multiparametric unenhanced liver magnetic resonance examination including MR spectroscopy (MRS), Proton Density Fat Fraction (PDFF) and T2* measurement with Multiecho technique, T1 mapping with Inversion Recovery method, and Intravoxel Incoherent Motion diffusion weighted imaging (IVIM-DWI), measuring different parameters.

Outcome Measures

Primary Outcome Measures:

1. False positives and False negatives [Time Frame: baseline]: false positives and false negatives of each imaging parameter, alone or in different combinations, compared to liver biopsy in the diagnosis of NASH/fibrosis.
2. Sensitivity and Specificity [Time Frame: baseline]: sensitivity and specificity of each imaging parameter, alone or in different combinations, compared to liver biopsy in the diagnosis of NASH/fibrosis.

Secondary Outcome Measures:

1. Correlation of each imaging parameter (US-FLI, US-SWE stiffness, MR-T1, MR-PDFF, MR-T2*, MR-IVIM-coefficients, MRS metabolites) with histopathological, demographic, anthropometric and clinical characteristics [Time Frame: baseline].
2. Number of patients with incomplete or unreliable imaging tests [Time Frame: baseline].

Recruitment setting and process

Eligible patients will be recruited by the liver specialist at the moment of referral for liver biopsy due to NAFLD assessment. The participation in the study will be proposed to patients who accept to undergo liver biopsy for clinical reasons. After the explanation of pros and cons, patients who accept to participate in the study should sign the informed consent.

Eligibility Criteria

Ages and sexes eligible: 18 years and older; both sexes.

Inclusion criteria:

- clinical indication to perform a liver biopsy for NAFLD assessment based on all of the following:
    1. presence of liver steatosis at ultrasound;
    2. at least one risk factor for NASH/fibrosis (obesity, or type 2 diabetes mellitus, or metabolic syndrome);
    3. increased liver enzymes (at least one of: GOT>40 U/l, GPT>49 U/l, GGT>75 U/l) or high NAFLD fibrosis score (>0.675), or intermediate NAFLD fibrosis score (between -1.455 and 0.675) and increased liver stiffness at transient elastography (>7 KPa).
- consent to participate in the study.

Exclusion criteria:

- age < 18 years;
- secondary causes of liver steatosis (moderate to severe alcohol consumption, steatogenic drugs);
- known diffuse liver diseases other than NAFLD (cirrhosis, viral or autoimmune hepatitis, hemochromatosis, amyloidosis, other) or previous primary or secondary liver neoplasms;
- contraindications to perform liver biopsy (ascites, platelet count<50.000/mmc, INR>1.5, PT>50%, serum bilirubin >3 mg/dL);
- contraindications to perform magnetic resonance (pace-maker, claustrophobia, pregnancy, MR-unsafe metallic implants).

Sample size calculation

To obtain a statistical power of 80%, in the hypothesis of a sensitivity (or specificity) of one or more imaging methods compared to liver biopsy equal to 96%, to exclude a sensitivity (or specificity) ≤89%, we need to enroll 50 patients with NASH or significant fibrosis and 50 patients

without NASH or significant fibrosis. If we estimate a NASH/fibrosis prevalence of 50%, we will need to enroll 100 participants.

Ethics

The study is conducted consistently with the protocol and the principles of the declaration of Helsinky, after the approval by the Ethic Committee.

References

1. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018 Jan;67(1):123-133. doi: 10.1002/hep.29466.

2. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, Harrison SA, Brunt EM, Sanyal AJ. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology. 2018 Jan;67(1):328-357. doi: 10.1002/hep.29367.

3. European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. J Hepatol. 2016 Jun;64(6):1388-402. doi: 10.1016/j.jhep.2015.11.004.

4. Sumida Y, Nakajima A, Itoh Y. Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. World J Gastroenterol. 2014 Jan 14;20(2):475-85. doi: 10.3748/wjg.v20.i2.475.

5. Issa D, Patel V, Sanyal AJ. Future therapy for non-alcoholic fatty liver disease. Liver Int. 2018 Feb;38 Suppl 1:56-63. doi: 10.1111/liv.13676.

6. Park CC, Nguyen P, Hernandez C, Bettencourt R, Ramirez K, Fortney L, Hooker J, Sy E, Savides MT, Alquiraish MH, Valasek MA, Rizo E, Richards L, Brenner D, Sirlin CB, Loomba R. Magnetic Resonance Elastography vs Transient Elastography in Detection of Fibrosis and Noninvasive Measurement of Steatosis in Patients With Biopsy-Proven Nonalcoholic Fatty Liver Disease. Gastroenterology. 2017 Feb;152(3):598-607.e2. doi: 10.1053/j.gastro.2016.10.026.

7. Besutti G, Valenti L, Ligabue G, Bassi MC, Pattacini P, Guaraldi G, Giorgi Rossi P. Accuracy of imaging methods for steatohepatitis diagnosis in non-alcoholic fatty liver disease patients: A systematic review. Liver Int. 2019 Aug;39(8):1521-1534. doi: 10.1111/liv.14118.

8. Ballestri S, Lonardo A, Romagnoli D, Carulli L, Losi L, Day CP, Loria P. Ultrasonographic fatty liver indicator, a novel score which rules out NASH and is correlated with metabolic parameters in NAFLD. Liver Int. 2012 Sep;32(8):1242-52. doi: 10.1111/j.1478-3231.2012.02804.x.

9. Eddowes PJ, McDonald N, Davies N, Semple SIK, Kendall TJ, Hodson J, Newsome PN, Flintham RB, Wesolowski R, Blake L, Duarte RV, Kelly CJ, Herlihy AH, Kelly MD, Olliff SP, Hübscher SG, Fallowfield JA, Hirschfield GM. Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. Aliment Pharmacol Ther. 2018 Mar;47(5):631-644. doi: 10.1111/apt.14469.

10. Pavlides M, Banerjee R, Tunnicliffe EM, Kelly C, Collier J, Wang LM, Fleming KA, Cobbold JF, Robson MD, Neubauer S, Barnes E. Multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease severity. Liver Int. 2017 Jul;37(7):1065-1073. doi: 10.1111/liv.13284.

11. Bedossa P; FLIP Pathology Consortium. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. Hepatology. 2014 Aug;60(2):565-75. doi: 10.1002/hep.27173.

5.2. Imaging biomarkers of fatty liver as prognostic factors in oncology

5.2.1 The effect of diffuse liver diseases on the occurrence of liver metastases in cancer patients: a systematic review and meta-analysis*

Summary: Since diffuse liver diseases (including steatosis) may alter liver microenvironment possibly influencing the development of liver metastases, we conducted a metanalysis on the relationship between diffuse liver disease and the occurrence of synchronous and metachronous liver metastases in patients with solid tumors.

Aim: Through a systematic review with meta-analysis, we aimed to assess the effect of the presence of diffuse liver diseases on the risk of having liver metastasis at diagnosis (synchronous) or developing liver metastases after the diagnosis (metachronous) in patients with solid neoplasms.

Methods: Relevant databases were searched for systematic reviews and cross-sectional or cohort studies published since 1990, comparing the risk of liver metastases in patients with and without diffuse liver diseases (including steatosis, chronic hepatitis virus infection, cirrhosis, and fibrosis) in solid cancers patients (excluding liver primary tumors). Primary outcomes were prevalence of synchronous metastases, cumulative risk of metachronous liver metastases and liver metastasis free survival. Overall survival (OS) and cumulative risk of synchronous + metachronous metastases were also considered. The Newcastle Ottawa Scale was used to assess the risk of bias (ROB). Separate meta-analyses were performed, reporting pooled relative risk (RR) for synchronous, and hazard ratios (HR) for metachronous metastases. Subgroup analyses were performed by type of diffuse liver disease, primary cancer site and continent where study was performed. Heterogeneity was evaluated using I2 statistics, and publication bias was assessed.

Results: Nineteen studies were included (n=37591 patients), the majority on colorectal cancer. ROB appraisal results were mixed; major concerns were patients' group selection and different follow-up between groups. Patients with diffuse liver diseases had lower risk of synchronous metastases (RR 0.50 95%CI 0.34-0.76), with higher effect for cirrhosis, and a slightly higher risk of metachronous liver metastases (HR 1.11 95%CI 1.03-1.19). In both cases, overall heterogeneity was considerable (I2 89.4% and 78%). The only result with low heterogeneity for metachronous metastases was a lower risk in patients with vs without viral hepatitis (HR 0.57 95%CI 0.40-0.82). Results were heterogeneous when stratifying by cancer site but not by country. Only 2 studies allowed to evaluate the cumulative incidence of both synchronous ad metachronous metastases on the same patients' groups, showing lower risk in patients with viral hepatitis. A publication bias was

suggested toward the studies reporting lower risk for patients with diffuse liver disease, while it is unlikely that differences in OS had introduced a serious bias.

Conclusions: Diffuse liver diseases seem to be protective for synchronous liver metastases. A slight protective effect was also found on metachronous liver metastases for viral hepatitis, while the presence of other diffuse liver diseases had no effect or resulted in a slight increase in risk of metachronous liver metastases.

***** *Manuscript to be submitted to Cancers*. Monelli F, Besutti G, Djuric O, Bonvicini L, Damato A, Bonelli C, Bassi MC, Farì R, Bonfatti S, Ligabue G, Pattacini P, Pinto C, Giorgi Rossi P. The effect of diffuse liver diseases on the occurrence of liver metastases in cancer patients: a systematic review and metanalysis.

INTRODUCTION

In most cases, cancers cause death through the growth of distant metastases in vital organs [1]. Due to its double venous system and to the complex lymphatic system, liver is one of the most frequent sites of distant metastases, particularly for colon cancer, for which half of metastases are in the liver, but also for rectal, breast, pancreatic and lung cancers. Thus, liver metastatic disease frequently affects patient's prognosis [2], management and therapeutic choices [3].

Diffuse liver diseases include various pathological conditions, some of them, such as liver steatosis, fibrosis and cirrhosis, very common worldwide. In fact, steatosis is the hallmark of non-alcoholic fatty liver disease (NAFLD), that has an overall global prevalence of 25% [4], while fibrosis and cirrhosis are the final results of different insults to the liver, including also chronic viral infections that reach a prevalence near to 5% [5].

In comparison to general population, cancer patients have at least the same probability of having diffuse liver diseases, probably also higher given the known association between metabolic syndrome, frequently characterized by liver steatosis, and cancer development [6]. Moreover, cancer patients have a higher risk of developing diffuse liver diseases as a consequence of therapies.

In the latest years, both pathogenetic research and the new therapeutic regimen for cancer patients have focused more on tumor and organ microenvironment than on the cancer cells themselves [7]. Since the microenvironment of the target organ is recognized as a central factor in the process of forming organ metastases, it is reasonable to think that diffuse liver diseases may have an impact on the occurrence of liver metastases. In fact, the ability of metastatic cells to survive and proliferate in

the liver is determined by the outcome of complex interactions between tumor and preexisting tissue cells, including the sinusoidal endothelium, stellate, Kupffer, and inflammatory cells [8]. This interplay may be modified by diffuse liver diseases, that influence liver microenvironment potentially favoring or either hindering the development of hepatic metastases [9, 10]. Similarly, also post-chemotherapy liver modifications, although temporary in some cases [11], may influence the probability of developing liver metastases.

Some studies analyzed the association between the presence of diffuse liver diseases and liver metastasis occurrence in cancer patients, but literature on the topic appears to be fragmented and somewhat contradictory.

Through this systematic review, we aimed to assess the effect of the presence of different diffuse liver diseases on the risk of having liver metastasis at diagnosis (synchronous) or developing liver metastases after the diagnosis (metachronous) in patients with solid neoplasms, excluding hepatic primary tumors.

MATERIALS AND METHODS

Study eligibility

Eligible studies were both systematic reviews and cross-sectional or cohort studies assessing if the risk and the timing of developing liver metastases in patients with solid cancers is different in patients with and without chronic liver injury. Studies specifically addressing recurrence of liver metastases after a R0 (i.e. without residual disease) liver resection for metastases of solid neoplasms in patients with and without liver injury at the moment of resection were included. Complete protocol has been registered in the PROSPERO database (ID CRD42019133519).

Considered diffuse liver diseases were liver steatosis, chronic viral hepatitis or chronic hepatitis virus infection, liver cirrhosis, and liver fibrosis. Considered primary tumors were all solid tumors excluding liver primary neoplasms.

Included studies should report a direct comparison between two solid neoplasm patients' groups: patients with chronic liver injury (exposed) and without chronic liver injury (non-exposed).

Exclusion criteria were hematologic neoplasms, primary liver neoplasms, absence of comparison between different exposure levels, animal studies, studies on solid malignancies with a follow up under 12 months in case of lung and pancreatic cancer or 24 months for every other type of tumors.

Studies reported only as abstracts or published in languages other than English, German, Spanish, French and Italian were excluded.

Outcomes and rationale of the comparisons

Primary endpoints that have been considered were presence of liver metastases at diagnosis (synchronous), development of liver metastases during follow-up (metachronous); for synchronous metastases the outcome was computed as prevalence of cases with liver metastases on total diagnosed cases; for metachronous metastases the considered outcomes were liver metastasis free survival (measured as hazard ratio from survival times) and cumulative risk of liver metastases (measure as proportion after a fixed follow up time or rate using person/time as denominator). Other considered outcomes were overall survival, considered only in studies reporting also liver metastasis free survival, and cumulative risk of metastases summarizing synchronous and metachronous, in studies in which both analyses were explicitly conducted on all incident patients coming from the same population base for the same time period. Overall survival was considered only to evaluate the possible bias introduced by differences in this outcome between patients with and without liver disease on the accuracy of liver metastasis free survival: in fact, differences in overall survival could introduce a bias due to competitive mortality. On the other hand, the cumulative incidence of synchronous and metachronous metastases was included to assess if differences in prevalence of synchronous metastases between patients with and without liver disease could be due to a difference in the detection of prevalent metastases. If the presence of liver disease could affect the probability of detecting liver metastases at diagnoses, because of the different number, type and accuracy of tests performed in patients with or without liver disease, undetected metastases would occur in follow up as metachronous metastases. Thus, by comparing the sum of metastases occurring at diagnosis and in follow up we should overcome this possible bias, if patients included in the two analyses represents an unselected sample of all incident cancers in the same period.

Study search and selection

A systematic search was conducted in MEDLINE, The Cochrane Library, EMBASE and Scopus, adapting the search algorithm to the requirement of each database including studies from 1990. This date limit was introduced because of the presence of a certain heterogeneity in “chronic liver injury” definition, which is more relevant in studies before 1990 [12]. This is particularly important for chronic viral infection, since hepatitis C virus was discovered in 1989 [13] and first clinical diagnostic tests were developed in 1990 [14].

The last search was conducted in September 2019; the search algorithm designed for MEDLINE was (("Hepatitis, Alcoholic"[Mesh terms] OR "Fatty Liver"[Mesh terms] OR "Hepatitis, Chronic"[Mesh]

OR "Hepatitis, Viral, Human"[Mesh terms] OR "Liver Cirrhosis, Alcoholic"[Mesh terms] OR "Liver Cirrhosis, Biliary"[Mesh terms] OR "Liver Disease"[title] OR "fatty liver"[title] OR Steatohepatitis[title] OR Steatosis[title] OR Steatoses[title] OR "Steatotic Liver"[title] OR "nonalcoholic fatty liver disease"[title] OR "non-alcoholic fatty liver disease"[title] OR NAFLD[title] OR "liver fibrosis"[title] OR "chronic hepatitis"[title] OR cirrhosis[title]) AND ("Neoplasm Metastasis"[Mesh] OR "Neoplasm Recurrence, Local"[Mesh] OR liver metastasis[title] OR liver metastases[title] OR metastatic liver[title] OR "Liver Neoplasms/secondary"[Mesh] OR recurrence[title] OR "Cancer Prognosis"[Title])).

One reviewer (FM) screened the search results based on title/abstract; a second reviewer (GB) screened a computer-generated random sample of 25% of the references to identify potential disagreement, resolved by consensus. Then, two reviewers (FM and GB) independently examined eligibility based on the full text of the relevant articles in a two-step procedure: initially removing articles not pertinent to the research question and secondly removing articles without a specific analysis of eligibility criteria and considered outcomes. In case of disagreement, inclusion was decided by group consensus involving a third reviewer (PGR).

Data extraction and synthesis

Two reviewers (FM and RF) extracted data on study design, country, objective, population (number and characteristics of included patients and controls), how the diagnosis of diffuse liver disease was performed, tumor type, outcomes, inclusion and exclusion criteria, presence, type and length of follow-up and results. Differences between reviewers were resolved by consensus and, when this was not possible, by a third reviewer (GB). These data were collected in a pre-designed data extraction sheet. The Newcastle Ottawa Scale [17] for risk of bias assessment in cohort studies was applied to selected studies by two reviewers (FM and SB) to assess the risk of bias by consensus, with a third reviewer involved in case of disagreement (PGR). In absence of a universally accepted assessment tool for cross sectional study design, to evaluate them we chose to remove non relevant fields from the NOS.

Statistical analysis

Descriptive analysis was done to summarize the distribution of synchronous and metachronous metastases among exposed and non-exposed patients as well as to summarize the survival among studies reporting metachronous metastases in order to assess whether survival bias could influence results.

Separate meta-analysis were performed for synchronous and metachronous metastases. For synchronous metastases relative risks (RR) were combined and the pooled RRs with the 95% CI was calculated while for the metachronous metastases hazard ratios (HR) with 95%CI were calculated. Both types of point estimates were calculated using the random-effects model described by DerSimonian and Laird. Missing HR and standard errors were imputed using methods for incorporating summary time-to-event data into meta-analysis, provided by Tierney et al [16].

Subgroup analyses were performed to investigate the risk of metastases by type of liver injury, primary cancer origin and continent where study was performed. In all analyses which were not divided by type of liver injury, only one liver injury per study was considered, i.e., if one study reported results for more than one type of liver injury, the classification considered for the primary objective of the study was included in the analysis, and in case of no clear definition of the main objective, the liver disease with the highest number of patients was included.

Forest plots were used to display the RRs or HR and corresponding 95% CI. Heterogeneity among the studies was evaluated using I^2 statistics. Values of I^2 can be interpreted as not important (0-40%), moderate (30-60%), substantial high (50-90%), considerable (75-100%) levels of heterogeneity [17]. The possibility of a publication bias was assessed visually using a funnel plot for asymmetry. Meta-analyses were performed using STATA 13.0, metan command.

RESULTS

Study selection and characteristics of the included studies

Study selection according to PRISMA flow diagram is reported in Figure 1.

Nineteen studies met inclusion criteria and were included in synthesis, for a total of 37591 patients, 6868 exposed and 29992 non-exposed (Table 1). Among the 19 included studies [18-36], 5 reported overall survival [21, 22, 29, 31, 33]. The majority of the selected studies focused on liver metastases from colorectal cancer, while four focused on other cancer, breast cancer, non-small cell lung cancer, pancreatic cancer and nasopharyngeal carcinoma. Analysing the aetiology of liver disease, among studies on colorectal cancer, three are on liver cirrhosis, four on liver steatosis, eight on hepatic viruses, one on liver fibrosis; two studies include two aetiologies, analysing both steatosis and cirrhosis in one case and hepatic viruses in the other. Among non-colorectal cancer studies, two are on hepatic viruses, pancreatic cancer and nasopharyngeal carcinoma, and two on hepatic steatosis, breast cancer and non-small cell lung cancer. Sixteen studies are on first occurrence of metastasis from primitive cancer and three are on metastasis recurrence after R0 liver metastasis resection, two

on colorectal cancer metastasis to steatotic liver and two on colorectal cancer metastasis to liver with chronic hepatic virus infection.

From twelve studies it was possible to extract quantitative information on synchronous liver metastases and from eleven studies it was possible to extract quantitative information on metachronous liver metastases; four works provided information on both metastases types, but only two of these were studies on all incident cases (Qian 2014 [29], Zeng 2013 [31]), while the other two included patients after liver metastasis resection.

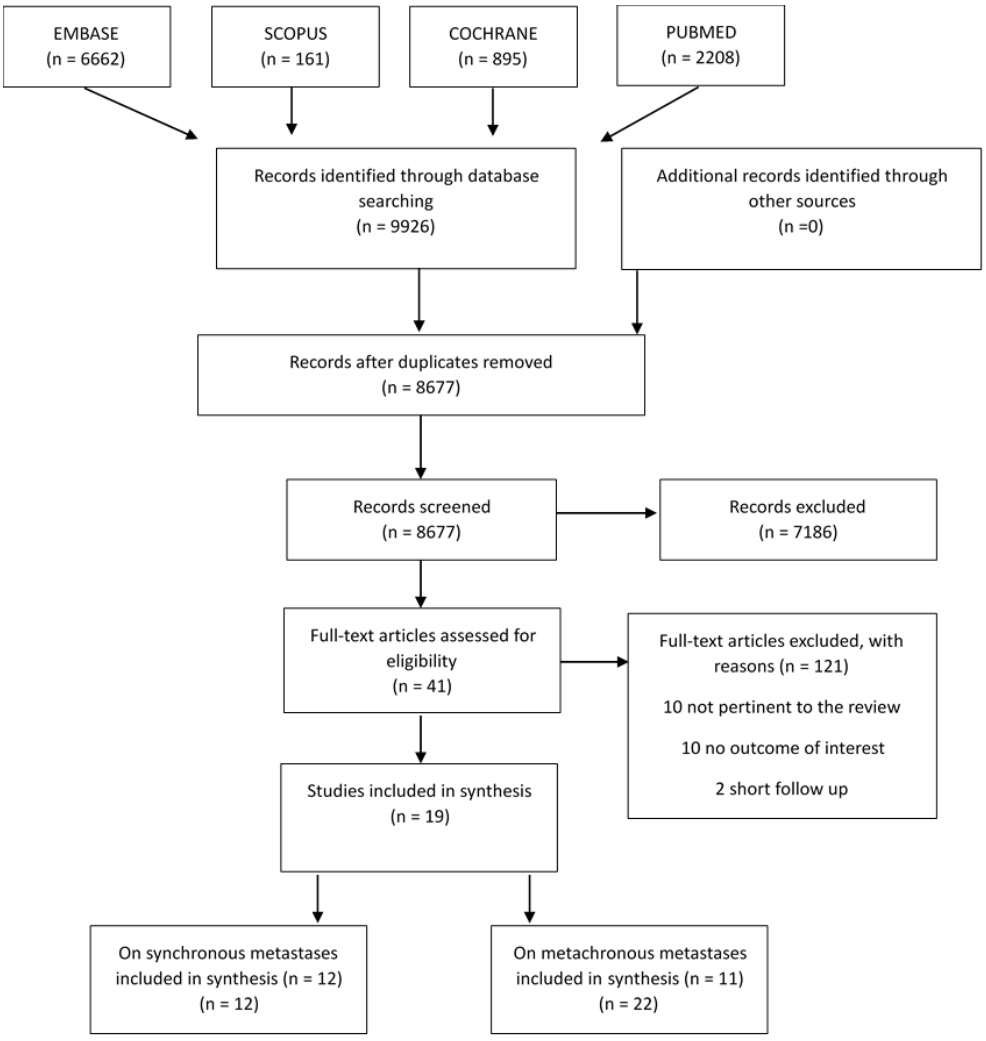


Figure 1: PRISMA flowchart representing records identified through literature search, screened, and included in synthesis.

Author / Year	Primary cancer site	Diffuse liver disease	Study design	Outcome	Primitive/ R0	Total	Exposed	Non-Exposed	LD diagnosis method	LD diagnosis timing	Follow-up EXP	Follow-up nonEXP	Age EXP	Age non-EXP	SYN EXP	SYN nonEXP	MET EXP	MET nonEXP
Chiou 2014 [18]	CRC	Cirrhosis	Cohort	MET	Primitive	14865	2973	11892	National register: ICD-9-CM codes 571.2 (alcoholic cirrhosis of liver), 571.5 (cirrhosis of liver without mention of alcohol) and 571.6 (biliary cirrhosis)	Before prim. surgery	-	-	67.6	67.6	-	-	516	2013
Uetsuji 1992 [19]	CRC	Cirrhosis	Cross-Sectional	SYN	Primitive	250	46	204	bilirubin>1.5, albumin<3.5, cholinesterase<3500, GGT>46, IG retention rate at 15'>15%	At prim. surgery	-	-	-	-	0	40	-	-
Iascone 2005 [20]	CRC	Cirrhosis	Cross-Sectional	SYN	Primitive	576	171	405	Biopsy	At prim. surgery	-	-	71.2	65.8	8	174	-	-
Iascone 2005 [20]	CRC	Steatosis	Cross-Sectional	SYN	Primitive	576	33	543	Biopsy	At prim. surgery	-	-	-	-	5	174	-	-
Hamady 2012 [21]	CRC	Steatosis	Cohort	MET	R0	2715	927	1788	Biopsy	At LM resection	34 months		>75: 16%	>75: 14,8 %	-	-	47.1 %	38.1 %
Ramos 2015 [22]	CRC	Steatosis	Cohort	SYN/MET	R0	943	421	513	Biopsy	At LM resection	47,05 months		62.6	62.9	194	317	48.0 %	47.9 %
Hayashi	CRC	Steatosis	Cohort	SYN	Primitive	839	121	718	Ultrasound	At prim.	-	-	58.5 ± 9.6	61.4 ± 12.4	2	115	-	-

1997 [23]										diagnosis								
Murono 2013 [24]	CRC	Steatosis	Cohort	SYN/MET	Primitive	604	63	541	CT (HU liv/spl<1.1)	Before prim. surgery	-	-	65 ± 11.3	67.2 ± 10.8	2	52	1	58
Uetsuji 1992 [19]	CRC	Virus	Cross-Sectional	SYN	Primitive	250	76	174	Blood Test	At prim. surgery	-	-	-	-	4	36	-	-
Kin Pan Au 2018 [25]	CRC	Virus	Cohort	SYN	R0	304	21	283	Blood Test	Before LM resection	-	-	61 (55–66)	60 (52–68)	13	172	-	-
Huo 2018 [26]	CRC	Virus	Cross-Sectional	SYN	Primitive	4033	244	3789	Blood Test	At prim. diagnosis	-	-	57,25	62,69	38	326	-	-
Iascone 2005 [27]	CRC	Virus	Cross-Sectional	SYN	Primitive	630	87	543	Blood Test	Before prim. surgery	-	-	69.8	65.8	15	174	-	-
Li Destri 2013 [28]	CRC	Virus	Cohort	MET	Primitive	488	31	457	Blood Test	Before prim. diagnosis	9 years		61	66	-	-	1	43
Qian 2014 [29]	CRC	Virus	Cohort	SYN/MET	Primitive	1413	138	1275	Blood Test	Before prim. surgery	72.3 months		58,5	59,2	7	126	6	179
Wang 2012 [30]	CRC	Virus	Cross-Sectional	SYN	Primitive	354	70	284	Blood Test	-	-	-	52,9	56,2	2	48	-	-
Zeng 2013 [31]	CRC	Virus	Cohort	SYN/MET	Primitive	2868	373	2495	Blood Test	-	-	-	57	61	14	197	19	268
Kondo 2016 [32]	CRC	Fibrosis	Cohort	MET	Primitive	953	77	876	NFS>0.676	-	51.2 months		75,3	64,9	-	-	8	46

Wu 2017 [33]	Breast	Steatosis	Cohort	MET	Primitive	1230	372	858	Ultrasound	At prim. diagnosis	30.7 months	32.4 months	>50: 49.7 %	>50: 34.7 %	-	-	27	96
Wu 2019 [34]	Lung	Steatosis	Cohort	MET	Primitive	1873	408	1465	Ultrasound	At prim. diagnosis	14.5 months		>60: 275	>60: 741	-	-	58	108
Wei 2013 [35]	Pancreas.	Virus	Cohort	SYN	Primitive	460	63	397	Blood Test	At prim. diagnosis	12 months		-	-	29	127	-	-
Li 2019 [36]	NPC	Virus/Steatosis	Cohort	MET	Primitive	1367	123	492	Blood Test	At prim. diagnosis	27.8 months		≥ 60: 9.8 %	≥ 60: 11.4 %	-	-	13	51
TOT CRC						32661	5872	26780										
TOT NON-CRC						4930	966	3212										
TOT						37591	6868	29992										

Table 1: Synopsis of the included studies subdivided by type of diffuse liver disease and reporting primary cancer site, considered outcomes, number of patients included and number of patients in exposed and non-exposed groups. Primitive/R0 refers to patients' condition at the moment of inclusion (at the diagnosis of primitive cancer or after R0 liver metastasis resection). CRC, colorectal cancer; NPC, nasopharyngeal cancer; MET, metachronous liver metastases; SYN, synchronous liver metastases; TOT, total number of patients; EXP, exposed patients; non-EXP, non-exposed patients,; NFS, NAFLD fibrosis score; LM, liver metastasis.

Risk of bias

The results of the consensus appraisal according to NOS and modified NOS tools are presented in Table 2. The scores attributed to the selection domain were generally high, reflecting a correct selection of exposed and unexposed cohort and a clear definition of exposure, while some concern for studies evaluating metachronous liver metastases derived from the uncertain absence of metastases at study start. As for comparability domain, the major concern was the unclear addressing of confounders which can undermine the comparison between groups, this being truer for hospital-based rather than population-based studies. Finally, scores of the outcome domain were frequently lowered by inadequate assessment of the occurrence of the metastases and duration of the follow-up.

Author / Year	ROB	SELECTION	Representativeness of exposed cohort	Selection of non exposed cohort	Ascertainment of exposure	Outcome of interest not present at study start	COMPARABILITY	Comparability on the basis of the design or analysis	OUTCOME	Assessment of outcome	Follow-up long enoughfor outcomes	Adequacy of follow up
Chiou 2014 [18]	5/9	3/4	☼	☼	☼		1/2	☼	1/3	☼		
Uetsuji 1992 [19]	4/6	3/3	☼	☼	☼		1/2	☼	0/1			
Iascone 2005 [20]	4/6	3/3	☼	☼	☼		1/2	☼	0/1			
Hamady 2012 [21]	8/9	4/4	☼	☼	☼	☼	2/2	☼☼	2/3	☼	☼	
Ramos 2015 [22]	8/9	4/4	☼	☼	☼	☼	2/2	☼☼	2/3	☼	☼	
Hayashi 1997 [23]	7/9	3/4	☼	☼	☼		1/2	☼	3/3	☼	☼	☼
Murono 2013 [24]	6/9	4/4	☼	☼	☼	☼	2/2	☼☼	0/3			
Kin Pan Au 2018 [25]	9/9	4/4	☼	☼	☼	☼	2/2	☼☼	3/3	☼	☼	☼
Huo 2018 [26]	6/6	3/3	☼	☼	☼		2/2	☼☼	1/1	☼		
Iascone 2005 [27]	4/6	3/3	☼	☼	☼		1/2	☼	1/1	☼		
Li Destri 2013 [28]	7/9	3/4	☼	☼	☼		1/2	☼	3/3	☼	☼	☼
Qian 2014 [29]	6/9	3/4	☼	☼	☼		0/2		3/3	☼	☼	☼
Wang 2012 [30]	4/6	2/3	☼	☼			1/2	☼	1/1	☼		
Zeng 2013 [31]	6/9	3/4	☼	☼	☼		2/2	☼☼	1/3	☼		
Kondo 2016 [32]	3/9	1/4			☼		0/2		2/3	☼	☼	
Wu 2017 [33]	8/9	4/4	☼	☼	☼	☼	1/2	☼	3/3	☼	☼	☼
Wu 2019 [34]	7/9	4/4	☼	☼	☼	☼	1/2	☼	2/3	☼		☼
Wei 2013 [35]	6/9	4/4	☼	☼	☼	☼	0/2		2/3	☼		☼
Li 2019 [36]	9/9	4/4	☼	☼	☼	☼	2/2	☼☼	3/3	☼	☼	☼

Table 2: Appraisal of included guidelines using NOS tool for cohort and case-control studies and modified NOS for cross-sectional studies. T, primary cancer site; CRC, colorectal cancer; NPC, nasopharyngeal cancer; MET, metachronous liver metastases; SYN, synchronous liver metastases.

Meta-analyses

Synchronous metastases

Meta-analysis of the overall risk of having synchronous metastases in patients with vs without diffuse liver diseases showed lower risk among patients with liver disease (RR 0.50 95%CI 0.34-0.76) with considerable overall heterogeneity (I^2 89.4%) (Figure 2). The difference between exposed and non-exposed was higher for patients with cirrhosis (RR 0.14 95%CI 0.07-0.27, I^2 0%) than for those with steatosis (RR 0.37 95%CI 0.15-0.93, I^2 80%). Although this result was compatible with random fluctuations, also patients with viral hepatitis had a slightly lower risk of synchronous metastases when compared with patients without viral hepatitis (RR 0.68 95%CI 0.42-1.10). In this last case heterogeneity was important (I^2 86.9%) with the three largest studies going in the opposite direction (Huo 2018 [26] colorectal cancer; Wei 2013 [35] pancreatic cancer) or showing no effect (Kin Pan Au 2018 [25] colorectal cancer).

When stratified by primary cancer site, the risk of developing synchronous metastases was higher for patients without liver injury if they had colorectal cancer as primary cancer site (RR 0.44 95%CI 0.29-0.69, I^2 88.5%). Only one study reported results on pancreatic cancer as the primary cancer site (Wei 2013 [35]) and showed a higher risk of synchronous metastases for those with liver injury (RR 1.44, 95%CI 1.06-1.95) (Figure 3).

When stratified by the study origin, pooled risk estimates obtained in the studies conducted in and out of Asia were similar to the overall estimate (Figure 4).

Metachronous metastases

Overall, people with diffuse liver diseases had a slightly higher risk of developing metachronous metastases (HR 1.11 95%CI 1.03-1.19), with considerable overall heterogeneity (I^2 78%) (Figure 5). However, the only result with low heterogeneity and showing a more convincing effect was on the opposite direction, i.e. a lower risk in patients with vs without viral hepatitis (HR 0.57 95%CI 0.40-0.82, I^2 0%). Pooled results for steatosis showed a slight increase in risk among exposed patients, with high heterogeneity (HR 1.12 95%CI 1.01-1.25, I^2 80.9%), and similar results were obtained for cirrhosis (HR 1.15 95%CI 1.04-1.28) and fibrosis (HR 2.87 95%CI 1.17-7-03), although estimated in one study only.

When stratifying by cancer site, results were heterogeneous between groups. Pooled analysis was possible only for colorectal cancer (HR 1.12 95%CI 1.04-1.21, I^2 78.7%), while for other cancer sites the risk of metachronous liver metastases in patients with liver disease was either higher (HR 1.43 95%CI 1.02-2.01 for lung cancer) or lower (HR 0.55 95%CI 0.35-0.86 for breast). For nasopharyngeal cancer the study was underpowered and inconclusive (HR 0.77 95%CI 0.38-1.58) (Figure 6).

When stratified by the study origin, pooled risk estimates obtained in the studies conducted in and out of Asia were similar to the overall estimate, without significant heterogeneity between groups (Figure 7).

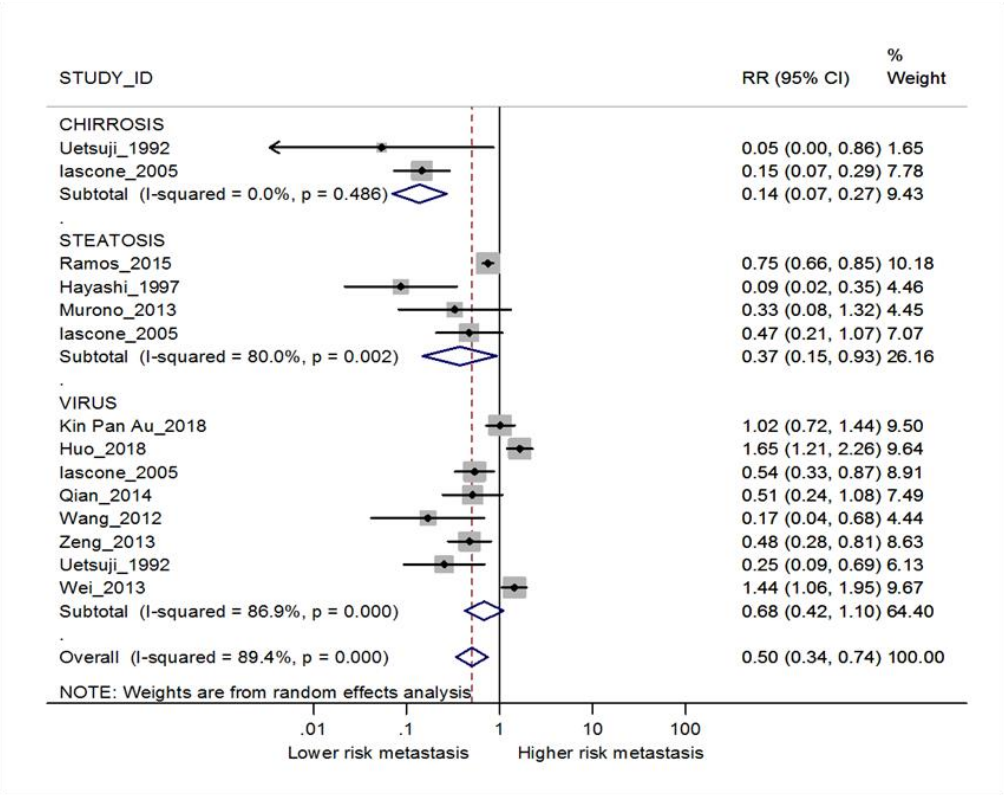


Figure 2: Forest plot displaying the RRs and corresponding 95% CI for synchronous metastases in exposed vs non-exposed patients for different diffuse liver diseases. I-squared is reported as a measure of heterogeneity. Overall RR and I-squared are calculated after excluding the comparison between patients with and without steatosis for Iascone 2005 [20] since this was the less represented liver disease in the study.

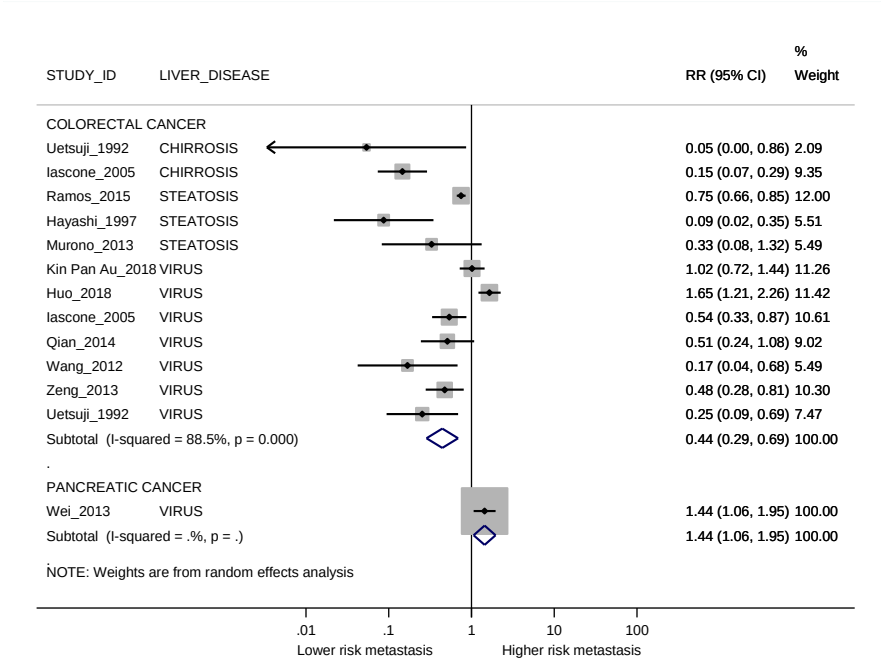


Figure 3: Forest plot displaying the RRs and corresponding 95% CI for synchronous metastases in exposed vs non-exposed patients stratified by cancer sites. I-squared is reported as a measure of heterogeneity. The comparison between patients with and without steatosis for Iascone 2005 [20] was not considered since this was the less represented liver disease in the study.

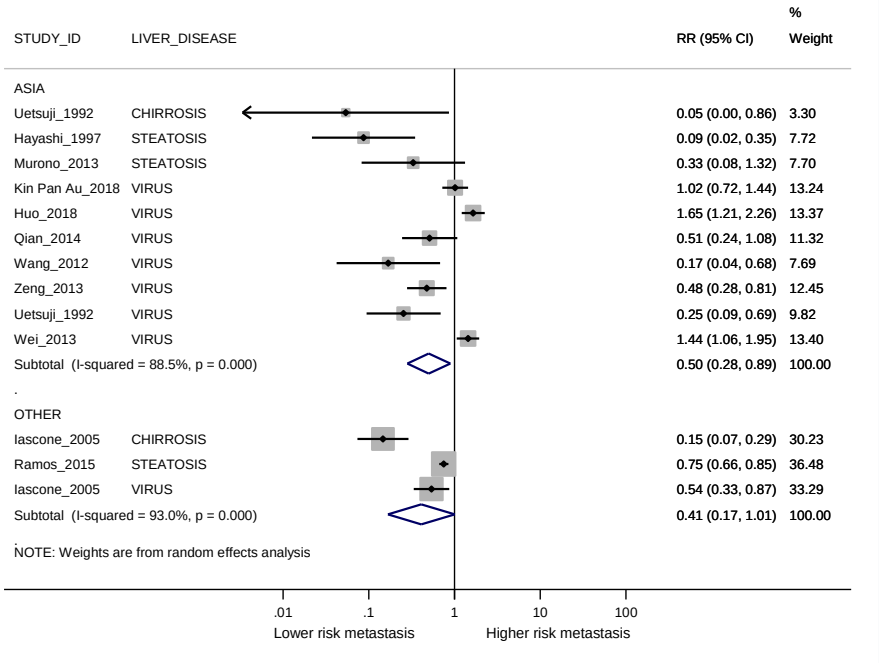


Figure 4: Forest plot displaying the RRs and corresponding 95% CI for synchronous metastases in exposed vs non-exposed patients stratified by country. I-squared is reported as a measure of

heterogeneity. The comparison between patients with and without steatosis for Iascone 2005 [20] was not considered since this was the less represented liver disease in the study.

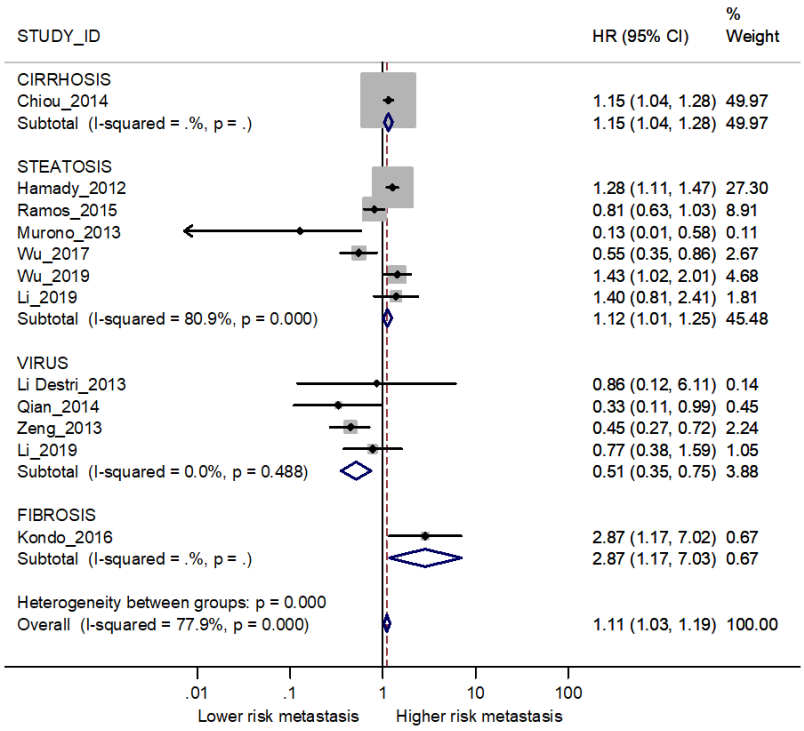


Figure 5: Forest plot displaying the HRs and corresponding 95% CI for metachronous metastases in exposed vs non-exposed patients for different diffuse liver diseases. I-squared is reported as a measure of heterogeneity. Overall RR and I-squared are calculated after excluding the comparison between patients with and without steatosis for Li 2019 [36] since this was a secondary outcome in the study.

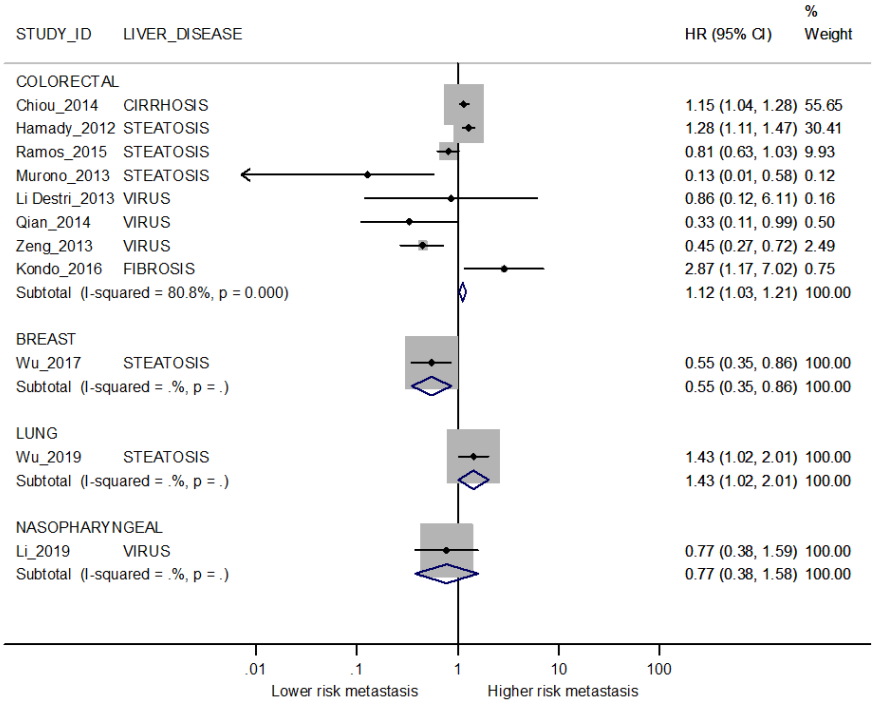


Figure 6: Forest plot displaying the HRs and corresponding 95% CI for metachronous metastases in exposed vs non-exposed patients stratified by cancer site. I-squared is reported as a measure of heterogeneity. The comparison between patients with and without steatosis for Li 2019 [36] is not considered since this was a secondary outcome in the study.

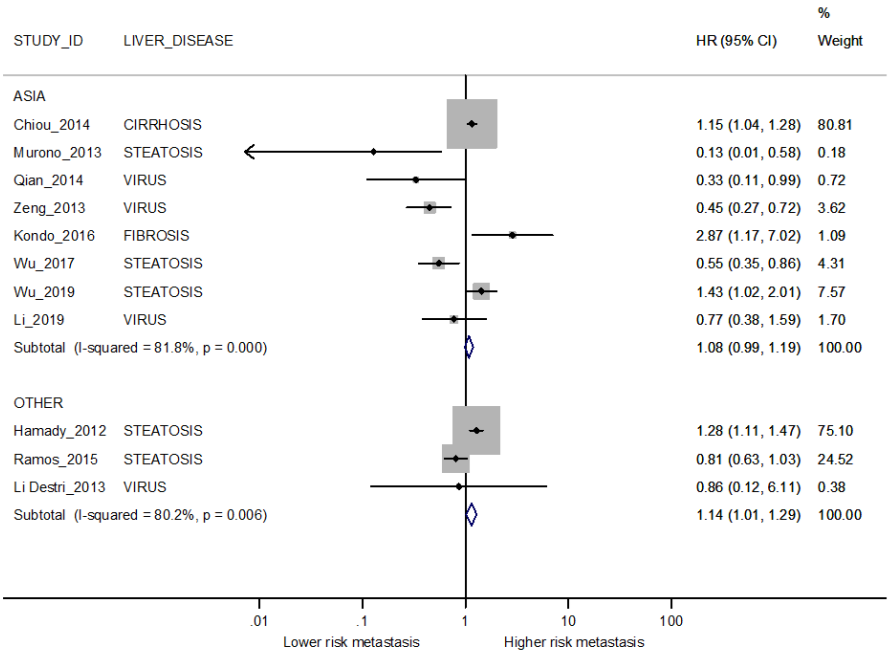


Figure 7: Forest plot displaying the HRs and corresponding 95% CI for metachronous metastases in exposed vs non-exposed patients stratified by country. I-squared is reported as a measure of heterogeneity. The comparison between patients with and without steatosis for Li 2019 [36] is not considered since this was a secondary outcome in the study.

Studies evaluating both synchronous and metachronous liver metastases

Two studies allowed to evaluate the cumulative incidence of both synchronous ad metachronous liver metastases on the same patients groups, with and without viral hepatitis (Qian 2014 [29] and Zeng 2013 [31]). Both showed lower risk of metastases in patients with liver disease: in Qian 2014 [29] cumulatively 9.7% vs. 25.5% and in Zeng 2013 [31] 9.0% vs. 19.6% (Figure 8).

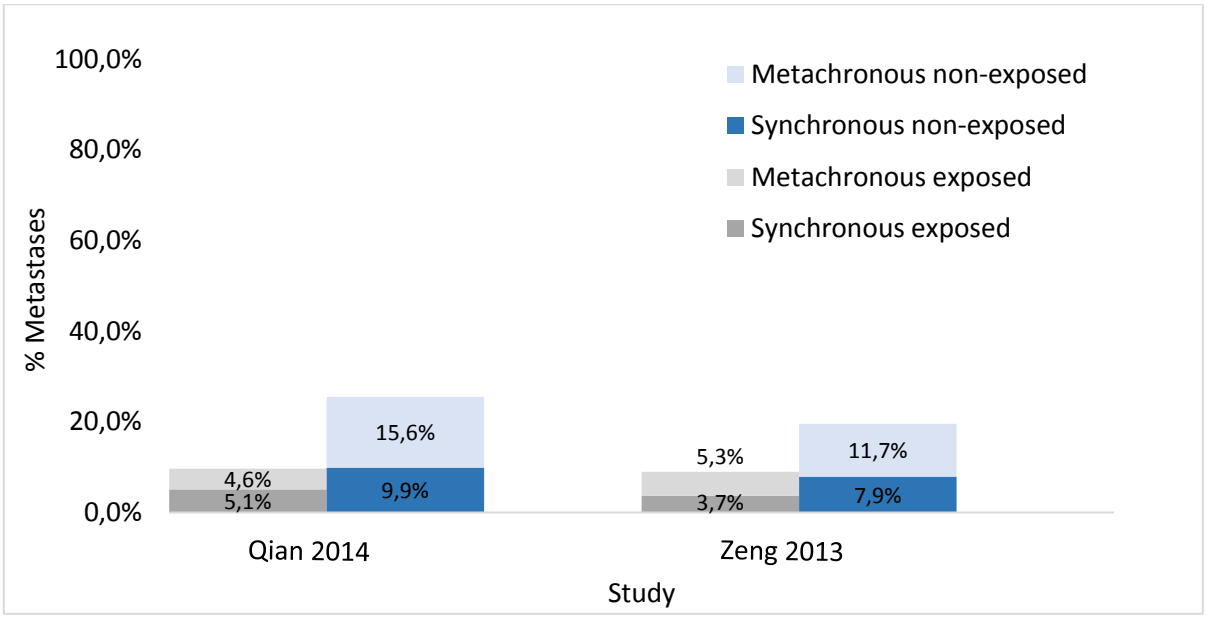
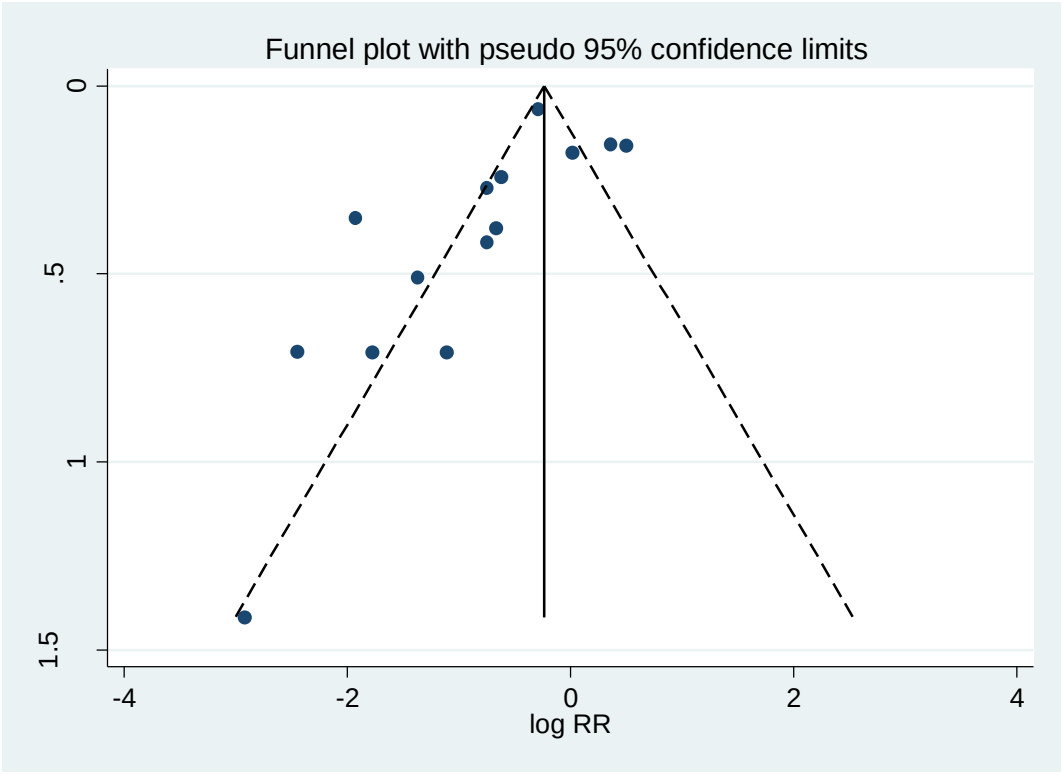


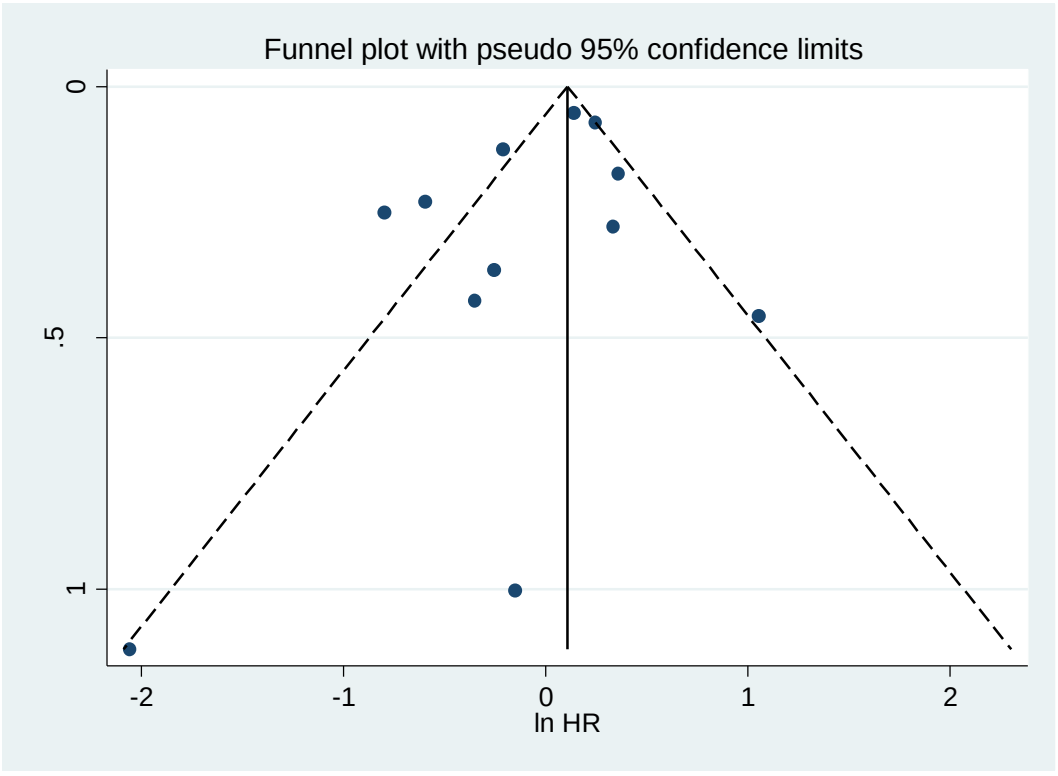
Figure 8: Graph representing the percentages of synchronous and metachronous liver metastases occurred in exposed and non-exposed patients in studies evaluating both types of metastases.

Publication bias

The funnel plot for both synchronous and metachronous metastases was not symmetrically distributed and several studies were outside of the 95%CI (Figure 9) suggesting a publication bias toward the studies reporting lower risk for patients with diffuse liver disease.



A



B

Figure 9: Funnel plots with pseudo 95% confidence intervals for studies on synchronous (A) and metachronous (B) liver metastases.

Overall survival in studies reporting liver disease free survival

Five-year overall survival ranged from 39.3% to 79.8% among exposed and from 39.3% to 92.2% among non-exposed patients. Two studies (Hamady 2012 [21]; Kondo 2016 [32]) reported lower 5-year survival among exposed patients while three studies (Ramos 2015 [22]; Qian 2014 [29]; Zeng 2013 [31]) reported higher 5-year survival among non-exposed patients (Table 3). Given the slight differences with opposite directions in overall survival within studies and compatible with random fluctuations, it is not likely that difference in overall survival could introduce a serious bias in the results of meta-analyses for liver metastasis survival.

Study	**Exposed**		**Non-exposed**	
	5-year survival	**Median survival**	**5-year survival**	**Median survival**
Hamady 2012 [21]	39.3%	22 months	42.8%	24 months
Ramos 2015 [22]	55.1%		45.2%	
Qian 2014 [29]	40.6%		39.3%	
Zeng 2013 [31]		56 months		49 months
Kondo 2016 [32]	79.8%		92.2%	

Table 3: Overall survival in exposed and non-exposed patients for studies reporting both overall and liver metastasis free survival.

DISCUSSION

The results of this systematic review and metanalysis show a slight protective effect of diffuse liver diseases on the presence of synchronous liver metastases. Since this protective effect was more apparent for the most severe liver injury (liver cirrhosis), a dose-response relationship can be hypothesized. A slight protective effect was also found on metachronous liver metastases for viral hepatitis, while the presence of other diffuse liver diseases had no effect or resulted in a slight increase in risk of developing metachronous liver metastases. Results of included studies were inconsistent. A considerable heterogeneity was found both overall and when stratifying for type of diffuse liver disease and primary cancer site, while no heterogeneity was found when stratifying for country of origin.

Our analyses suggest a strong publication bias, with smaller studies often supporting extreme results particularly in favor of lower risk of metastases in patients with liver disease. Hospital-based studies may under-estimate the risk of metastases in patients with liver disease due to selection bias. In fact, research hospitals can attract more complex cases, and complexity could be due to the preexisting conditions as liver disease, or to cancer severity. Thus, in hospital-based rather than population-based studies, patients with diffuse liver diseases could be centralized to regional research hospitals at less advanced cancer stages than patients without liver disease.

The high inconsistency and heterogeneity may partially derive by the design and patient inclusion criteria applied by the majority of the studies found in the literature. In fact, most studies assessing metachronous metastases did not evaluate synchronous metastases as well. By excluding patients with baseline liver metastases, the risk is to look at only one part of the picture. For instance, patients with a diffuse liver disease may undergo a higher number and a different type of diagnostic examinations. This may lead to anticipating the diagnosis of a primary tumor, consequently shifting stage at diagnosis towards non-metastatic primary tumor, in exposed patients. This anticipated diagnosis may explain the finding of a protective effect of cirrhosis on synchronous but not on metachronous liver metastases. Besides, this protective effect is seen in colorectal cancer patients, who may benefit from an earlier cancer detection, and not in pancreatic cancer patients, seemingly confirming the hypothesis of an anticipated diagnosis. Of course, for colorectal cancer this hypothesis would be unconvincing in the era of screening, when the vast majority of cancers in the general population are diagnosed in early stage, but it is worth to note that included studies on cirrhosis and synchronous metastases included cancers diagnosed before the implementation of screening programs.

Only two small studies including viral hepatitis allowed to compare cumulative occurrence of metastases from diagnosis to follow up in an unbiased population of patients (Qian 2014 [29]; Zeng 2013 [31]) and do not suggest that differences in synchronous or metachronous metastases are due to differential assessment of prevalent metastases in the two groups, in fact both studies found less metastases at diagnosis and during follow up in patients with viral hepatitis. Unfortunately, no study allowed this comparison for steatosis or cirrhosis.

Previous metanalyses have reported a lower risk of liver metastases in injured livers, however focusing on only one type of diffuse liver disease (cirrhosis) or only one type of primary cancer site (colorectal cancer) [37]. Compared with previous metanalyses, our study pays more attention to the issue of synchronous and metachronous liver metastases, both by conducting clearly separated

analyses, and on the other hand by trying to consider them together for studies allowing it, since the two phenomena can be two communicating vessels.

According to our results, the previously hypothesized mechanisms for a lower risk of liver metastases in injured liver seem to be confirmed only partially, possibly more for viral hepatitis than for other diffuse liver diseases. These include the higher concentration of metalloproteinase inhibitor [38], decreased neovascularization [39], changes in liver-related immunity [40].

This systematic review has some limitations. Firstly, since we excluded papers in Chinese and Japanese, and liver disease prevalence is particularly high in Asia resulting in a particularly rich literature from these countries, we could have missed some relevant papers. Moreover, we considered together studies with different assessment of the diffuse liver disease. Liver steatosis could be diagnosed either by imaging techniques or by pathological examination, while for viral hepatitis some studies included only active disease and others included any serological sign of present or past infection; only one study (Li 2019 [36]) stratified for different types of viral infection status suggesting that inactive or resolved infection but not chronic infection had a protective effect on liver metastases occurrence.

Conclusions

Diffuse liver diseases seem to be protective for synchronous liver metastases. This could be the result of earlier cancer diagnoses due to opportunistic screening in patients treated and followed for liver diseases.

A slight protective effect was also found on metachronous liver metastases for viral hepatitis, while the presence of other diffuse liver diseases had no effect or resulted in a slight increase in risk of developing metachronous liver metastases.

To give a clear answer to the question if and which diffuse liver diseases influence the probability of developing liver metastases, future studies should be population-based, including all incident cases from stage II (at least) to IV, and assess simultaneously, but separately, the prevalence of synchronous metastases and the incidence of metachronous metastases.

References

1 Helling TS, Martin M. Cause of death from liver metastases in colorectal cancer. Ann Surg Oncol. 2014 Feb;21(2):501-6. doi: 10.1245/s10434-013-3297-7. Epub 2013 Oct 1. PMID: 24081807.

2 Zaorsky NG, Churilla TM, Egleston BL, Fisher SG, Ridge JA, Horwitz EM, Meyer JE. Causes of death among cancer patients. Ann Oncol. 2017 Feb 1;28(2):400-407. doi: 10.1093/annonc/mdw604. PMID: 27831506; PMCID: PMC5834100.

3 Stewart CL, Warner S, Ito K, Raoof M, Wu GX, Kessler J, Kim JY, Fong Y. Cytoreduction for colorectal metastases: liver, lung, peritoneum, lymph nodes, bone, brain. When does it palliate, prolong survival, and potentially cure? Curr Probl Surg. 2018 Sep;55(9):330-379. doi: 10.1067/j.cpsurg.2018.08.004. Epub 2018 Oct 4. PMID: 30526930; PMCID: PMC6422355.

4 Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018 Jan;67(1):123-133. doi: 10.1002/hep.29466. Epub 2017 Dec 1. PMID: 28802062; PMCID: PMC5767767.

5 Razavi H. Global Epidemiology of Viral Hepatitis. Gastroenterol Clin North Am. 2020 Jun;49(2):179-189. doi: 10.1016/j.gtc.2020.01.001. PMID: 32389357.

6 Uzunlulu M, Telci Caklili O, Oguz A. Association between Metabolic Syndrome and Cancer. Ann Nutr Metab. 2016;68(3):173-9. doi: 10.1159/000443743. Epub 2016 Feb 20. PMID: 26895247.

7 Langley RR, Fidler IJ. The seed and soil hypothesis revisited--the role of tumor-stroma interactions in metastasis to different organs. Int J Cancer. 2011 Jun 1;128(11):2527-35. doi: 10.1002/ijc.26031. Epub 2011 Mar 25. PMID: 21365651; PMCID: PMC3075088.

8 Brodt P. Role of the Microenvironment in Liver Metastasis: From Pre- to Prometastatic Niches. Clin Cancer Res. 2016 Dec 15;22(24):5971-5982. doi: 10.1158/1078-0432.CCR-16-0460. Epub 2016 Oct 19. PMID: 27797969.

9 Milette S, Sicklick JK, Lowy AM, Brodt P. Molecular Pathways: Targeting the Microenvironment of Liver Metastases. Clin Cancer Res. 2017 Nov 1;23(21):6390-6399. doi: 10.1158/1078-0432.CCR-15-1636. Epub 2017 Jun 14. PMID: 28615370; PMCID: PMC5668192.

10 Masaki S, Hashimoto Y, Kunisho S, Kimoto A, Kitadai Y. Fatty change of the liver microenvironment influences the metastatic potential of colorectal cancer. Int J Exp Pathol. 2020 Oct;101(5):162-170. doi: 10.1111/iep.12371. Epub 2020 Aug 11. PMID: 32783302; PMCID: PMC7495750.

11 Vigano L, De Rosa G, Toso C, Andres A, Ferrero A, Roth A, Sperti E, Majno P, Rubbia-Brandt L. Reversibility of chemotherapy-related liver injury. J Hepatol. 2017 Jul;67(1):84-91.

12 Lonardo A, Leoni S, Alswat KA, Fouad Y. History of Nonalcoholic Fatty Liver Disease. Int J Mol Sci. 2020 Aug 16;21(16):5888. doi: 10.3390/ijms21165888. PMID: 32824337; PMCID: PMC7460697.

13 Kuo G, Choo QL, Alter HJ, Gitnick GL, Redeker AG, Purcell RH, Miyamura T, Dienstag JL, Alter MJ, Stevens CE, et al. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science. 1989 Apr 21;244(4902):362-4. doi: 10.1126/science.2496467. PMID: 2496467.

14 Aach RD, Stevens CE, Hollinger FB, Mosley JW, Peterson DA, Taylor PE, Johnson RG, Barbosa LH, Nemo GJ. Hepatitis C virus infection in post-transfusion hepatitis. An analysis with first- and second-generation assays. N Engl J Med. 1991 Nov 7;325(19):1325-9. doi: 10.1056/NEJM199111073251901. PMID: 1656258.

15 Wells, G., B. Shea, D. L. O'Connell, J. Peterson, Welch, M. Losos, P. Tugwell, Sb Wells Ga, G. Zello and J. Petersen. "The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses." (2014).

16 Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials. 2007 Jun 7;8:16. doi: 10.1186/1745-6215-8-16. PMID: 17555582; PMCID: PMC1920534.

17 Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane, 2020. Available from www.training.cochrane.org/handbook.

18 Chiou WY, Chang CM, Tseng KC, Hung SK, Lin HY, Chen YC, Su YC, Tseng CW, Tsai SJ, Lee MS, Li CY. Effect of liver cirrhosis on metastasis in colorectal cancer patients: a nationwide population-based cohort study. Jpn J Clin Oncol. 2015 Feb;45(2):160-8. doi: 10.1093/jjco/hyu178. Epub 2014 Nov 6. PMID: 25378650.

19 Uetsuji S, Yamamura M, Yamamichi K, Okuda Y, Takada H, Hioki K. Absence of colorectal cancer metastasis to the cirrhotic liver. Am J Surg. 1992 Aug;164(2):176-7. doi: 10.1016/s0002-9610(05)80380-0. PMID: 1636899.

20 Iascone C, Ruperto M, Barillari P. Metastasi epatiche sincrone da carcinoma colo-rettale in pazienti con cirrosi epatica o steatosi [Occurrence of synchronous colorectal cancer metastasis in the cirrhotic or fatty liver]. Minerva Chir. 2005 Jun;60(3):185-90. Italian. PMID: 15985994.

21 Hamady ZZ, Rees M, Welsh FK, Toogood GJ, Prasad KR, John TK, Lodge JP. Fatty liver disease as a predictor of local recurrence following resection of colorectal liver metastases. Br J Surg. 2013 May;100(6):820-6. doi: 10.1002/bjs.9057. Epub 2013 Jan 28. PMID: 23354994.

22 Ramos E, Torras J, Lladó L, Rafecas A, Serrano T, Lopez-Gordo S, Busquets J, Fabregat J. The influence of steatosis on the short- and long-term results of resection of liver metastases from colorectal carcinoma. HPB (Oxford). 2016 Apr;18(4):389-96. doi: 10.1016/j.hpb.2015.12.002. Epub 2016 Jan 29. PMID: 27037210; PMCID: PMC4814618.

23 Hayashi S, Masuda H, Shigematsu M. Liver metastasis rare in colorectal cancer patients with fatty liver. Hepatogastroenterology. 1997 Jul-Aug;44(16):1069-75. PMID: 9261601.

24 Murono K, Kitayama J, Tsuno NH, Nozawa H, Kawai K, Sunami E, Akahane M, Watanabe T. Hepatic steatosis is associated with lower incidence of liver metastasis from colorectal cancer. Int J Colorectal Dis. 2013 Aug;28(8):1065-72. doi: 10.1007/s00384-013-1656-2. Epub 2013 Feb 8. PMID: 23392476.

25 Au KP, Chok KSH, Chan ACY, Dai WC, Cheung TT, Lo CM. Impact of Hepatitis B Carrier Status on the Outcomes of Surgical Treatment of Colorectal Liver Metastases. World J Surg. 2018 Aug;42(8):2642-2650. doi: 10.1007/s00268-018-4483-3. PMID: 29352337.

26 Huo T, Cao J, Tian Y, Shi X, Wu L, Zhang M, Wong LL, Zhao L. Effect of Concomitant Positive Hepatitis B Surface Antigen on the Risk of Liver Metastasis: A Retrospective Clinical Study of 4033 Consecutive Cases of Newly Diagnosed Colorectal Cancer. Clin Infect Dis. 2018 Jun 1;66(12):1948-1952. doi: 10.1093/cid/cix1118. PMID: 29293940.

27 Iascone C, Ruperto M, Barillari P. Metastasi epatiche da carcinoma colo-rettale in pazienti con infezione da epatite B o C [Colorectal carcinoma metastasis in livers infected with hepatitis B or C virus]. Minerva Chir. 2005 Apr;60(2):77-81. Italian. PMID: 15973212.

28 Li Destri G, Castaing M, Ferlito F, Minutolo V, Di Cataldo A, Puleo S. Rare hepatic metastases of colorectal cancer in livers with symptomatic HBV and HCV hepatitis. Ann Ital Chir. 2013 May-Jun;84(3):323-7. PMID: 22722140.

29 Qian HG, Hao CY. Hepatitis B virus infection is an independent factor influencing the occurrence of liver metastasis in colorectal cancer: a retrospective analysis of 1413 cases. Hepatogastroenterology. 2014 Oct;61(135):1908-14. PMID: 25713887.

30 Wang FS, Shao ZG, Zhang JL, Liu YF. Colorectal liver metastases rarely occur in patients with chronic hepatitis virus infection. Hepatogastroenterology. 2012 Jul-Aug;59(117):1390-2. doi: 10.5754/hge11747. PMID: 22172376.

31 Zeng Y, He XW, He XS, Wu XJ, Ma JP, Wang JP, Lan P. Chronic hepatic viral infection could be a protective factor for colorectal cancer liver metastases: analysis in a single institute. Hepatogastroenterology. 2013 Jan-Feb;60(121):37-41. doi: 10.5754/hge12564. PMID: 22944343.

32 Kondo T, Okabayashi K, Hasegawa H, Tsuruta M, Shigeta K, Kitagawa Y. The impact of hepatic fibrosis on the incidence of liver metastasis from colorectal cancer. Br J Cancer. 2016 Jun 28;115(1):34-9. doi: 10.1038/bjc.2016.155. Epub 2016 Jun 9. PMID: 27280634; PMCID: PMC4931372.

33 Wu W, Chen J, Ye W, Li X, Zhang J. Fatty liver decreases the risk of liver metastasis in patients with breast cancer: a two-center cohort study. Breast Cancer Res Treat. 2017 Nov;166(1):289-297. doi: 10.1007/s10549-017-4411-5. Epub 2017 Jul 24. PMID: 28741275.

34 Wu W, Liao H, Ye W, Li X, Zhang J, Bu J. Fatty liver is a risk factor for liver metastasis in Chinese patients with non-small cell lung cancer. PeerJ. 2019 Mar 14;7:e6612. doi: 10.7717/peerj.6612. PMID: 30886784; PMCID: PMC6421062.

35 Wei XL, Qiu MZ, Chen WW, Jin Y, Ren C, Wang F, Luo HY, Wang ZQ, Zhang DS, Wang FH, Li YH, Xu RH. The status of HBV infection influences metastatic pattern and survival in Chinese patients with pancreatic cancer. J Transl Med. 2013 Oct 8;11:249. doi: 10.1186/1479-5876-11-249. PMID: 24099678; PMCID: PMC3851713.

36 Li X, Wu W, Chen J, Ye W, Zhang J, Wei T. Effect of hepatitis B virus infection status on liver metastasis of nasopharyngeal carcinoma: a cohort study. Transl Cancer Res 2019;8(1):262-272. doi: 10.21037/tcr.2019.01.25AB

37 Cai B, Liao K, Song XQ, Wei WY, Zhuang Y, Zhang S. Patients with chronically diseased livers have lower incidence of colorectal liver metastases: a meta-analysis. PLoS One. 2014 Sep 29;9(9):e108618. doi: 10.1371/journal.pone.0108618. PMID: 25265536; PMCID: PMC4181318.

38 Barsky SH, Gopalakrishna R. High metalloproteinase inhibitor content of human cirrhosis and its possible conference of metastasis resistance. J Natl Cancer Inst. 1988 Mar 16;80(2):102-8. doi: 10.1093/jnci/80.2.102. PMID: 2830405.

39 Karube H, Masuda H, Hayashi S, Ishii Y, Nemoto N. Fatty liver suppressed the angiogenesis in liver metastatic lesions. Hepatogastroenterology. 2000 Nov-Dec;47(36):1541-5. PMID: 11148998.

40 Okuno K, Hirai N, Lee YS, Kawai I, Shigeoka H, Yasutomi M. Involvement of liver-associated immunity in hepatic metastasis formation. J Surg Res. 1998 Mar;75(2):148-52. doi: 10.1006/jsre.1998.5272. PMID: 9655087.

5.2.2 Baseline liver steatosis has no impact on liver metastases and overall survival in rectal cancer patients*

Summary: Proceeding in the evaluation of the prognostic value of fatty liver biomarkers in oncologic patients, we designed a retrospective study to evaluate the impact of CT-assessed liver steatosis on liver metastases and overall survival in patients with rectal cancer.

Background: The liver is one of the most frequent sites of metastases in rectal cancer. This study aimed to evaluate how the development of synchronous or metachronous liver metastasis and overall survival are impacted by baseline liver steatosis and chemotherapy-induced liver damage in rectal cancer patients.

Methods: Patients diagnosed with stage II to IV rectal cancer between 2010 and 2016 in our province with suitable baseline CT scan were included. Data on cancer diagnosis, staging, therapy, outcomes and liver function were collected. CT scans were retrospectively reviewed to assess baseline steatosis (liver density <48 HU and/or liver-to-spleen ratio <1.1). Among patients without baseline steatosis and treated with neoadjuvant chemotherapy, chemotherapy-induced liver damage was defined as steatosis appearance, ≥ 10% liver volume increase, or significant increase in liver function tests.

Results: We included 283 stage II to IV rectal cancer patients with suitable CT scan (41% females; mean age 68±14 years). Steatosis was present at baseline in 90 (31.8%) patients, synchronous liver metastasis in 42 (15%) patients and metachronous liver metastasis in 26 (11%); 152 (54%) deaths were registered. The prevalence of synchronous liver metastasis was higher in patients with steatosis (19% vs 13%), while the incidence of metachronous liver metastasis was similar. After correcting for age, sex, stage, and year of diagnosis, steatosis was not associated with metachronous liver metastasis nor with overall survival. In a small analysis of 63 patients without baseline steatosis and treated with neoadjuvant chemotherapy, chemotherapy-induced liver damage was associated with higher incidence of metachronous liver metastasis and worse survival, results which need to be confirmed by larger studies.

Conclusions: Our data suggest that rectal cancer patients with steatosis had a similar occurrence of metastases during follow-up, even if the burden of liver metastases at diagnosis was slightly higher, compatible with chance.

*Besutti G, Damato A, Venturelli F, Bonelli C, Vicentini M, Monelli F, Mancuso P, Ligabue G, Pattacini P, Pinto C, Giorgi Rossi P. Baseline liver steatosis has no impact on liver metastases and

overall survival in rectal cancer patients. BMC Cancer. 2021 Mar 9;21(1):253. doi: 10.1186/s12885-021-07980-9.

Supplementary Materials in Appendix C.

BACKGROUND

The estimated global prevalence of liver steatosis is around 25% and projected to be 33.5% in 2030 [1]. Since non-alcoholic fatty liver disease and metabolic syndrome are associated with increased incidence of colorectal cancer (CRC) [2,3], the expected prevalence of liver steatosis in CRC patients is at least similar or possibly higher than that estimated in the general population. Moreover, cancer patients have a strong risk of developing liver steatosis and steatohepatitis as a consequence of anticancer therapies, especially chemotherapy drugs such as 5-Fluorouracil, an anti-metabolite (anti-pyrimidine), and irinotecan, a cytotoxic anti-tumour molecule of the DNA topoisomerase inhibitor class. This liver side effect is particularly described in patients with metastatic CRC [4,5], who develop steatosis in 30%-47% cases and steatohepatitis in about 20% of those cases treated with irinotecan [6,7].

CRC is the third most incident cancer worldwide in western countries [8]; rectal cancer accounts for about 27% of the total colorectal cancer incidence in Italy [9]. The most frequent sites for metastases in rectal cancer are liver and lung (12.3% and 5.6%, respectively) [10]. In the first decade of 2000, during the implementation of screening programmes in Italy, the proportion of patients presenting with synchronous distant metastases was 16% [9]. Despite the improvement in locally-advanced rectal cancer treatment, including neoadjuvant therapy and standardized surgical treatment, distant metastases after curative-intent treatment are still common (25-40% in the first 5 years), with lung and liver as the most common sites [11-13]. Liver recurrence is one of the most important prognostic factors for rectal cancer patients, influencing survival more than does the occurrence of lung metastases [14]. Personalized biomarkers to correctly stratify the risk of liver recurrence could be useful to improving the management of those patients.

Several studies have investigated the relationship between liver steatosis and liver metastasis occurrence in patients with solid tumours. Microenvironment changes induced by steatosis, such as inflammation and stellate cells activation, may influence neoangiogenesis [14-16] and may either favour or interfere with the process of metastasizing in the liver. In a study of CRC patients who underwent liver resection for synchronous metastases, steatosis was an independent risk factor for liver recurrence [17], while in another study of CRC patients without baseline liver metastases,

steatosis was associated with a lower incidence of liver metastases during follow-up [18]. Discordant results on this topic have also been reported for other primitive cancer sites, including breast cancer and non-small cell lung cancer [19,20].

Studying the complex association between liver steatosis and metastases is even more complicated if we take into account two possible diverging biases induced by steatosis: 1) the masking effect of liver steatosis on the detection of liver metastases, i.e., hypovascular metastases, may be difficult to detect by computed tomography (CT) scan in a fatty liver [21]; 2) hepatic diffuse disease may lead to performing more and different imaging tests, which can increase the probability of detecting liver metastases and thus classifying them as synchronous instead of metachronous. Our hypothesis is that these biases may be responsible for the discordant results found in previous studies.

The aim of this study was to determine whether baseline imaging-defined liver steatosis is a risk factor for liver metastasis occurrence in patients with stages II-IV rectal cancer. We also aimed at evaluating the impact of the presence of baseline liver steatosis on overall survival. Finally, in patients with locally advanced rectal cancer, we evaluated the association between chemotherapy-induced liver damage and liver metastasis occurrence and overall survival after the end of chemotherapy.

METHODS

Study design and population

In this retrospective observational study, all consecutive patients diagnosed with stage II-IV rectal cancer between 2010 and 2016 in the province of Reggio Emilia, with available CT scan performed at the time of diagnosis for staging purposes, were included. Patients with baseline CT not suitable for evaluation of presence/absence of liver steatosis, including unavailability of liver unenhanced CT images or diffuse liver parenchymal derangement, were excluded.

The rationale of including both locally advanced and metastatic patients (stages II-IV) derives from the possible masking effect of liver steatosis on liver metastasis detection. Since liver steatosis may affect the sensitivity of CT scan in the detection of liver metastases, patients with and without liver steatosis may have a different probability of having undetected metastases at baseline, with a consequent possible excess of liver metastasis incidence in the group with undetected metastases. By excluding stage IV patients, we would have introduced a bias in the comparison of cumulative incidence of liver metastases in patients with and without steatosis. This bias is overcome by the comparison of both liver metastasis prevalence at baseline (including stage IV patients) and cumulative incidence during follow-up.

Clinical data

Data on patient health status, rectal cancer diagnosis, staging, therapy and outcomes and liver function tests were extracted from the local population-based Cancer Registry and from the electronic medical records of all the Local Health Authority hospitals of the province, which report all inpatient and outpatients procedures. The Reggio Emilia Cancer Registry includes all malignant cancer cases diagnosed in the Reggio Emilia province since January 1, 1996. Cancer site and morphology are coded according to International Classification of Diseases for Oncology – 3rd Edition (ICD-O-3). Clinical TNM staging for rectal cancer in our centre is based on magnetic resonance (MR) imaging and endorectal ultrasonography (US) for local staging and on CT scan for evaluation of distant metastases, followed by other exams such as contrast-enhanced US, MR or positron emission tomography (PET) if needed. CT images were retrieved from the radiology information system - picture archiving and communication system (RIS-PACS) of the Local Health Authority.

CT scan evaluation of liver steatosis

CT scans performed at the time of diagnosis were retrospectively reviewed by a single radiologist with 7 years of experience in liver imaging, blinded to clinical data and patient outcomes. Mean liver and spleen attenuation values (Hounsfield units, HU) were obtained by placing eight regions of interest (ROIs) in the liver and three ROIs in the spleen [22], being careful to exclude vessels, bile ducts, focal lesions, focal fatty changes or focal fatty sparing and visceral margins. Steatosis was defined as present for absolute liver density < 48 HU and/ or liver-to-spleen ratio < 1.1. A moderate-to-severe degree of steatosis was defined for absolute liver density < 40 HU and/ or liver-to-spleen ratio < 0.8 and/ or liver-spleen difference < -10 HU [23].

Outcome measures

Liver metastases already present at the time of diagnosis and those occurring during a follow-up period of at least 2 years were considered separately. Follow-up duration for the occurrence of metastases was considered from disease diagnosis to the last imaging exam, excluding the presence of metastases. Overall survival (OS) was defined as the time from disease diagnosis to all-cause death or the last clinical evaluation.

Subgroup analysis on post-chemotherapy liver damage

The effect of chemotherapy-induced liver damage was evaluated in a subgroup of patients with locally advanced rectal cancer treated with neoadjuvant chemotherapy and with an available CT scan performed within 4 months from the cessation of chemotherapy, without baseline liver steatosis. At

our institution, the standard neoadjuvant chemotherapy treatment for locally advanced rectal cancer patients includes fluorouracil and oxaliplatin, while irinotecan is rarely used.

By comparing baseline and post-neoadjuvant chemotherapy CT scans and liver function tests, liver damage was defined as at least one of the following: 1) appearance of liver steatosis, 2) liver volume change of ≥ 10%, 3) increase in liver function tests (doubled in values and AST > 40 U/L; ALT > 49 U/L; GGT > 73 U/L).

In this subgroup of patients, baseline and post-neoadjuvant chemotherapy liver volume was retrospectively assessed by a blinded dedicated post-processing technologist supervised by a blinded abdominal radiologist through liver manual segmentation on portal venous phase CT scans.

Patients with baseline liver steatosis were excluded from this analysis since it was impossible to apply one of the three criteria defining damage in those patients.

Follow-up was defined as starting from 4 months after neoadjuvant chemotherapy end date up to event or last imaging or oncologic evaluation for metastasis occurrence, or to death or last update of mortality registry (August 31, 2019) for overall survival.

Ethics

This observational retrospective study was performed in accordance with the Declaration of Helsinki and was approved by the local ethics committee ("Comitato Etico dell'Area Vasta Emilia Nord") with protocol number 2019/0079373. The need for informed consent was waived due to the retrospective nature of the study. Nevertheless, the investigators asked all patients presenting to any of the participating institutions for any clinical reasons for consent.

Statistical analyses

We present the distribution of patient characteristics, therapies received and completeness of follow-up by stage at diagnosis. Cumulative incidence of metastasis is presented in graphs using Nelson-Aalen cumulative hazard function

The association between steatosis and metastasis at baseline was assessed with a logistic model and reported as odds ratio with relative 95% confidence intervals (95% CI). All cases were included in the model. The association between steatosis and metastasis detection during follow-up was assessed with Cox proportional hazards regression model and reported as hazard ratios with relative 95% CI. Only cases free from liver metastases at baseline were included in the model. Models were adjusted for variables identified a priori: age (as continuous variable), sex, calendar period (i.e. year of

diagnosis), and stage. We chose not to adjust for risk factors that may lie on the same causal chain linking liver steatosis and liver metastases.

Median OS was computed by using the reverse Kaplan Meier method. Cox models for OS were constructed as those for liver metastases, but in this case, we performed two analyses, one including and the other excluding patients with liver metastases at baseline.

The association between post-chemotherapy liver damage and incidence of metastases was assessed with an exact logistic model [24]. The model included only cases that were free from steatosis at baseline, had completed chemotherapy, and were free of metastases at the first post-treatment imaging assessment (i.e. a CT performed between February 23, 2011 and May 4, 2017). To select the variables to include in final models, we built a standard logistic model including age, sex, calendar period, and stage: variables that were not associated with the outcome were excluded (p >0.5), while the remaining variables were included in the exact logistic model. Odds ratios (ORs) were compared with hazard ratios (HRs) obtained with similar Cox models in order to exclude that differences in length of follow-up influenced the results.

We did not perform any formal statistical tests; p-values as well as 95% confidence intervals should be interpreted as an indication of the probability that the observed differences occurred under the null hypothesis without a pre-fixed threshold of significance.

RESULTS

Study population

From 2010 to 2016, 465 patients were diagnosed with rectal cancer in the Reggio Emilia province. After excluding patients with stage I rectal cancer and patients with unavailable or unsuitable CT scan at diagnosis, 283 patients were included in the main analysis (Figure 1).

Of these 283 patients, 116 (41%) were females. Mean age was 68 ± 14 years. Disease stage at diagnosis was II in 77 (27%) patients, III in 125 (44%) patients and IV in 72 (25%) patients. For nine patients the stage was unknown but distant metastases were excluded (stage XM0). Patient characteristics in different disease stages are reported in Supplementary Table S1, Additional File.

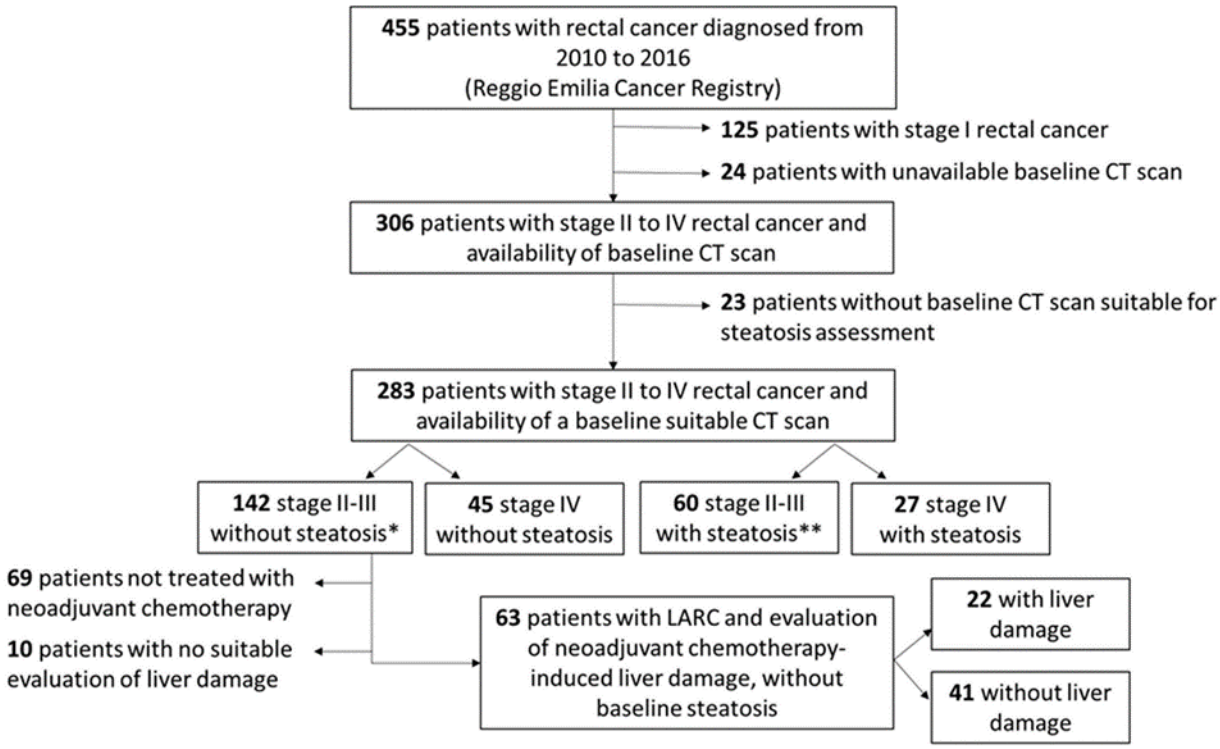


Figure 1: Patient flowchart for the main study and subgroup analysis. * + 6 patients stage XM0; ** + 3 patients stage XM0. TNM stage refers to clinical staging. LARC (locally advanced rectal cancer).

Liver steatosis was present at baseline in 90 (31.8%) patients, of whom 19 (21.1%) had severe steatosis Table 1). Baseline liver metastases were present in 42 (15%) patients, while 26 (11%) patients developed liver metastases during follow-up. Median follow-up duration for liver metastasis detection was 31 months. Median follow-up duration for overall survival was 45 months. The overall number of deaths occurred during follow-up was 152 (54%) and among them 27 (18%) were for causes different from CRC. Baseline characteristics in patients with and without steatosis are reported in Table 2, while follow-up and treatment characteristics are listed in Supplementary Table S2, Additional File, and baseline characteristics in patients with moderate/severe steatosis are reported in Supplementary Table S3, Additional File.

Table 1: Baseline liver characteristics including CT indices of steatosis and liver function tests.

Table 1	**No Liver Steatosis (n=193; 68.2%)**	**Mild liver steatosis (n=71; 25.1%)**	**Moderate/Severe liver steatosis (n=19; 6.7%)**
Liver density (HU); mean (SD)	58.35 (5.98)	49.38 (5.31)	34.84 (10.20)
Liver-to-spleen density ratio; median (range)	1.25 (1.09; 2.1)	1.09 (0.89;1.47)	0.78 (0.21; 1.03)
Liver-spleen density difference (HU); median (range)	12 (5; 33)	3.54 (-6; 15)	-10.53 (-38; 1)
AST (U/L); median (range) *	17 (10; 124)	25.57 (16; 48)	22.46 (7; 118)
AST > 40 U/L; n (%) *	3 (2.7)	4 (6.6)	1 (7.1)
ALT (U/L); median (range) *	14 (7; 94)	22.11 (7; 143)	30.29 (8; 76)
ALT > 49 U/L; n (%) *	2 (1.8)	2 (3.2)	2 (14.3)
GGT (U/L); median (range) *	17 (6; 280)	55.95 (7; 763)	49.21 (13; 171)
GGT > 73 U/L; n (%) *	4 (4.0)	9 (16.4)	3 (21.4)

Table 1: CT liver characteristics and liver function tests in patients with and without CT-defined liver steatosis. SD, standard deviation; HU, Hounsfield Units. * missing values were: in no liver steatosis group 83, 81, and 92 for AST, ALT, and GGT, respectively, in mild steatosis group 10, 9, and 16 for AST, ALT, and GGT, respectively, and in moderate/severe steatosis group 5 for all the three liver function tests.

Table 2: Baseline characteristics in patients with and without baseline CT-defined liver steatosis.

Table 2		**No Liver Steatosis (n= 193; 68.20%)**				**Liver Steatosis (n= 90; 31.80%)**			
		Overall	**Synchronous liver metastases (n=25)**	**Metachronous liver metastases (n=17)**	**Deaths (n=106)**	**Overall**	**Synchronous liver metastases (n=17)**	**Metachronous liver metastases (n=9)**	**Deaths (n=46)**
Age; mean (SD)		68.27 (14.3)				66.81 (13.23)			
Sex; n (%)	Male	106 (54.9)	15	9	62	61 (67.8)	11	4	30
	Female	87 (45.1)	10	8	44	29 (32.2)	6	5	16
Stage; n (%)	II	56 (29.0)	-	5	29	21 (23.3)	-	1	3
	III	86 (44.6)	-	8	32	39 (43.3)	-	7	17
	IV	45 (23.3)	25	4	41	27 (30.0)	17	1	23
	XM0	6 (3.1)	-	-	4	3 (3.3)	-	-	3
Grade; n (%)	Well	1 (0.5)	0	0	0	2 (2.2)	1	0	2
	Moderately	64 (33.2)	4	2	25	37 (41.1)	6	4	13
	Poorly	79 (40.9)	10	11	47	29 (32.2)	3	1	16
	Missing	49 (25.4)	11	4	34	22 (24.4)	7	4	15

Table 2: TNM stage refers to clinical staging. Stage XM0 was considered for patients with unknown local staging but exclusion of distant metastases with CT scan. SD: Standard deviation.

Association between liver steatosis and liver metastases

The prevalence of liver metastases at baseline was slightly higher in patients with liver steatosis (17/90, 19%) than in those without (25/168, 13%), while the incidence of liver metastases during follow-up was similar in patients with and without liver steatosis: 9 out of 73 (12%) and 17 out of 168 (10%), respectively (Figure 2). The final cumulative incidence of metastases was slightly higher in patients with steatosis.

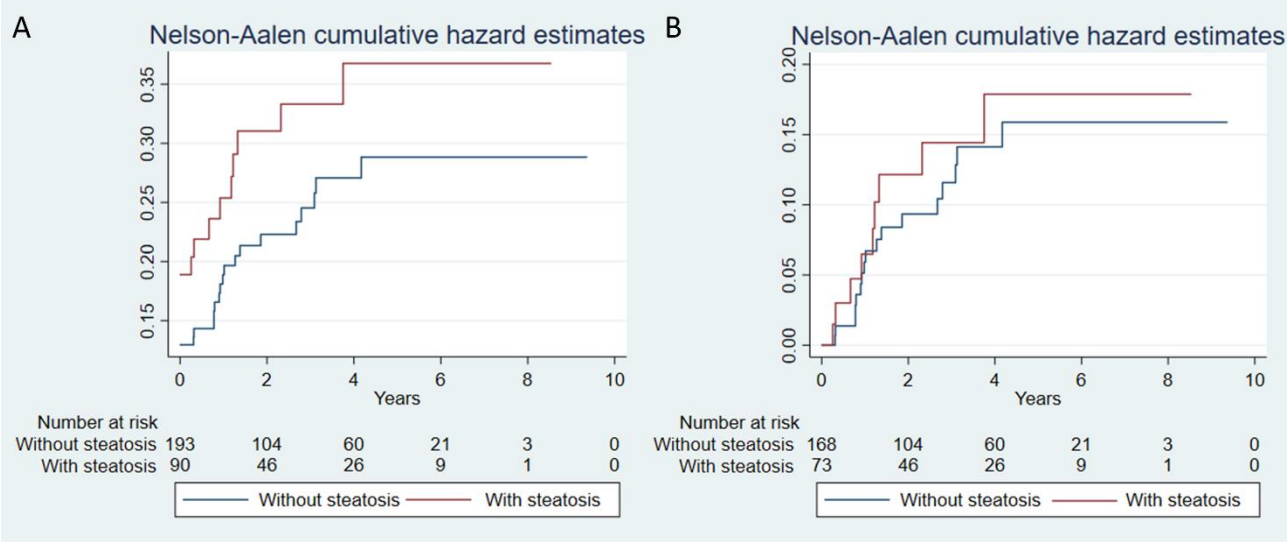


Figure 2: Cumulative incidence of liver metastases in all patients subdivided according to absence/presence of liver steatosis, including all metastases (A) and only metastases occurring during follow-up (B). Note that in graph A, cumulative hazard curves start from the observed values of synchronous metastases, i.e. 13.0% in patients without steatosis and 18.9% in patients with steatosis.

After correcting for possible confounders, liver steatosis was slightly associated with the presence of liver metastases at baseline, even if the difference in prevalence may have been due to chance; the association was almost null when considering the metachronous liver metastases occurring during follow-up (Table 3).

Table 3: Impact of liver steatosis on patient outcomes

Table 3	***Synchronous metastases*** **(n=283 patients; n=42 metastases)**			***Metachronous metastases*** **(n=241 patients; n=26 metastases)**			***Overall survival*** **(n=283 patients; n=152 deaths)**			***Overall survival excluding patients with baseline liver metastases*** **(n=241 patients; n=73 deaths)**		
Variables	***OR***	***95%CI***	***p***	***HR***	***95%CI***	***p***	***HR***	***95%CI***	***p***	***HR***	***95%CI***	***P***
Liver steatosis	1.58	0.79-3.14	0.20	1.25	0.55-2.85	0.60	0.92	0.64-1.32	0.65	0.66	0.35-1.22	0.19
Age at diagnosis	1.01	0.99-1.04	0.43	0.99	0.97-1.02	0.61	1.03	1.01-1.04	< 0.001	1.05	1.03-1.07	<0.001
Sex	0.90	0.45-1.77	0.75	1.66	0.76-3.66	0.21	1.03	0.74-1.44	0.89	0.76	0.43-1.36	0.36
Stage	-	-	-	1.78	1.00-3.20	0.05	2.15	1.76-2.62	< 0.001	2.51	1.64-3.82	<0.001
Year of diagnosis	1.01	0.85-1.19	0.91	0.88	0.73-1.08	0.22	0.98	0.90-1.06	0.64	0.82	0.71-0.94	0.006

Table 1: Logistic multivariate models for liver metastasis presence at baseline (model includes patients with liver metastases at baseline) and Cox proportional hazards regression models for occurrence of liver metastases during follow-up (after exclusion of patients with baseline liver metastases) and for overall survival (both including all patients and after exclusion of patients with baseline liver metastases). The variable of interest was steatosis at baseline; adjusting variables (age, sex, stage, and calendar period) included in the models were selected a priori for their known impact on disease. TNM stage refers to clinical staging.

Impact of liver steatosis on overall survival

No difference was found in terms of overall survival between patients with and without baseline liver steatosis (p = 0.46) (Figure 3). After correcting for age at diagnosis, sex, year of diagnosis and stage, liver steatosis did not affect overall survival (Table 3).

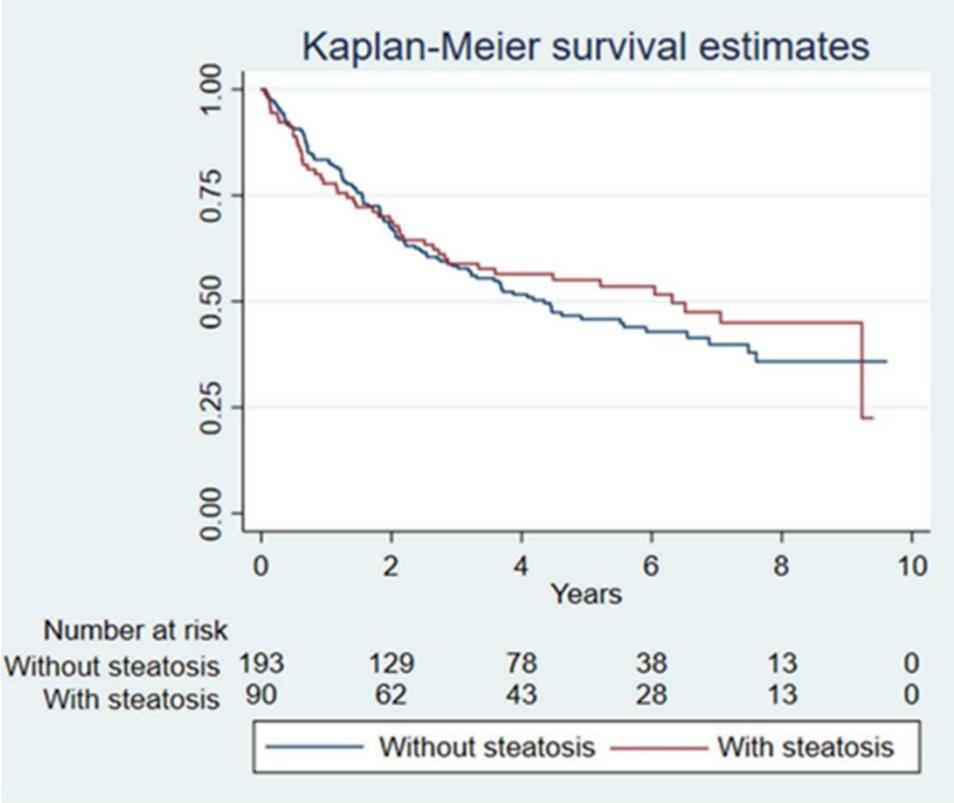


Figure 3: Overall survival in patients with and without baseline steatosis. All patients were included in this analysis.

Post-chemotherapy liver damage

Of the 202 patients with locally advanced rectal cancer, 142 had no baseline liver steatosis. Among these, patients who were not treated with neoadjuvant chemotherapy, who did not complete it, or who did not have a suitable assessment of damage (liver function tests or CT scan within 4 months from chemotherapy end date) were excluded. After also excluding patients who died (n=1) or had liver metastases (n=2) before the start of follow-up (i.e. 4 months after the end of neoadjuvant chemotherapy), 62 patients were included in the subgroup analysis on post-chemotherapy liver damage (Supplementary Tables S4, S5, and S6, Additional File).

Liver damage was found in 24 (38.7%) patients, 17 (70.8%) of whom had CT-defined liver steatosis appearance or liver volumetric change and 11 (45.8%) of whom had significant increase in liver function tests.

In this small group of patients, those experiencing post-chemotherapy liver damage had a higher occurrence of liver metastases during follow-up, even if based only on 5 vs 2 events in the group without vs with liver damage, respectively, and worse overall survival, with 10 vs 5 events, respectively (Figure 3). After adjusting for potential confounders, exact logistic regressions showed that the ORs of post-chemotherapy liver damage were 6.0 (95% CI 0.82 to 74.0) for liver metastases, and 5.9 (95% CI 1.37-30.9) for all-cause mortality. Point estimates were similar when using Cox proportional hazard models (HR 6.2 for liver metastases, HR 4.4 for OS).

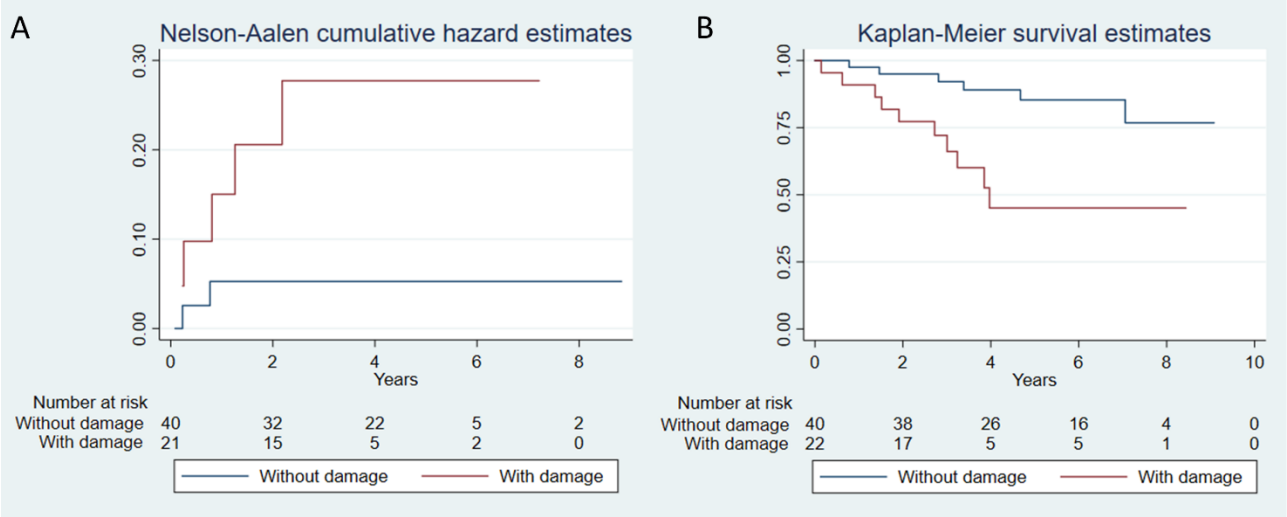


Figure 4: Cumulative incidence of liver metastases (A) and overall survival (B) in patients with and without post-neoadjuvant chemotherapy liver damage.

DISCUSSION

In this cohort of patients with stage II-IV rectal cancer, the prevalence of CT-defined liver steatosis was 31.8%, similar or slightly higher than that reported in the general population [1]. The prevalence of steatosis was slightly higher in stage IV.

The prevalence of liver metastases at diagnosis was slightly higher in patients with steatosis (OR 1.6, 95% CI 0.8-3.0). Although the difference could be due to chance, it is interesting that the higher prevalence of steatosis in stage IV is completely justified by the excess of liver metastases present in patients with steatosis.

Our data show that steatosis is not associated with a higher probability of developing a metastasis during follow-up (HR 1.25, 95% CI 0.55-2.85). Nevertheless, cumulative incidence was higher in patients with steatosis due to the higher prevalence at diagnosis. This observation has some implications that should be considered when comparing our study with others, because some of the

metastases detected at baseline in our study may have been detected during follow-up in other studies and vice versa, depending on the intensity of diagnostic imaging procedures adopted for staging. The consequence is that studies including only patients without metastases at diagnosis may show only one part of the phenomenon.

Existing studies on the effect of liver steatosis on the risk of developing liver metastases in CRC patients present ambiguous results. In fact, the results from studies on patients who underwent primary CRC resection [18,25-26] suggest a protective role of steatosis in terms of both synchronous and metachronous liver metastasis. On the other hand, studies on patients in follow-up after resection of liver metastases [17,27] show an equal or higher risk of developing a new metastasis in those with steatosis. The main differences between our study and the existing literature are the inclusion of all stage II-IV rectal cancers and the focus exclusively on rectal cancer, which is a biologically distinct entity compared to colon cancer, with different characteristics also in terms of the probability of liver metastases [10,13] Moreover, while some other studies used CT for the assessment of liver steatosis [17], others adopted pathological examination on liver biopsy specimens [25,28]. Although the use of CT for the assessment of liver steatosis has been validated [29], the comparability between studies which adopted different techniques may be affected.

Based on our results, it seems that both the biases that we hypothesized when designing this study were not confirmed. A masking effect of liver steatosis on the detection of metastases would have resulted in a lower prevalence of metastases at baseline, followed by a liver metastasis incidence excess in patients with steatosis. On the other hand, the increased detection of metastases due to more and different imaging tests in patients with steatosis would have induced an anticipation in liver metastasis diagnosis, resulting in a higher prevalence of baseline metastasis followed by a reduced occurrence during follow-up in patients with steatosis. Our results do not reflect any of these conditions. As our curves of metastasis occurrence after diagnosis were perfectly parallel in the two groups, if we had had undetected metastases, these would be similar in the two groups. Consequently, either no masking effect or detection bias was present or the two compensated each other perfectly.

We observed lower overall survival and a higher risk of liver metastases in patients with post-chemotherapy liver damage. Due to the small sample size, our results are very imprecise and could be due to chance. A previous study reported an association between chemotherapy-induced liver damage and improved survival after CRC liver metastasis resection [30]. The discrepancy between these two results, despite the different contexts and study designs, should stimulate the conception of larger studies on this topic. Further studies should be designed to confirm such association with larger numbers, but also to understand what the direction of causality is. In fact, as most of the liver

metastases in patients with liver damage in our study were diagnosed a few months after the end of chemotherapy, we cannot exclude that liver damage may be a sign of existing micrometastases [31] and not a risk factor for developing a metastasis. In other words, the liver changes that we classified as chemotherapy-induced liver damage may represent generic liver characteristics (increase in liver function tests and liver volume and decreased liver density) linked to liver metastases, as a sort of early biomarker of undetectable metastasization. This reverse causality interpretation could also explain the discrepancies with previous studies on patients surviving after liver metastasis resection.

This study has several limitations. Firstly, it was conducted retrospectively. Nevertheless, starting from a population-based registry allowed us to understand whether and how some of the biases of a retrospective cohort study could occur. The study flowchart shows that very few cases were excluded due to the lack of adequate imaging. We cannot be certain that procedures for assessing metastases were conducted in the same way in all patients, and differences could be linked to liver conditions. Secondly, our sample size was small, particularly for the secondary aim of assessing liver damage as a risk factor. These were pre-planned analyses and therefore reported results are not driven by data, meaning that the reported p-values can be interpreted as true test of hypothesis; we therefore decided not to define a threshold to reject the hypothesis, but only to present a confidence interval to reflect on the possible implication of this association. Data on factors associated with liver steatosis, such as BMI, presence of metabolic syndrome, viral infections, or alcohol intake, were available only for few patients. However, by adjusting for these factors, which may lie on the same causal chain linking steatosis with liver metastases, there would be a high chance of hiding the effect of the more distal risk factors in favour of the proximal ones. Finally, steatosis was assessed by CT scan, which is known to be highly specific but presents low sensitivity, especially for mild steatosis [23], and not by MR or histological examination, which are more precise. However, CT scan was the only test available for the vast majority of rectal cancer patients at baseline.

CONCLUSION

Our data suggest that rectal cancer patients with and without steatosis have a similar occurrence of metastases during follow-up, despite our observing a slightly higher burden of liver metastases at diagnosis in patients with steatosis, compatible with chance. In this preliminary analysis, liver damage after chemotherapy was associated with a higher occurrence of liver metastases; this association and its direction should be further assessed by larger prospective studies.

REFERENCES

1 Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018 Jan;67(1):123-133.

2 Bitzur R, Brenner R, Maor E, Antebi M, Ziv-Baran T, Segev S, Sidi Y, Kivity S. Metabolic syndrome, obesity, and the risk of cancer development. Eur J Intern Med. 2016 Oct;34:89-93.

3 Kant P, Hull MA. Excess body weight and obesity--the link with gastrointestinal and hepatobiliary cancer. Nat Rev Gastroenterol Hepatol. 2011 Apr;8(4):224-38.

4 Chun YS, Laurent A, Maru D, Vauthey JN. Management of chemotherapy-associated hepatotoxicity in colorectal liver metastases. Lancet Oncol. 2009 Mar;10(3):278-86.

5 Vigano L, De Rosa G, Toso C, Andres A, Ferrero A, Roth A, Sperti E, Majno P, Rubbia-Brandt L. Reversibility of chemotherapy-related liver injury. J Hepatol. 2017 Jul;67(1):84-91.

6 Vauthey JN, Pawlik TM, Ribero D, Wu TT, Zorzi D, Hoff PM, Xiong HQ, Eng C, Lauwers GY, Mino-Kenudson M, Risio M, Muratore A, Capussotti L, Curley SA, Abdalla EK. Chemotherapy regimen predicts steatohepatitis and an increase in 90-day mortality after surgery for hepatic colorectal metastases. J Clin Oncol. 2006 May 1;24(13):2065-72.

7 Robinson SM, Wilson CH, Burt AD, Manas DM, White SA. Chemotherapy-associated liver injury in patients with colorectal liver metastases: a systematic review and meta-analysis. Ann Surg Oncol. 2012 Dec;19(13):4287-99.

8 Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015 Mar 1;136(5):E359-86.

9 Vicentini M, Zorzi M, Bovo E, Mancuso P, Zappa M, Manneschi G, Mangone L, Giorgi Rossi P; Colorectal Cancer Screening IMPATTO study working group. Impact of screening programme using the faecal immunochemical test on stage of colorectal cancer: Results from the IMPATTO study. Int J Cancer. 2019 Jul 1;145(1):110-121.

10 Qiu M, Hu J, Yang D, Cosgrove DP, Xu R. Pattern of distant metastases in colorectal cancer: a SEER based study. Oncotarget. 2015 Nov 17;6(36):38658-66.

11 Rödel C, Sauer R. Neoadjuvant radiotherapy and radiochemotherapy for rectal cancer. Recent Results Cancer Res. 2005;165:221-30.

12 Tjandra JJ, Chan MK. Follow-up after curative resection of colorectal cancer: a meta-analysis. Dis Colon Rectum. 2007 Nov;50(11):1783-99.

13 Riihimäki M, Hemminki A, Sundquist J, Hemminki K. Patterns of metastasis in colon and rectal cancer. Sci Rep. 2016 Jul 15;6:29765.

14 Wu Z, Wei D, Gao W, Xu Y, Hu Z, Ma Z, Gao C, Zhu X, Li Q. TPO-Induced Metabolic Reprogramming Drives Liver Metastasis of Colorectal Cancer CD110+ Tumor-Initiating Cells. Cell Stem Cell. 2015 Jul 2;17(1):47-59.

15 Mikuriya Y, Tashiro H, Kuroda S, Nambu J, Kobayashi T, Amano H, Tanaka Y, Ohdan H. Fatty liver creates a pro-metastatic microenvironment for hepatocellular carcinoma through activation of hepatic stellate cells. Int J Cancer. 2015 Feb 15;136(4):E3-13.

16 Eveno C, Hainaud P, Rampanou A, Bonnin P, Bakhouche S, Dupuy E, Contreres JO, Pocard M. Proof of prometastatic niche induction by hepatic stellate cells. J Surg Res. 2015 Apr;194(2):496-504.

17 Hamady ZZ, Rees M, Welsh FK, Toogood GJ, Prasad KR, John TK, Lodge JP. Fatty liver disease as a predictor of local recurrence following resection of colorectal liver metastases. Br J Surg. 2013 May;100(6):820-6.

18 Murono K, Kitayama J, Tsuno NH, Nozawa H, Kawai K, Sunami E, Akahane M, Watanabe T. Hepatic steatosis is associated with lower incidence of liver metastasis from colorectal cancer. Int J Colorectal Dis. 2013 Aug;28(8):1065-72.

19 Wu W, Chen J, Ye W, Li X, Zhang J. Fatty liver decreases the risk of liver metastasis in patients with breast cancer: a two-center cohort study. Breast Cancer Res Treat. 2017 Nov;166(1):289-297.

20 Wu W, Liao H, Ye W, Li X, Zhang J, Bu J. Fatty liver is a risk factor for liver metastasis in Chinese patients with non-small cell lung cancer. PeerJ. 2019 Mar 14;7:e6612.

21 Nakai H, Arizono S, Isoda H, Togashi K. Imaging Characteristics of Liver Metastases Overlooked at Contrast-Enhanced CT. AJR Am J Roentgenol. 2019 Apr;212(4):782-787.

22 Lee SW, Park SH, Kim KW, Choi EK, Shin YM, Kim PN, Lee KH, Yu ES, Hwang S, Lee SG. Unenhanced CT for assessment of macrovesicular hepatic steatosis in living liver donors: comparison of visual grading with liver attenuation index. Radiology. 2007 Aug;244(2):479-85.

23 Esterson YB, Grimaldi GM. Radiologic Imaging in Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. Clin Liver Dis. 2018 Feb;22(1):93-108.

24 Hirji KF, Mehta CR, Patel NR. Computing distributions for exact logistic regression. J Am Stat Assoc. 1986 82: 1110-1117.

25 Iascone C, Ruperto M, Barillari P. Metastasi epatiche sincrone da carcinoma colo-rettale in pazienti con cirrosi epatica o steatosi [Occurrence of synchronous colorectal cancer metastasis in the cirrhotic or fatty liver]. Minerva Chir. 2005 Jun;60(3):185-90. Italian.

26 Hayashi S, Masuda H, Shigematsu M. Liver metastasis rare in colorectal cancer patients with fatty liver. Hepatogastroenterology. 1997 Jul-Aug;44(16):1069-75.

27 Ramos E, Torras J, Lladó L, Rafecas A, Serrano T, Lopez-Gordo S, Busquets J, Fabregat J. The influence of steatosis on the short- and long-term results of resection of liver metastases from colorectal carcinoma. HPB (Oxford). 2016 Apr;18(4):389-96.

28 Au KP, Chok KSH, Chan ACY, Dai WC, Cheung TT, Lo CM. Impact of Hepatitis B Carrier Status on the Outcomes of Surgical Treatment of Colorectal Liver Metastases. World J Surg. 2018 Aug;42(8):2642-2650.

29 Byun J, Lee SS, Sung YS, Shin Y, Yun J, Kim HS, Yu ES, Lee SG, Lee MG. CT indices for the diagnosis of hepatic steatosis using non-enhanced CT images: development and validation of diagnostic cut-off values in a large cohort with pathological reference standard. Eur Radiol. 2019 Aug;29(8):4427-4435.

30 Viganò L, Capussotti L, De Rosa G, De Saussure WO, Mentha G, Rubbia-Brandt L. Liver resection for colorectal metastases after chemotherapy: impact of chemotherapy-related liver injuries, pathological tumor response, and micrometastases on long-term survival. Ann Surg. 2013 Nov;258(5):731-40; discussion 741-2.

31 Yokoyama N, Shirai Y, Ajioka Y, Nagakura S, Suda T, Hatakeyama K. Immunohistochemically detected hepatic micrometastases predict a high risk of intrahepatic recurrence after resection of colorectal carcinoma liver metastases. Cancer. 2002 Mar 15;94(6):1642-7.

6. SARCOPENIA AND FAT COMPARTMENTS DISTRIBUTION AS PROGNOSTIC FACTORS

6.1 Prognostic impact of muscle quantity and quality and fat distribution in diffuse large B-cell lymphoma patients: an observational study*

Summary: Since body composition is associated with long-term health outcomes in many oncologic diseases, particularly in elderly patients, we conducted a retrospective study to evaluate the association of CT biomarkers of muscle quantity and quality and fat distribution with disease outcomes in patients with diffuse large B-cell lymphoma.

Baseline CT scans of 116 patients (48% female, median 64 years) with diffuse large B-cell lymphoma (DLBCL) were retrospectively reviewed to investigate the prognostic role of sarcopenia and fat compartment distributions on overall survival (OS), progression-free survival (PFS), and early therapy termination. Skeletal muscle index (SMI), skeletal muscle density (SMD), and intermuscular adipose tissue (IMAT) were quantified at the level of the third lumbar vertebra (L3) and proximal thigh (PT). Low L3-SMD, but not low L3-SMI, was associated with early therapy termination (p=0.028), shorter OS (HR=6.29; 95% CI=2.17-18.26; p<0.001), and shorter PFS (HR=2.42; 95% CI=1.26-4.65; p=0.008). After correction for sex, International Prognostic Index (IPI), BMI, and R-CHOP therapy (rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone), low L3-SMD remained associated with poor OS (HR=3.54; 95% CI=1.10-11.40; p=0.034) but not with PFS. Increased PT-IMAT was prognostic for poor OS and PFS after correction for sex, IPI, BMI, and R-CHOP therapy (HR=1.35; CI=1.03-1.7; p=0.03, and HR=1.30; CI=1.04-1.64; p=0.024, respectively). Reduced muscle quality (SMD) and increased intermuscular fat (IMAT), rather than low muscle quantity (SMI), are associated with poor prognosis in DLBCL, when measured at the L3 level, and particularly at the level of the proximal thigh. The proximal thigh represents a novel radiological landmark to study body composition.

*Manuscript accepted in *Frontiers in Nutrition*. Besutti G, Massaro F, Bonelli E, Braglia L, Casali M, Versari A, Ligabue G, Pattacini P, Cavuto S, Merlo FD, Luminari S, Merli F, Vaccaro S, Pellegrini M. Prognostic Impact of Muscle Quantity and Quality and Fat Distribution in Diffuse Large B-Cell Lymphoma Patients.

Supplementary Material in Appendix D.

1 Introduction

Non-Hodgkin lymphoma is the fifth most common neoplasia in Western countries [1]. Among the various histological subtypes, DLBCL accounts for 25-35% of all cases [2]. With the advent of modern chemoimmunotherapy regimens, more than 50% of DLBCL patients achieve complete remission [3]. Nevertheless, many patients still fail to achieve an optimal response or experience relapse, and the most used prognostic tools to identify these subjects comprise only few clinical features (age, stage, performance status, serum lactate dehydrogenase level, number of involved extranodal sites) that were validated more than 20 years ago [4].

Body composition parameters are associated with long-term health outcomes in many diseases and have been recently studied as prognostic factors in DLBCL patients. Researchers have reported a negative impact of sarcopenia on survival outcomes in DLBCL patients treated with the R-CHOP regimen [5-7], particularly among elderly patients [5].

However, a conclusive definition of sarcopenia has not yet been established, and its assessment often remains beyond clinical practice. A CT scan is one of the most widely accepted tools for the assessment of sarcopenia, defined as severe systemic loss of skeletal muscle mass leading to progressive functional impairment. CT or PET-CT scan is routinely performed in DLBCL patients during initial staging and subsequent follow-up, representing an accessible source of data for the identification of sarcopenia or abnormal fat accumulation.

Sarcopenia, defined as a decreased skeletal muscle index at the level of the third lumbar vertebra, yielded ambiguous results as a prognostic marker. Some authors found that a decreased muscle mass was independently associated with lower overall survival and progression-free survival in DLBCL patients [5, 7]. Others confirmed this result in male patients only [8, 9], while Chu et al. [10] described a trend toward an improved survival in patients with lower SMI.

Besides muscle depletion, sarcopenia is also characterized by an increased proportion of inter- and intramuscular fat, which are markers of muscle quality deterioration [11, 12]. Skeletal muscle density, evaluated by the mean attenuation on CT imaging, represents the accumulation of intramuscular fat and water, while intermuscular adipose tissue is the quantity of visible fat that can be measured beneath the fascia and within the muscles. Both SMD and IMAT are predictors of muscle function [13, 14]. Furthermore, these ectopic fat depots are associated with chronic inflammation, metabolic impairment, and insulin resistance [15]. As described for non-hematologic malignancies [16, 17], a reduced SMD is a better predictor of poor survival than a reduced SMI in DLBCL patients [6, 10].

During cancer trajectory, inflammation, metabolic derangements, poor nutrition, physical inactivity, and cancer treatment may result in simultaneous skeletal muscle loss and fat gain that culminate in sarcopenic obesity, which is strongly related to reduced survival in cancer patients [18]. Several studies have addressed the association between body mass index (BMI) and DLBCL prognosis with discordant results [19-22], while a higher visceral adiposity measured by means of a CT scan resulted in a poorer prognosis [23]. Recently, in lymphoma patients undergoing stem cell transplantation, the combined effect of visceral adiposity and sarcopenia defined as low SMI has been correlated with higher mortality. However, the potential combined prognostic effect of low muscle quality and visceral adiposity has never been investigated in DLBCL patients.

The aim of this study was to explore the impact of baseline PET-CT indices of sarcopenia and adipose tissue distribution on OS and PFS in patients with DLBCL. Secondarily, we aimed to evaluate the impact of these indices on early therapy termination.

We have explored body composition metrics in PET-CT cross-sectional images at the L3 level and at the proximal thigh (PT) level. While the L3 cross-sectional skeletal muscle area (SMA) is a standard landmark linearly related to whole-body muscle mass and total adipose tissue [24], the skeletal muscle cross-sectional area at the PT level might be a new and sensitive prognostic marker of sarcopenia.

2 Materials and Methods

2.1. Patients and study design

All consecutive patients diagnosed with DLBCL between January 2014 and December 2017 at our institution were considered eligible for inclusion in this retrospective study. All patients were diagnosed through lymph node excision biopsy or core biopsy of the presenting extranodal sites and underwent a unilateral bone marrow biopsy. DLBCL patients routinely undergo whole-body PET-CT and CT imaging at the time of diagnosis for disease staging. The unavailability of baseline PET-CT scan for retrospective review was considered as an exclusion criterion. Patients informed consent to participate in the study was obtained whenever possible, due to the retrospective nature of the study. The study was approved by the appropriate local ethics committee of Area Vasta Emilia Nord (protocol number: 2018/0111083) and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

2.2. Clinical chart

Patient records were retrospectively reviewed to collect demographic data (age, sex), anthropometric data at the time of diagnosis (height, weight, and BMI, defining obesity as a BMI≥30), and baseline known clinical prognostic factors (Ann Arbor disease staging, serum lactate dehydrogenase level, Eastern Cooperative Oncology Group (ECOG) performance status, extranodal sites). IPI was calculated for all patients. Data on chemoimmunotherapy regimen, drug-induced toxicity, and response to therapy were collected. After evaluation of disease features, age, and comorbidities, the patients received chemoimmunotherapy treatment with different regimens according to international guidelines and local practice.

2.3. PET-CT image analysis

PET data were acquired using a Discovery GE PET/CT scanner (General Electric Medical Systems, Cleveland, OH, USA), which combines a helical 16-slice CT and a 3-dimensional (3-D) PET scanner. Body composition parameters were measured on the non-enhanced CT helical scan (scan field 500 mm, increment 3.75 mm, slice thickness 3.75 mm, pitch 1. 0.8 seconds per rotation, matrix 512×512 pixels, 120 kV, 80 mA). Images were retrospectively analyzed by a single trained image analyzer (B.E.) supervised by a senior radiologist (P.P.), both blinded to clinical data and outcomes, using the OSIRIX-Lite software V5.0 (Pixmeo, Sarl, Switzerland).

A single slice at the L3 level with both transverse processes visible was selected. Skeletal muscle and abdominal fat compartments were selected after applying Hounsfield Unit (HU) thresholds of -29 to +150 and -190 to -30, respectively. The cross-sectional SMA and the total adipose tissue, subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), and IMAT at the L3 level were obtained through autosegmentation and manual contour correction when necessary. The muscles segmented included the rectus abdominis, abdominal wall, psoas, and paraspinal muscle groups. The SMI was calculated by dividing the cross-sectional SMA by squared height in meters. Decreased muscle quantity was defined according to previously used L3-SMI cut-off values (<43 cm^2/m^2 for men with BMI <25, <53 cm^2/m^2 for men with BMI ≥25, and <41 cm^2/m^2 for women) [25]. Other previously used cut-off values were also investigated: L3-SMI-2 (<55.8 cm^2/m^2 for men and <38.9 cm^2/m^2 for women) [5] and L3-SMI-3 (<52.4 cm^2/m^2 for men and <38.5 cm^2/m^2 for women) [26]. Unless otherwise specified, the analysis of L3-SMI using the cut-off value of Martin et al. [25] is reported. The analysis using other cut-off parameters (L3-SMI-2 and L3-SMI-3) is reported in the Electronic Supporting Information. The Mean SMD (L3-SMD) was registered in the same region of interest used for SMA measurement. Poor skeletal muscle quality was defined according to previously used L3-SMD cut-off values (<41 HU for patients with BMI <25 and <33

HU for patients with BMI >25) [10, 25, 26]. The ratio between visceral and subcutaneous adipose tissue areas (VAT/SAT) was also reported. Using the same software and analytical method, IMAT, SMA, SMI, and SMD were also obtained at the PT level. The CT slice selected for this purpose was immediately below the last slice including the gluteal muscle. Representative images are reported in Figure 1.

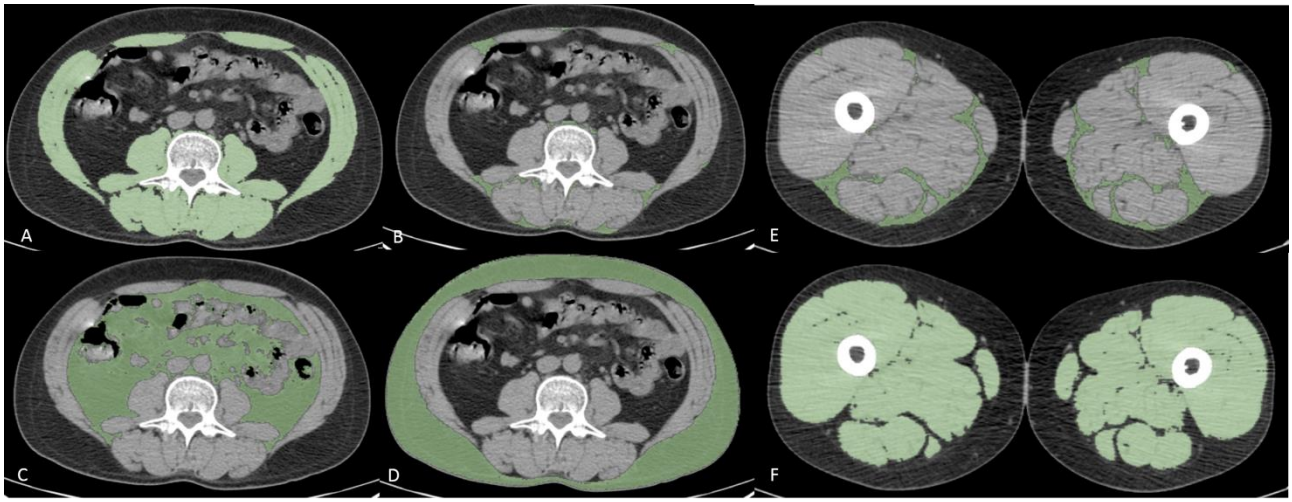


Figure 1. *Examples of PET-CT body composition measurements. Muscle and adipose tissue segmentations were obtained after applying specific density thresholds. In these axial slices at the level of the third lumbar vertebra (L3) (**A-D**) and at the level of proximal thigh (PT) (**E-F**), the colored regions of interest represent L3- and PT-skeletal muscle area (SMA) (**A, F**), L3- and PT-intermuscular adipose tissue (IMAT) (**B, E**), visceral adipose tissue (VAT) (**C**), and subcutaneous adipose tissue (SAT) (**D**)*

2.4. Statistical analyses

In the absence of a priori hypothesis, given the exploratory nature of the study, no formal sample size calculation was performed. Clinical and demographic data were expressed in terms of frequency and percentage for categorical variables, mean ± standard deviation for approximately normally distributed quantitative variables, and median (and interquartile range (IQR)) otherwise. Proportion estimates were accompanied by Wilson [27] confidence intervals (CIs). Proportions were compared between independent groups using the chi-square test or Fisher's exact test if needed. Quantitative variables were compared across independent groups using the Mann-Whitney U test; correlation was assessed with the Pearson correlation coefficient (after visual inspection of scatterplots).

In terms of prognostic analyses, PFS was defined as the time from the diagnosis to first progression or death from any cause, whichever came first. OS was defined as the time from the diagnosis to death from any cause. The median follow-up time was evaluated using the reverse Kaplan-Meier method [28]. Prognostic analyses entailed both estimates of survival function (using the Kaplan-Meier method, compared using the log rank test) and estimates of effects of covariates via Cox's regression models (proportional hazard assumption was assessed by testing scaled Schoenfeld residuals' correlation with time; no violation of the assumption was found).

Main prognostic analyses included PET-CT indices, both continuous and dichotomized, according to the criteria proposed in the literature; in multivariate analyses, we used IPI, sex, BMI, and therapy as covariates. Unless otherwise specified, CIs for estimates were two-tailed and calculated considering a 0.95 confidence level. Performed tests were considered statistically significant for p-values <0.05. Statistical analysis was performed using R 3.5.2 (2020) (R Core Team, Vienna, Austria).

3 Results

3.1. Population

Of 120 patients diagnosed with DLBCL between January 2014 and December 2017 at our institution, 4 were excluded because of unavailability of PET-CT examination performed at baseline for staging purposes, leaving a total of 116 patients included in the study. The mean age at diagnosis was 63.7 (± 16.4) years and 56 patients (48.3%) were female. Demographic, anthropometric, and clinical characteristics at the time of diagnosis and data on therapy and response to therapy are reported in Table 1.

TABLE 1. Demographic, anthropometric, and clinical data of the patients at the time of diagnosis and chemoimmunotherapy regimen

Covariates		**All patients (n=116)**
Age at diagnosis (years)		63.7 ± 16.4
Female		56 (48.3%)
Weight (kg)		70.6 ± 15.1
Height (cm)		166.0 ± 10
BMI (kg/m^2)		25.4 ± 4.4
Stage	I	14 (12.1%)
	II	24 (20.7%)
	III	17 (14.6%)
	IV	61 (52.6%)
ECOG Performance Status	0	80 (69.0%)
	1	12 (10.3%)
	2	24 (20.7%)
Extranodal sites	0	37 (32.2%)
	1	45 (39.1%)
	2	25 (21.7%)
	3	6 (5.2%)
	4	2 (1.7%)
LDH (U/l)		468 (338-656)
IPI	0	8 (6.9%)
	1	24 (20.7%)
	2	24(20.7%)
	3	30 (25.9%)
	4	23 (19.8%)
	5	7 (6%)
Chemo-immunotherapy regimens	R-CHOP	70 (60.3%)
	Others[a]	46 (39.7%)
Early therapy termination		16 (13.8%)
Response to therapy	Not evaluated	9 (7.8%)
	Progressive Disease	6 (5.2%)
	Complete Remission	86 (74.1%)

	Partial Remission	15 (12.9%)

Data are reported as frequencies and percentage for categorical variables, means and standard deviations for normally distributed continuous variables, and median and interquartile range for non-normally distributed continuous variables. BMI, body mass index; ECOG, Eastern Cooperative Oncology Group; IPI, international prognostic index.

[a]Includes chemoimmunotherapy regimens R-CVP (n=5), R-miniCHOP (n=18), R-MACOP-B (n=9), dose-adjusted EPOCH-R (n=3), R-CODOX-M/R-IVAC (n=3), autologous stem cell transplantation with FEAM conditioning (n=8) [29-31]. R-CHOP, rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone; R-CVP, rituximab, cyclophosphamide, vincristine, prednisone; R-miniCHOP, rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone; R-MACOP-B, rituximab, methotrexate, doxorubicin, cyclophosphamide, vincristine, prednisone, bleomycin; Dose-adjusted EPOCH-R, etoposide, vincristine, doxorubicin, cyclophosphamide, prednisone, rituximab; R-CODOX-M/R-IVAC, rituximab, cyclophosphamide, vincristine, doxorubicin, methotrexate / rituximab, ifosfamide, etoposide, cytarabine; FEAM, fotemustine, etoposide, cytarabine, melphalan.

3.2. PET-CT indices

Body composition PET-CT indices at the time of diagnosis are reported in Supplementary Table 1. The prevalence of L3-SMI-defined decreased muscle mass was 25%, while the prevalence of poor muscle quality based on L3-SMD was 52.6%. In terms of the associations between baseline PET-CT indices and serologic data such as glycemia, albumin, total serum protein, C-reactive protein, and vitamin D, we found correlations between L3-SMD and albumin (r=0.38, CI: 0.22-0.53), L3-SMD and total serum protein level (r=0.41, CI: 0.24-0.55), and VAT and glycemia (r=0.35, CI: 0.18-0.50). Upon examining the associations between different PET-CT indices, we identified a weak correlation between L3-SMD and L3-SMI (r=0.36, CI: 0.19-0.51). Muscle quantity and quality indices (SMD, SMI, and IMAT) measured at the L3 level and PT level showed strong correlations, with r ranging between 0.63and 0.86, the lowest for IMAT and the highest for SMD, respectively. These results are reported in Supplementary Table 2.

3.3. Outcome measures

The median follow-up was 30 months (IQR, 24-34 months). During the follow-up, we observed 43 progressions, of which 28 were deaths. The median OS and PFS were 64.8 and 53.6 months,

respectively. Survival curves for the whole study population are reported in Supplementary Figure 1.

3.4. Impact of body composition PET-CT indices on survival

Considering PET-CT indices as continuous variables (Table 2), an increased SMD was found to be a protective factor both for OS and PFS, which was stronger at the PT level, with significant or borderline significant p-values in the multivariate analysis after correction for sex, BMI, IPI, and therapy (HR=0.46; CI=0.23-0.90; p=0.025, and HR=0.62; CI=0.37-1.04; p=0.068, respectively). The SMI showed discordant effects; however, an increased L3-SMI was a significant independent poor prognostic factor for OS in the multivariate analysis (HR=2.06; CI=1.11-3.83; p=0.023). An increased IMAT had a significant independent poor prognostic role both in OS and PFS when measured at the PT level (HR=1.35; CI=1.03-1.7; p=0.03, and HR=1.30; CI=1.04-1.64; p=0.024, respectively), but not at the L3 level. Other indices including VAT and VAT/SAT did not show any significant association with neither OS nor PFS.

TABLE 2. Univariate and multivariate associations of different PET-CT indices with OS and PFS

Overall survival						
	Univariate analysis			**Multivariate analysis**		
	HR	95%CI	p	HR	95%CI	p
L3-SMI	1.24	0.86-1.79	0.240	2.06	1.11-3.83	0.023
L3-SMD	0.46	0.29-0.74	0.001	0.59	0.33-1.05	0.075
L3-IMAT	1.22	0.89-1.68	0.211	0.94	0.65-1.36	0.742
PT-SMI	0.85	0.72-1.01	0.069	1.02	0.82-1.27	0.872
PT-SMD	0.28	0.16-0.50	<0.001	0.46	0.23-0.90	0.025
PT-IMAT	1.41	1.12-1.78	0.003	1.35	1.03-1.76	0.030
VAT	1.01	0.97-1.04	0.795	1.01	0.95-1.07	0.799
VAT/SAT	1.14	0.71-1.83	0.596	0.87	0.47-1.62	0.658
Progression free survival						
	Univariate analysis			**Multivariate analysis**		
	HR	95%CI	p	HR	95%CI	p
L3-SMI	0.96	0.69-1.33	0.797	1.16	0.71-1.90	0.561
L3-SMD	0.64	0.44-0.91	0.014	0.83	0.55-1.26	0.387
L3 IMAT	1.09	0.85-1.41	0.487	0.89	0.67-1.18	0.405
PT-SMI	0.86	0.74-0.98	0.028	1.00	0.83-1.19	0.982
PT-SMD	0.41	0.26-0.63	<0.001	0.62	0.37-1.04	0.068
PT-IMAT	1.33	1.09-1.62	0.004	1.30	1.04-1.64	0.024
VAT	1.01	0.98-1.04	0.584	1.03	0.98-1.08	0.197
VAT/SAT	1.26	0.89-1.80	0.192	1.12	0.73-1.72	0.595

The multivariate models include IPI, sex, BMI, and therapy (R-CHOP vs other regimens) as covariates. PET-CT, positron emission tomography-computed tomography; OS, overall survival; PFS, progression-free survival; IPI, international prognostic index; BMI, body mass index; HR, hazard ratio; CI, confidence interval; L3, third lumbar vertebra; PT, proximal thigh; SMI, skeletal muscle index; SMD, skeletal muscle density; IMAT, intramuscular adipose tissue; VAT, visceral adipose tissue; SAT, subcutaneous adipose tissue. HRs for SMI, SMD, IMAT, and VAT are reported considering a 10-unit increase; R-CHOP, rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone; R-CVP, rituximab, cyclophosphamide, vincristine, prednisone; for other regimens see Table 1

3.5. Effect of poor muscle quality and muscle mass depletion on survival

Poor muscle quality defined according to L3-SMD was significantly associated with shorter OS and PFS (HR=6.29; 95% CI=2.17-18.26; p<0.001, and HR=2.42; 95% CI=1.26-4.65; p=0.008, respectively). When defined according to L3-SMI, decreased muscle mass was not associated with OS and PFS (HR=0.70; 95% CI=0.26-1.85; p=0.469, and HR=1.47; 95% CI=0.77-2.82; p=0.248) (Figure 2). No association with OS and PFS was observed for L3-SMI-2 and L3-SMI-3 (Supplementary Figure 2).

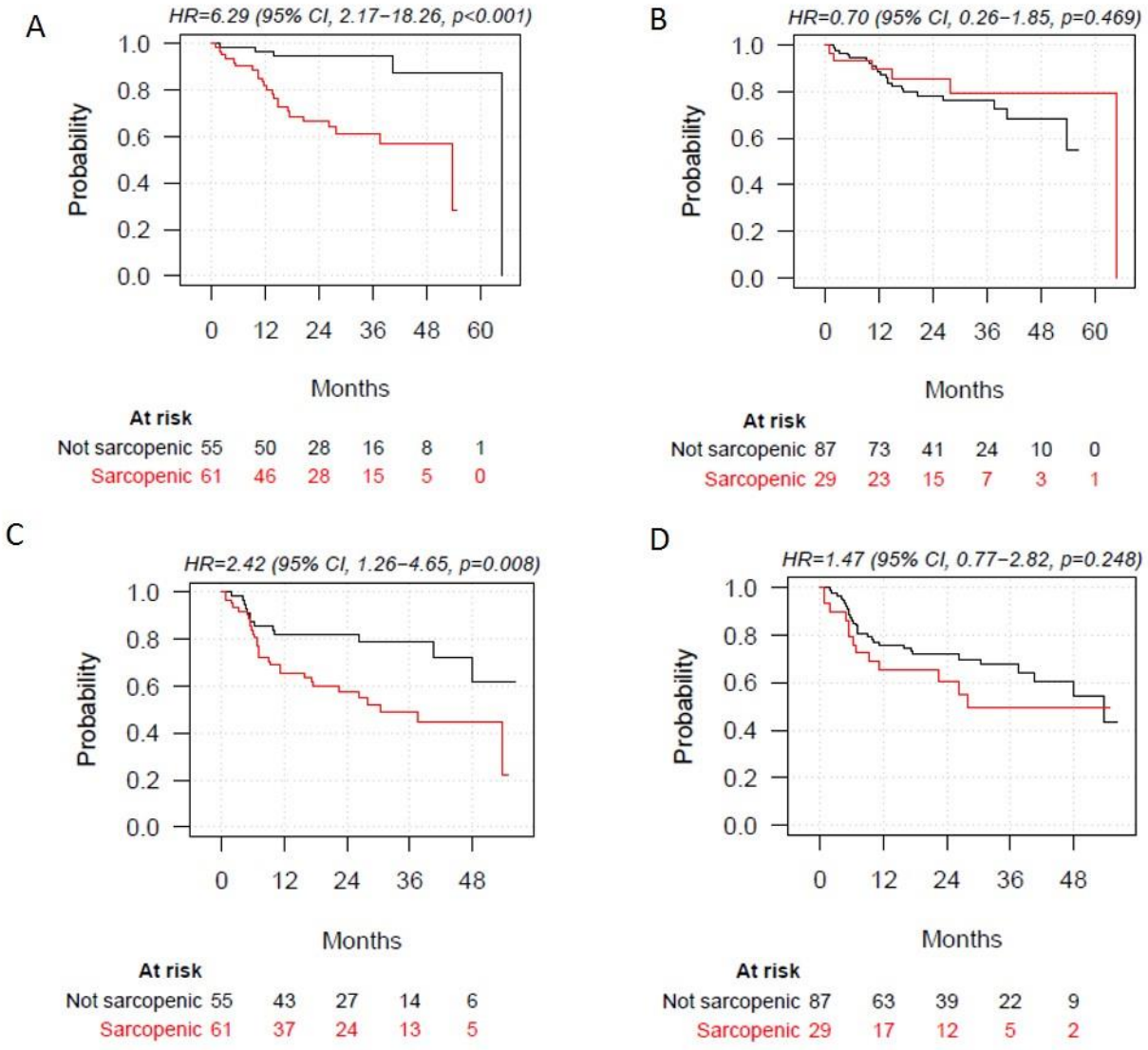


Figure 2. *Overall survival stratified by poor muscle quality defined according to skeletal muscle density at the level of the third lumbar vertebra (L3-SMD),* ***(A)*** *and decreased muscle mass according to skeletal muscle index at the level of the third lumbar vertebra (L3-SMI), (cut-off values: <43 cm^2/m^2 for men with BMI <25, <53 cm^2/m^2 for men with BMI ≥25, and BMI<41 cm^2/m^2 for women)* ***(B)****. Progression-free survival (PFS) stratified by poor muscle quality defined*

*according to L3-SMD (<41 HU for patients with BMI <25 and <33 HU for patients with BMI >25) (**C**) and decreased muscle mass according to L3-SMI (**D**). BMI, body mass index*

Multivariate Cox proportional hazard analyses, with BMI, male sex, IPI, and R-CHOP therapy as covariates, confirmed the lack of significant association of L3-SMI-defined muscle mass depletion with OS and PFS. The association between poor muscle quality defined according to L3-SMD and PFS was not statistically significant (HR=1.40, 95% CI=0.69-2.86, p=0.356), while a significant association remained for OS (HR=3.54; 95% CI=1.10-11.40; p=0.034) (Table 3).

TABLE 3. Multivariate Cox models for OS and PFS including muscle modification according to SMI or SMD

Overall survival							
	HR	95%CI	p		HR	95%CI	p
L3-SMI-based muscle depletion [a]	0.67	0.25-1.82	0.431	L3-SMD-based poor muscle quality	3.54	1.10-11.40	0.034
BMI	0.98	0.90-1.06	0.603	BMI	1.01	0.92-1.10	0.874
Male sex	1.04	0.47-2.28	0.932	Male sex	1.29	0.57-2.92	0.547
IPI	2.42	1.66-3.53	<0.001	IPI	2.03	1.35-3.04	<0.001
R-CHOP therapy	1.03	0.46-2.27	0.952	R-CHOP therapy	1.12	0.51-2.45	0.785
Progression free survival							
	HR	95%CI	p		HR	95%CI	p
L3-SMI-based muscle depletion [a]	1.53	0.78-3.00	0.216	L3-SMD-based poor muscle quality	1.40	0.69-2.86	0.356
BMI	0.98	0.91-1.05	0.509	BMI	0.98	0.91-1.05	0.503
Male sex	1.16	0.62-2.17	0.646	Male sex	1.16	0.62-2.16	0.649
IPI	1.92	1.46-2.53	<0.001	IPI	1.84	1.38-2.46	<0.001

R-CHOP therapy	0.81	0.43-1.51	0.504	R-CHOP therapy	0.90	0.48-1.68	0.737

Multivariate Cox models for overall and progression-free survival with BMI, male sex, and IPI as covariates. OS, overall survival; PFS, progression-free survival; SMD, skeletal muscle density; SMI, skeletal muscle index;
BMI, body mass index; IPI, international prognostic index; L3, third lumbar vertebra.

[a]*Results were similar when defining muscle mass depletion according to the other sets of L3-SMI (L3-SMI-2 and L3-SMI-3) cut-off values reported in the Methods section, PET-CT image analysis.*

3.6. Combined prognostic effect of poor muscle quality and obesity on survival

When stratifying patients according to BMI groups (<25; 25-30; ≥30), no significant difference was observed between groups in terms of OS (p=0.606) and PFS (p=0.437). Similar results were found when stratifying patients according to VAT tertiles (p=0.540 and p=0.549, respectively, for OS and PFS) (Supplementary Figure 3).

When further stratifying patients into four categories according to poor muscle quality defined based on L3-SMD and obesity (BMI≥30), a significant difference was found between categories in terms of OS (p=0.002) and PFS (p=0.033), with obese sarcopenic patients having the worst survival (Figure 3). Similar trends were found when stratifying patients according to poor muscle quality and visceral obesity defined based on VAT or VAT/SAT tertiles, with the worst survivals for sarcopenic patients with higher visceral adiposity (Supplementary Figure 4).

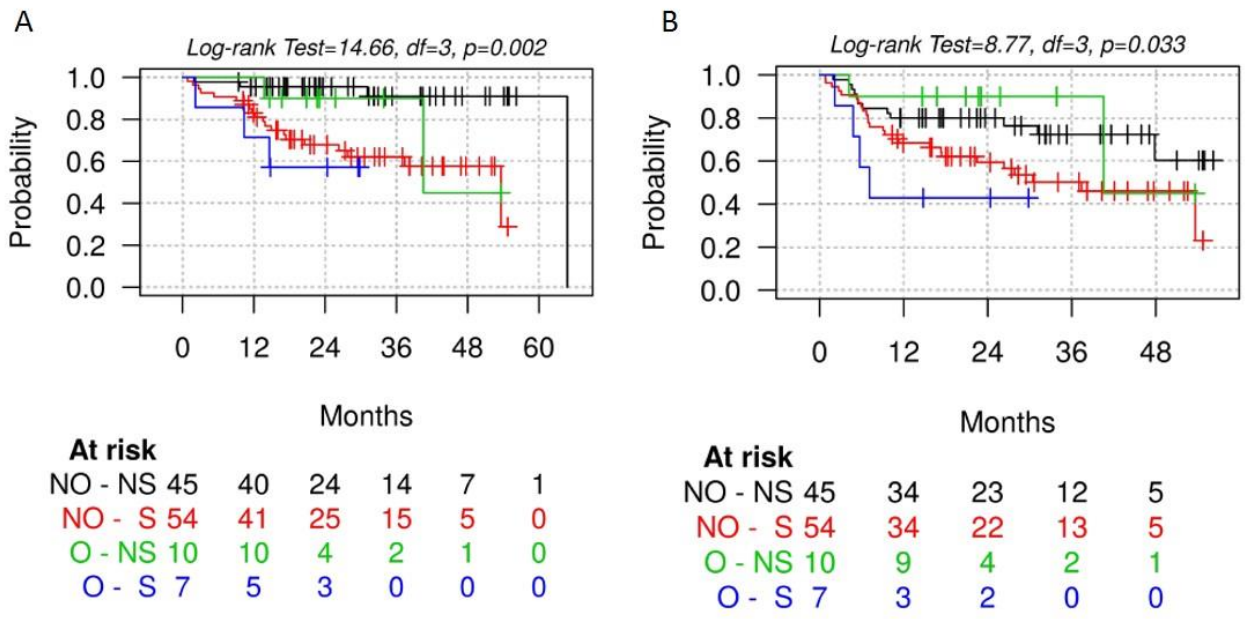


Figure 3. *Overall survival (**A**) and progression-free survival (**B**) stratified by poor muscle quality (according to L3-SMD cut-off values) and obesity (defined based on BMI ≥30). S, sarcopenic patients defined according to skeletal muscle density at the level of the third lumbar vertebra (L3-SMD); O, obese patients based on BMI ≥30; NS, non-sarcopenic patients according to L3-SMD; NO, non-obese patients*

3.7. Impact of PET-CT indices on early therapy termination

Early therapy termination was experienced by 16 patients (13.8%). In 3 patients, treatment was stopped for lack of response (2 disease progressions, 1 partial response), while in the remaining 13 patients, it was terminated due to grade >2 toxicity (8 infectious complications, 1 deep venous thrombosis, 1 intestinal occlusion, 1 intestinal perforation, 1 neuropathy, 1 cardiac impairment), which led to death in 4 patients.

Poor muscle quality defined according to L3-SMD was significantly associated with early therapy termination (p=0.028): 13/16 (81.3%) patients who had premature therapy termination due to toxicity belonged to the group with poorer muscle quality. This association was not found for muscle depletion based on L3-SMI cut-offs. However, patients who experienced early therapy termination had significantly lower PT-SMI (median: 65.2 vs 78.2 cm^2/m^2, p=0.041) and PT-SMD (41 vs 48 HU, p=0.023), with borderline significant PT-IMAT (17 vs 25, p=0.064).

4 Discussion

Body composition is known to affect clinical outcome in cancer patients. The assessment of sarcopenia and other body composition indices by means of routinely performed pretreatment diagnostic imaging is of great importance in clinical oncology. We have studied body composition based on DLBCL patients' basal PET-CT images in relation to the clinical outcome. Our findings show that a reduced L3-SMD or PT-SMD or increased PT-IMAT is associated with a low PFS and OS in DLBCL. The measurement of sarcopenia indices not only at the L3 level, as explored in the previous literature, but also at the PT level represents a novelty. Considering the absence of cut-off values for PT measurements, these indices were evaluated as continuous variables, and PT-SMD performed as a superior survival prognostic index over L3-SMD. Furthermore, when measured at the PT level, the expansion of intermuscular fat (PT-IMAT) was found to be associated with a significant reduction in OS and PFS. The prognostic value of IMAT has never been described previously in patients with DLBCL.

As a continuous variable, the reduction of PT-SMD showed to be a robust and independent prognostic index of PFS and OS in a multivariate analysis. At the L3 level, if we consider a cut-off of 41 HU for patients with BMI <25 and 33 HU for patients with BMI >25 [10, 25], SMD was an independent negative prognostic factor for OS but not for PFS. The negative correlation between the L3-SMD, serum albumin, and total serum protein values and the positive correlation with PCR suggest an association between muscle fat depots and the systemic inflammatory status of the patients.

The etiology of cancer-associated muscle alterations is multifactorial. A higher IMAT and lower SMD are caused by an increase in inter- and intramuscular fat, respectively, which is often associated with systemic inflammation and metabolic derangement [13]. Lymphoma cells or the body's immune response to cancer cells can aggravate the systemic inflammatory response. The inflammatory status and concomitant alterations in mitochondrial function can reduce the ability of the muscle fiber to oxidize lipids, thus increasing fat depots and decreasing the cell energy reserve [14] with the consequence of a reduction in muscle strength and physical performance [32]. Besides an ectopic fat deposition, these muscle changes might be indices of a more severe disease or more intense host pro-inflammatory response leading to worse clinical outcomes. We can only speculate why the indices of fat accumulation in the muscles of the PT are stronger predictors than those at the L3 level. A decreased muscle radio-density has been related to reduced strength and performance [33-35], and a reduced quality of the muscles of the lower limb and thigh might be a

better predictor of poor muscle function and physical performance status, indicating a more severe sarcopenic status [36, 37]. In addition, the muscles of the thigh might be more sensitive to the effects of metabolic or inflammatory derangement. Whatever the cause may be, SMD and IMAT at the thigh level represent new prognostic indices of clinical outcome or predictors of response to therapy in patients with DLBCL.

In our study, a decreased SMD plays a more potent role than obesity, defined by a high BMI, or visceral obesity, defined by increased VAT, in influencing patients' prognosis. However, sarcopenia and obesity can interact and aggravate each other, leading to sarcopenic obesity. The obese sarcopenic patients with a higher BMI or VAT had poorer survival outcomes.

Sarcopenia defined according to a low SMD was also significantly associated with early therapy termination due to toxicity. This result emphasizes the important potential implications that body composition has on the prescription of immunochemotherapy.

We have found that sarcopenia influenced patient survival only when it was defined as low SMD but not when defined as low SMI, regardless of the use of different cut-offs. This confirms the findings of Chu et al. [6] suggesting that poor muscle quality rather than muscle mass depletion affects survival. In the DLBCL patients that we have investigated, a reduction of SMI had no adverse prognostic value. In fact, in the multivariate analysis, a decreased SMI was associated with significantly improved OS among patients. There is a debate concerning whether sarcopenia due to muscle loss is directly associated with reduced response to anticancer therapy. While Lanic et al. indicated that sarcopenia assessed by a low SMI is predictive of a worse PFS and OS in DLBC patients [5], Chu et al. showed that DLBCL patients with low SMI demonstrated a trend toward improved PFS [6]. A possible explanation for the better outcomes in our patients with a low muscle mass and reduced L3-SMI may be related to the exposure to relatively higher plasma concentrations of rituximab due to the lower volumes of distribution and reduced clearance of this anticancer drug [6, 38].

This study has several limitations. First, due to the retrospective nature of the research, prospective confirmation of the results is required. In future investigations, the performance status of the patients should be tested in order to correlate muscle function with muscle fat infiltrations at the thigh level. Second is the therapeutic regimen. While the majority of patients have been treated with R-CHOP, few patients received different therapies. In order to reduce the effect of the therapy variable, the multivariate regression analyses for survival were adjusted for R-CHOP therapy.

In conclusion, we have shown that muscle quality, and particularly the inter- and intramuscular fat infiltration at baseline, at the PT level rather than at the L3 level, are associated with a worse prognosis in patients with DLBCL. A staging PET-CT scan, performed as part of standard clinical practice, may serve as a powerful and inexpensive tool to assess body composition metrics. These indicators can complement classical prognostic indices, such as the IPI, which are based on patient characteristics directly associated with the disease. This information, if confirmed in prospective studies, will pave the way for nutritional and physical activity interventions with the aim of improving the body composition, and accordingly the clinical outcomes, of patients with DLBCL.

References

1. Miranda-Filho A, Piñeros M, Znaor A, Marcos-Gragera R, Steliarova-Foucher E, Bray F. Global patterns and trends in the incidence of non-Hodgkin lymphoma. *Cancer Causes Control CCC* (2019) **30**:489–499. doi:10.1007/s10552-019-01155-5

2. Morton LM, Wang SS, Devesa SS, Hartge P, Weisenburger DD, Linet MS. Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. *Blood* (2006) **107**:265–276. doi:10.1182/blood-2005-06-2508

3. Coiffier B, Lepage E, Brière J, Herbrecht R, Tilly H, Bouabdallah R, Morel P, Van Den Neste E, Salles G, Gaulard P, et al. CHOP Chemotherapy plus Rituximab Compared with CHOP Alone in Elderly Patients with Diffuse Large-B-Cell Lymphoma. *N Engl J Med* (2002) **346**:235–242. doi:10.1056/NEJMoa011795

4. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med* (1993) doi:10.1056/nejm199309303291402

5. Lanic H, Kraut-Tauzia J, Modzelewski R, Clatot F, Mareschal S, Picquenot JM, Stamatoullas A, Leprêtre S, Tilly H, Jardin F. Sarcopenia is an independent prognostic factor in elderly patients with diffuse large B-cell lymphoma treated with immunochemotherapy. *Leuk Lymphoma* (2014) **55**:817–823. doi:10.3109/10428194.2013.816421

6. Chu MP, Lieffers J, Ghosh S, Belch A, Chua NS, Fontaine A, Sangha R, Turner RA, Baracos VE, Sawyer MB. Skeletal muscle density is an independent predictor of diffuse large B-cell lymphoma outcomes treated with rituximab-based chemoimmunotherapy. *J Cachexia Sarcopenia Muscle* (2017) **8**:298–304. doi:10.1002/jcsm.12161

7. Go S-I, Park MJ, Song H-N, Kim H-G, Kang MH, Lee HR, Kim Y, Kim RB, Lee SI, Lee G-W. Prognostic impact of sarcopenia in patients with diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone: Sarcopenia in diffuse large B-cell lymphoma. *J Cachexia Sarcopenia Muscle* (2016) **7**:567–576. doi:10.1002/jcsm.12115

8. Nakamura N, Hara T, Shibata Y, Matsumoto T, Nakamura H, Ninomiya S, Kito Y, Kitagawa J, Kanemura N, Goto N, et al. Sarcopenia is an independent prognostic factor in male patients with diffuse large B-cell lymphoma. *Ann Hematol* (2015) **94**:2043–2053. doi:10.1007/s00277-015-2499-4

9. Burkart M, Schieber M, Basu S, Shah P, Venugopal P, Borgia JA, Gordon L, Karmali R. Evaluation of the impact of cachexia on clinical outcomes in aggressive lymphoma. *Br J Haematol* (2019) **186**:45–53. doi:10.1111/bjh.15889

10. Chu MP, Lieffers J, Ghosh S, Belch AR, Chua NS, Fontaine A, Sangha R, Turner AR, Baracos VE, Sawyer MB. Skeletal muscle radio-density is an independent predictor of response and outcomes in follicular lymphoma treated with chemoimmunotherapy. *PloS One* (2015) **10**:e0127589. doi:10.1371/journal.pone.0127589

11. Goodpaster BH, Kelley DE, Thaete FL, He J, Ross R. Skeletal muscle attenuation determined by computed tomography is associated with skeletal muscle lipid content. *J Appl Physiol Bethesda Md 1985* (2000) **89**:104–110.

12. Yamashita M, Kamiya K, Matsunaga A, Kitamura T, Hamazaki N, Matsuzawa R, Nozaki K, Tanaka S, Nakamura T, Maekawa E, et al. Prognostic Value of Psoas Muscle Area and Density in Patients Who Undergo Cardiovascular Surgery. *Can J Cardiol* (2017) **33**:1652–1659. doi:10.1016/j.cjca.2017.10.009

13. Addison O, Marcus RL, Lastayo PC, Ryan AS. Intermuscular fat: a review of the consequences and causes. *Int J Endocrinol* (2014) **2014**:309570. doi:10.1155/2014/309570

14. Tanaka M, Okada H, Hashimoto Y, Kumagai M, Nishimura H, Oda Y, Fukui M. Relationship between metabolic syndrome and trunk muscle quality as well as quantity evaluated by computed tomography. *Clin Nutr* (2019) doi:10.1016/j.clnu.2019.07.021

15. Shulman GI. Ectopic fat in insulin resistance, dyslipidemia, and cardiometabolic disease. *N Engl J Med* (2014) **371**:1131–1141. doi:10.1056/NEJMra1011035

16. Sabel MS, Lee J, Cai S, Englesbe MJ, Holcombe S, Wang S. Sarcopenia as a prognostic factor among patients with stage III melanoma. *Ann Surg Oncol* (2011) **18**:3579–3585. doi:10.1245/s10434-011-1976-9

17. Antoun S, Lanoy E, Iacovelli R, Albiges-Sauvin L, Loriot Y, Merad-Taoufik M, Fizazi K, di Palma M, Baracos VE, Escudier B. Skeletal muscle density predicts prognosis in patients with metastatic renal cell carcinoma treated with targeted therapies. *Cancer* (2013) **119**:3377–3384. doi:10.1002/cncr.28218

18. Kazemi-Bajestani SMR, Mazurak VC, Baracos V. Computed tomography-defined muscle and fat wasting are associated with cancer clinical outcomes. *Semin Cell Dev Biol* (2016) **54**:2–10. doi:10.1016/j.semcdb.2015.09.001

19. Carson KR, Bartlett NL, McDonald JR, Luo S, Zeringue A, Liu J, Fu Q, Chang S-H, Colditz GA. Increased Body Mass Index Is Associated With Improved Survival in United States Veterans With Diffuse Large B-Cell Lymphoma. *J Clin Oncol* (2012) **30**:3217. doi:10.1200/JCO.2011.39.2100

20. Geyer SM, Morton LM, Habermann TM, Allmer C, Davis S, Cozen W, Severson RK, Lynch CF, Wang SS, Maurer MJ, et al. Smoking, alcohol use, obesity, and overall survival from non-Hodgkin lymphoma. *Cancer* (2010) **116**:2993–3000. doi:10.1002/cncr.25114

21. Hong F, Habermann TM, Gordon LI, Hochster H, Gascoyne RD, Morrison VA, Fisher RI, Bartlett NL, Stiff PJ, Cheson BD, et al. The role of body mass index in survival outcome for lymphoma patients: US intergroup experience. *Ann Oncol* (2014) **25**:669–674. doi:10.1093/annonc/mdt594

22. Jones JA, Fayad LE, Elting LS, Rodriguez MA. Body mass index and outcomes in patients receiving chemotherapy for intermediate-grade B-cell non-Hodgkin lymphoma. *Leuk Lymphoma* (2010) **51**:1649–1657. doi:10.3109/10428194.2010.494315

23. Shin D-Y, Kim A, Byun BH, Moon H, Kim S, Ko Y-J, Kim M-J, Lee H-R, Kang H-J, Na II, et al. Visceral adipose tissue is prognostic for survival of diffuse large B cell lymphoma treated with frontline R-CHOP. *Ann Hematol* (2016) **95**:409–416. doi:10.1007/s00277-015-2571-0

24. Mourtzakis M, Prado CMM, Lieffers JR, Reiman T, McCargar LJ, Baracos VE. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. *Appl Physiol Nutr Metab Physiol Appl Nutr Metab* (2008) **33**:997–1006. doi:10.1139/H08-075

25. Martin L, Birdsell L, MacDonald N, Reiman T, Clandinin MT, McCargar LJ, Murphy R, Ghosh S, Sawyer MB, Baracos VE. Cancer Cachexia in the Age of Obesity: Skeletal Muscle Depletion Is a Powerful Prognostic Factor, Independent of Body Mass Index. *J Clin Oncol* (2013) **31**:1539–1547. doi:10.1200/JCO.2012.45.2722

26. Prado CM, Lieffers JR, McCargar LJ, Reiman T, Sawyer MB, Martin L, Baracos VE. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. *Lancet Oncol* (2008) **9**:629–635. doi:10.1016/S1470-2045(08)70153-0

27. Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. *J Am Stat Assoc* (1927) **22**:209–212. doi:10.1080/01621459.1927.10502953

28. Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials* (1996) **17**:343–346. doi:10.1016/0197-2456(96)00075-x

29. Wilson WH, Grossbard ML, Pittaluga S, Cole D, Pearson D, Drbohlav N, Steinberg SM, Little RF, Janik J, Gutierrez M, et al. Dose-adjusted EPOCH chemotherapy for untreated large B-cell lymphomas: a pharmacodynamic approach with high efficacy. *Blood* (2002) **99**:2685–2693. doi:10.1182/blood.v99.8.2685

30. Maruyama D, Watanabe T, Maeshima AM, Nomoto J, Taniguchi H, Azuma T, Mori M, Munakata W, Kim S-W, Kobayashi Y, et al. Modified cyclophosphamide, vincristine, doxorubicin, and methotrexate (CODOX-M)/ifosfamide, etoposide, and cytarabine (IVAC) therapy with or without rituximab in Japanese adult patients with Burkitt lymphoma (BL) and B cell lymphoma, unclassifiable, with features intermediate between diffuse large B cell lymphoma and BL. *Int J Hematol* (2010) **92**:732–743. doi:10.1007/s12185-010-0728-0

31. Musso M, Scalone R, Marcacci G, Lanza F, Di Renzo N, Cascavilla N, Di Bartolomeo P, Crescimanno A, Perrone T, Pinto A. Fotemustine plus etoposide, cytarabine and melphalan (FEAM) as a new conditioning regimen for lymphoma patients undergoing auto-SCT: a

multicenter feasibility study. *Bone Marrow Transplant* (2010) **45**:1147–1153. doi:10.1038/bmt.2009.318

32. Heymsfield SB, Gonzalez MC, Lu J, Jia G, Zheng J. Skeletal muscle mass and quality: evolution of modern measurement concepts in the context of sarcopenia. *Proc Nutr Soc* (2015) **74**:355–366. doi:10.1017/S0029665115000129

33. Goodpaster BH, Carlson CL, Visser M, Kelley DE, Scherzinger A, Harris TB, Stamm E, Newman AB. Attenuation of skeletal muscle and strength in the elderly: The Health ABC Study. J Appl Physiol Bethesda Md 1985 (2001) 90:2157–2165.

34. Visser M, Kritchevsky SB, Goodpaster BH, Newman AB, Nevitt M, Stamm E, Harris TB. Leg muscle mass and composition in relation to lower extremity performance in men and women aged 70 to 79: the health, aging and body composition study. J Am Geriatr Soc (2002) 50:897–904. doi:10.1046/j.1532-5415.2002.50217.x

35. Beavers KM, Beavers DP, Houston DK, Harris TB, Hue TF, Koster A, Newman AB, Simonsick EM, Studenski SA, Nicklas BJ, et al. Associations between body composition and gait-speed decline: results from the Health, Aging, and Body Composition study. *Am J Clin Nutr* (2013) **97**:552–560. doi:10.3945/ajcn.112.047860

36. Addison O, Young P, Inacio M, Bair W-N, Prettyman MG, Beamer BA, Ryan AS, Rogers MW. Hip but Not Thigh Intramuscular Adipose Tissue is Associated with Poor Balance and Increased Temporal Gait Variability in Older Adults. *Curr Aging Sci* (2014) **7**:137–143. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4480674/ [Accessed January 15, 2020]

37. Cruz-Jentoft AJ, Bahat G, Bauer J, Boirie Y, Bruyère O, Cederholm T, Cooper C, Landi F, Rolland Y, Sayer AA, et al. Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing* (2018) doi:10.1093/ageing/afy169

38. Müller C, Murawski N, Wiesen MHJ, Held G, Poeschel V, Zeynalova S, Wenger M, Nickenig C, Peter N, Lengfelder E, et al. The role of sex and weight on rituximab clearance and serum elimination half-life in elderly patients with DLBCL. *Blood* (2012) **119**:3276–3284. doi:10.1182/blood-2011-09-380949

6.2 The impact of chest CT body composition parameters on clinical outcomes in COVID-19 patients*

Summary: Body composition is likely to have an impact on COVID-19 course, since ectopic fat is linked to systemic inflammation which may favor the COVID-19 cytokine storm, and sarcopenia may affect respiratory efficiency. Moreover, elderly patients, typically sarcopenic, are known to have a very poor prognosis. Therefore, we conducted a retrospective study to evaluate the impact of CT biomarkers of body composition on COVID-19 outcomes.

We assessed the impact of chest CT body composition parameters on outcomes and disease severity at hospital presentation of COVID-19 patients, focusing also on the possible mediation of body composition in the relationship between age and death in these patients. Chest CT scans performed at hospital presentation by consecutive COVID-19 patients (02/27/2020-03/13/2020) were retrospectively reviewed to obtain pectoralis muscle density and total, visceral, and intermuscular adipose tissue areas (TAT, VAT, IMAT) at T7-T8 vertebrae. Primary outcomes were: hospitalization, mechanical ventilation (MV) and/or death, death alone; secondary outcomes: C-reactive protein (CRP), oxygen saturation (SO2), CT disease extension at hospital presentation. The mediation of body composition in the effect of age on death was explored. Of the 318 patients included in the study (median age 65.7 years, females 37.7%), 205 (64.5%) were hospitalized, 68 (21.4%) needed MV, and 58 (18.2%) died. Increased muscle density was a protective factor while increased TAT, VAT, and IMAT were risk factors for hospitalization and MV/death. All these parameters except TAT had borderline effects on death alone. All parameters were associated with SO2 and extension of lung parenchymal involvement at CT; VAT was associated with CRP. Approximately 3% of the effect of age on death was mediated by decreased muscle density. In conclusion, low muscle quality and ectopic fat accumulation were associated with COVID-19 outcomes, VAT was associated with baseline inflammation. Low muscle quality partly mediated the effect of age on mortality

*Manuscript Under Review in *Plos One* after the first round of revisions. Besutti G, Pellegrini M, Ottone M, Cantini M, Milic J, Bonelli E, Dolci G, Cassone G, Ligabue G, Spaggiari L, Pattacini P, Fasano T, Canovi S, Massari M, Salvarani C, Guaraldi G, Giorgi Rossi P.

Supplementary Material in Appendix E

INTRODUCTION

A novel severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) has recently emerged as a global health threat [1-2]. As of 15 February, over 108 million people had been affected with more than 2.390.000 deaths reported worldwide so far [3]. The case fatality rate varies dramatically across countries and phases of the epidemic, ranging from 2% to 20%, depending on the characteristics of the population and the ability of the health system to identify less severe cases. Most severe COVID-19 patients develop acute respiratory distress syndrome (ARDS) or sepsis with multiorgan dysfunction [1, 4-5], often associated with an uncontrolled cytokine-mediated immune response called the *cytokine storm*. Of these patients, 71-75% need assisted mechanical ventilation and about 50% die [1-2, 5-7]. Obesity and advanced age are among the most important recognized risk factors for an unfavorable outcome in COVID-19 patients [1-2, 8-10].

Obesity was considered a risk factor during the previous H1N1 virus outbreak as well [11], and it is not surprising that SARS-CoV-2 pneumonia is also negatively affected by overweight [12]. A higher percentage of body fat mass is associated with a greater cardiometabolic risk, but not all body fat deposits have the same significance. While body mass index (BMI) represents a useful but rough index of general adiposity, ectopic visceral, hepatic, and muscular fat depots are associated with increased production of pro-inflammatory cytokines and with a higher incidence of cardiovascular events, insulin resistance, and type 2 diabetes [13-14]. In this *milieu,* the onset of the COVID-19 cytokine storm may be favored.

Advanced age, another important risk factor for the more severe forms of COVID-19, is characterized by a progressive reduction of body muscle mass and function (sarcopenia) associated with a progressive accumulation of fat deposits in the muscle (myosteatosis) [15-16]. Both sarcopenia and myosteatosis impact medical and surgical outcomes and are reliable predictors of all-cause mortality [15, 17]. Body composition parameters are commonly studied by collecting CT cross-sectional areas at the level of L3 vertebra, linearly related to whole body muscle and fat mass [18]. Myosteatosis can be apparent within muscle fibers and is assessed through CT scans by measuring skeletal muscle density (SMD). In addition, it can be detected across muscle fibers and within the fascia, where it is assessed through CT scans by measuring the adipose tissue between muscles (intermuscular adipose tissue).

Recent publications have reported that higher BMI, higher abdominal visceral adipose tissue, higher intermuscular adipose tissue, reduced liver density, reduced lumbar SMD, and reduced pectoral muscle area are associated with worse clinical outcomes in COVID-19 patients in terms of disease severity and death [5, 19-25]. While CT scans of the abdomen are rarely available for unselected series of COVID-19 patients, chest CT has been widely used in some centers to rapidly assess the

presence of pneumonia and stratify patients with different disease severity [26]. Body composition parameters measured on chest CT scans show moderate to high correlation both with abdominal CT fat compartments and skeletal muscle mass measured with bioelectrical impedance analysis [27-28]. Moreover, like abdominal VAT, intrathoracic VAT (epicardial and extracardiac) is associated with the production of systemic inflammatory markers [29], and thoracic and pectoral muscle quantity and quality may be predictive of clinical outcomes in different respiratory diseases, especially in patients requiring mechanical ventilation [30-32].

The possible role of body composition parameters as prognostic factors for COVID-19 severity has been initially explored [5, 19-20, 22-25]. However, the underlying causal relationship remains to be determined and contextualized in the complex pathogenetic pathways involved in COVID-19 progression. Therefore, we first investigated the association of chest CT-derived body composition parameters with clinical outcomes in COVID-19 patients, including hospitalization, mechanical ventilation (MV) or death, and death alone. We also explored the association between body composition parameters and biomarkers of disease progression at emergency room presentation: oxygen saturation and extension of parenchymal involvement at CT for the lung damage, and C-reactive protein for the inflammatory reaction. Through a mediation analysis of the factors associated with age and death, secondly, we evaluated whether the effect of age on death is partly mediated by body composition.

MATERIALS AND METHODS

Setting

In the Reggio Emilia province (Northern Italy, 532,000 inhabitants, six hospitals), the first case of SARS-CoV-2 infection was diagnosed on 27 February 2020. As of 13 March 2020, there were 1,154 RT-PCR-confirmed COVID-19 patients in the province, with the daily number of new cases rising steadily.

Study design and population

This observational study was approved by the Area Vasta Emilia Nord Ethics Committee on 7 April 2020 (protocol number 2020/0045199) and performed in accordance with the ethical standards of the Declaration of Helsinki. Patients' informed consent to participate in the study was obtained whenever possible, given the retrospective nature of the study.

All consecutive patients were included who presented to the provincial emergency rooms (ERs) between 27 February and 13 March 2020 for suspected COVID-19: these underwent chest CT at ER presentation and tested positive on RT-PCR for SARS-CoV-2 within 10 days. During the COVID-19 outbreak, virtually all symptomatic patients with suspected COVID-19 pneumonia were referred

to CT. Patients with CT scans not suitable for different post-processing evaluations were excluded from specific study analyses, e.g., CTs with a small field of view were not suitable for evaluation of subcutaneous adipose tissue, CTs of patients with thoracic lipomas were excluded from the evaluation of fat compartments, and CTs with artifacts due to pacemakers or other implants were not suitable for pectoral muscle segmentation.

Outcomes

The main outcomes considered were death, hospitalization, and death or mechanical ventilation while being a COVID-19 patient. We included all outcomes occurring between ER presentation and before symptom remission and two negative RT-PCR tests or end of follow up, i.e., 21 April 2020.

Data collection

Date of symptom onset, diagnosis, hospitalization, and death were retrieved from the COVID-19 Surveillance Registry, coordinated by the Italian National Institute of Health and implemented in each Local Health Authority [33]. Registry data were linked with the hospital radiology information system to search for CTs performed at or after the onset of COVID symptoms and with hospital discharge databases to collect information on comorbidities. The Charlson Index was calculated based on hospital admissions in the previous 10 years [34]. BMI was calculated whenever patient height and weight registered within six months preceding COVID-19 diagnosis were available from the hospital information systems. Diabetes was ascertained through linkage with the local Diabetes Registry [35]. The need for invasive or non-invasive MV during hospitalization was manually collected from medical records.

Blood tests

At ER presentation the levels of C-reactive protein, lactate dehydrogenase (LDH), white blood cell, lymphocyte, neutrophil, and platelet counts were routinely collected. Oxygen saturation level was also recorded for patients who had an arterial blood gas analysis before being provided with oxygen support. The tests were carried out in the Hospital Clinical Laboratories with routine automated methods.

CT acquisition technique

CT scans were performed using one of three scanners (128-slice Somatom Definition Edge, Siemens Healthineers; 64-slice Ingenuity, Philips Healthcare; 16-slice GE Brightspeed, GE Healthcare) without contrast media injection, with the patient in supine position during end-inspiration. Scanning parameters were tube voltage 120 KV, automatic tube current modulation, collimation width 0.625 or 1.25 mm, acquisition slice thickness 2.5 mm, and interval 1.25 mm. Images were reconstructed with a high-resolution algorithm at slice thickness 1.0/1.25 mm.

CT retrospective analysis

To evaluate COVID-19 pneumonia extension, CT scans were retrospectively reviewed by a chest radiologist with 15-year experience (LS), who graded extension of pulmonary lesions using a visual scoring system (< 20%, 20-39%, 40-59%, ≥ 60%) [26].

To evaluate body composition parameters, CT images were retrospectively analyzed by a single trained image analyzer (EB) supervised by a senior radiologist (PP), both blinded to clinical data and outcomes, by using the OSIRIX-Lite software V5.0 (Pixmeo, Sarl, Switzerland) (S1 Fig).

As measures of sarcopenia, pectoralis muscle cross-sectional area (cm^2) and mean density (Hounsfield Unit, HU) were obtained selecting a single axial slice directly superior to the aortic arch and manually contouring both pectoralis major and minor on the right side (or on the left side when a defibrillator was present on the right), after applying a density range of -29 to 150 HU [36].

For total, subcutaneous, visceral, and intermuscular adipose tissue areas (TAT, SAT, VAT, and IMAT), a single slice at the level of the seventh to eighth thoracic vertebrae (T7-T8) was selected and a density range from -190 to -30 HU was applied. Fat compartment was measured through autosegmentation, with manual contour correction when necessary [27].

Mean liver and spleen attenuation values (HU) were obtained by drawing nine regions of interest (ROIs) in the liver and three ROIs in the spleen, paying attention to avoid vessels, bile ducts, focal lesions, focal fatty changes, and visceral margins.

For all retrospective measures, a second measurement was obtained in a sample of 15 consecutive patients by the same reader after two months, in order to test intrareader agreement.

Statistical analyses

Continuous variables are reported as median and interquartile range, and categorical variables as proportions. CT body composition parameters were considered as continuous variables. We calculated Spearman correlation to assess the association among different fat distribution indices as well as between age and CT body composition parameters.

We checked the linearity between continuous predictor variables and the logit of the outcome, and univariate logistic regression analyses were performed to identify the main CT body composition parameters influencing adverse outcomes (hospitalization, MV or death, death alone) in COVID-19 patients. For these parameters and for each outcome, we applied a multivariate logistic model adjusted for sex, age, and calendar period (in weeks since the beginning of the outbreak). We choose not to adjust for patient conditions at disease onset since these could be causally linked to body composition. Furthermore, we did not adjust for cardiovascular and metabolic pre-existing conditions because they can be mediators in the relationship between body composition and outcomes. Hospitalization, mechanical ventilation (MV) and/or death, and mortality at 40 days odds ratios (OR) with 95%

confidence intervals (95% CI) are reported for unit increase of CT body composition parameters (HU for pectoral density and cm^2 for adipose tissue variables). Only the OR of IMAT for mortality is reported for IMAT quartiles.

As sensitivity analyses, we restricted the sample to patients with no comorbidities, diabetes only, and cardiovascular comorbidities only.

We also tested the association between the body composition and disease severity at ER presentation using the following biomarkers: CRP as an indicator of cytokine storm intensity; SO2 and CT disease extension as indices of the degree of lung parenchyma involvement. The associations were investigated using multivariate linear regression models adjusted for sex, age, and calendar period.

Lastly, we analyzed the relationship between age and body composition parameters. A mediation analysis was conducted to assess to what degree body composition parameters could explain the effect of age on death by using logit model adjusted for sex, age, and calendar period. This analysis subdivided the total effect into indirect effects representing the causal mechanism through body composition, as opposed to direct effects represented by all other mechanisms [37].

Intrareader agreement was evaluated by Spearman's rank correlation coefficient and respective p-value. Data analysis was performed using Stata 13.0 SE (Stata Corporation, Texas, TX).

RESULTS

Study population

Of the 488 RT-PCR-positive patients presenting to the ER in the time period under study, we included 318 consecutive patients (median age 65.7 years, females 37.7 %) satisfying the inclusion criteria (Fig 1). The remaining 170 patients did not undergo CT scan primarily because chest X-rays and clinical presentation did not suggest pneumonia, and none of them died or received MV during follow up. Patient characteristics are reported in Table 1. During follow up, 205 (64.47%) hospitalizations and 58 (18.24%) deaths were registered; 68 (21.4%) patients were treated with invasive or non-invasive MV, and a total of 97 (30.5%) patients died or needed MV.

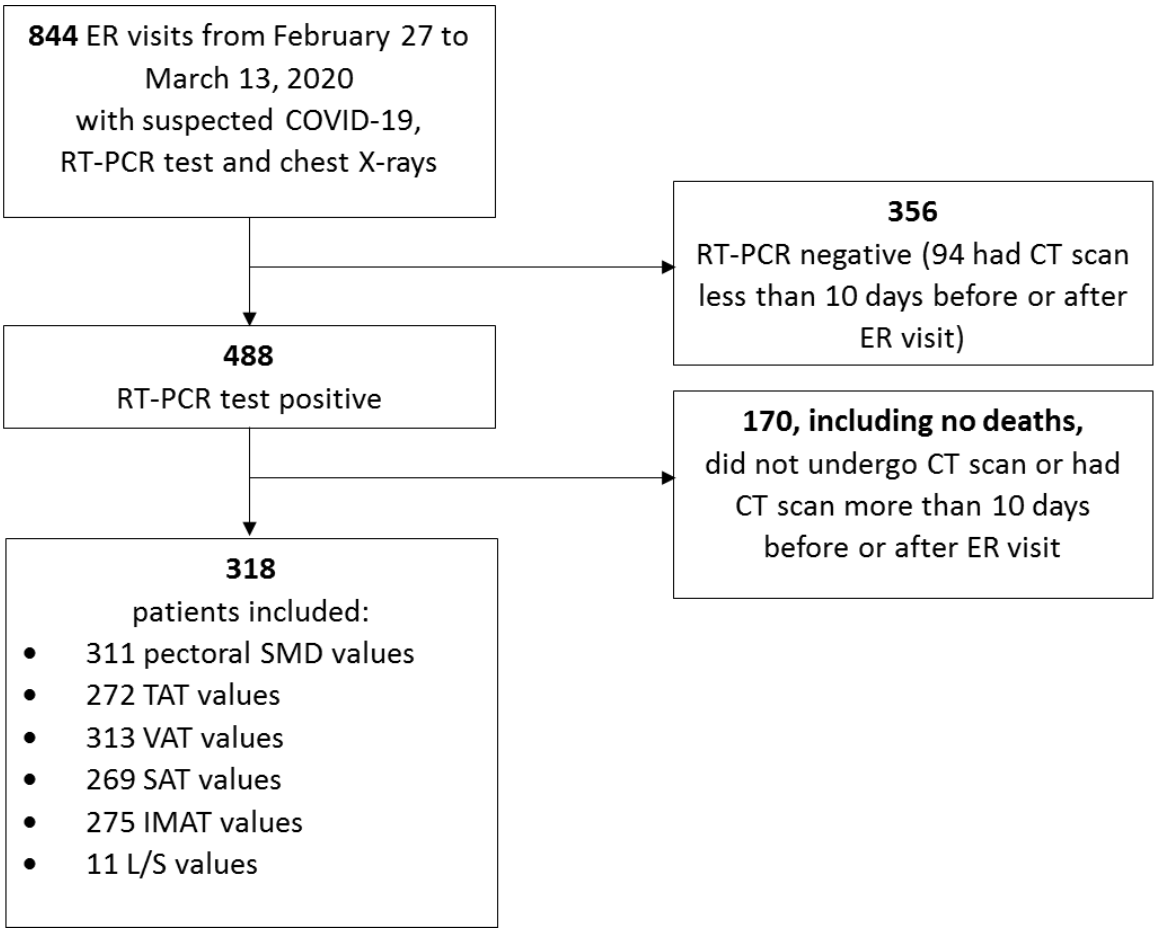


Figure 1. Flowchart describing patient selection.

Table 1. Clinical and body composition parameters in the population as a whole and in patients experiencing different outcomes

Variables	All Patients	Hospitalization	Mechanical Ventilation	Death	Mechanical Ventilation or Death
		N (%)	N (%)	N (%)	N (%)
	318	**205 (64.47)**	**68 (21.38)**	**58 (18.24)**	**97 (30.50)**
Age (years)	65.7 (52.8; 75.7)	71.8 (61.4; 79.8)	69.8 (63.2; 77.6)	79.8 (72.5; 85.0)	73.8 (66.4;82.5)
Females	120 (37.7)	69 (57.5)	16 (13.3)	13 (10.8)	27 (22.5)
Calendar period (Week 1)	36 (11.3)	27 (75.0)	15 (41.7)	8 (22.2)	17 (47.2)
(Week 2)	167 (52.5)	123 (73.7)	42 (25.2)	40 (24.0)	61 (36.5)
(Week 3)	115 (36.16)	55 (47.8)	11 (9.6)	10 (8.7)	19 (16.5)
Charlson Comorbidity Index (0)	239 (75.16)	134 (56.1)	45 (18.8)	27 (11.3)	58 (24.3)
(1)	22 (6.92)	18 (81.8)	7 (31.8)	7 (31.8)	11 (50.0)
(2)	20 (6.29)	18 (90.0)	6 (30.0)	5 (25.0)	8 (40.0)
(3)	37 (11.64)	35 (94.6)	10 (27.0)	19 (51.4)	20 (54.1)
Diabetes	43 (13.52)	41 (95.4)	20 (46.5)	11 (25.6)	23 (53.5)
COPD	10 (3.14)	10 (100)	3 (30.0)	7 (70.0)	9 (90.0)
Dementia	1 (0.31)	1 (100)	1 (100)	1 (100)	1 (100)
Chronic kidney failure	3 (0.94)	3 (100)	2 (66.7)	2 (66.7)	3 (100)
Previous cancer diagnosis	51 (16.04)	43 (84.3)	15 (29.4)	14 (27.5)	20 (39.2)
Hypertension	56 (17.61)	49 (87.5)	21 (37.5)	20 (35.7)	27 (48.2)
Arrhythmias	24 (7.55)	22 (91.7)	7 (29.2)	12 (50.0)	14 (58.3)
Cardiovascular diseases	47 (14.78)	42 (89.4)	16 (34.0)	22 (46.8)	26 (55.3)
Days from symptom onset	7 (4; 8)	6 (4;8)	6 (5; 7)	5 (2;7)	5 (3;7)
White blood cells (10^9/L)	5.22 (4.14; 6.63)	5.59 (4.11; 6.87)	5.82 (4.17; 7.18)	6.27 (4.54; 8.05)	5.86 (4.31; 7.58)

Lymphocytes (10^9/L)	0.96 (0.71; 1.34)	0.88 (0.68; 1.25)	0.84 (0.63; 1.00)	0.78 (0.49; 0.92)	0.83 (0.61; 1)
Neutrophils (10^9/L)	3.84 (2.95; 4.75)	4.10 (2.83; 5.29)	4.57 (2.94; 5.80)	4.69 (3.50; 6.33)	4.62 (3.27; 5.82)
Platelets (10^9/L)	176 (142; 219)	171 (133.27; 219)	156.5 (129; 190)	160.1 (124; 201.5)	159.5 (124; 197.9)
C-reactive protein (mg/dL)	5.34 (2.10; 11.58)	7.94 (3.60; 13.62)	11.68 (6.40; 16.00)	11.35 (4.18; 15.91)	11.05 (4.79; 15.87)
LDH (U/L)	514.7 (471.0; 594)	533.8 (482.5; 665.0)	584.9 (514.6; 742.7)	534.9 (468.0; 745.2)	558.0 (499.0; 734.4)
SO2 (%)	94.8 (92.8; 96.1)	93.7 (91.7; 95.3)	91.8 (90.0; 94.2)	92.6 (89.6; 94.5)	92.4 (90; 94.5)
CT extension <20%	109 (34.28)	37 (33.9)	8 (7.3)	7 (6.4)	13 (11.9)
20-39%	115 (36.16)	82 (71.3)	21 (18.3)	14 (12.2)	30 (26.1)
40-59%	60 (18.87)	52 (86.7)	20 (33.3)	16 (26.7)	27 (45.0)
≥60%	34 (10.69)	34 (100)	19 (55.9)	21 (61.8)	27 (79.4)
Pectoral muscle area (cm^2)	17 (12; 21)	16 (12; 21)	15 (12; 20)	15 (11; 19)	15 (11; 20)
Pectoral muscle density (HU)	34 (27; 41)	33 (26; 39)	32 (22; 40)	30 (23; 37)	32.5 (23; 39)
L/S ratio	223.5 (159; 292.5)	230 (167; 311)	250.5 (190; 346)	215.5 (160; 291)	246.5 (168; 314)
TAT (cm^2)	34 (23; 47)	38 (27; 51)	46 (33; 57)	45 (30; 58)	43.5 (30; 56)
VAT (cm^2)	152 (102; 210)	152 (108.5; 211.5)	152 (115.5; 220.5)	122 (99; 179)	147.5 (112; 210)
SAT (cm^2)	27 (18; 37)	30.5(21; 42)	35 (26;45)	35 (21; 49)	34 (25; 45)
IMAT (cm^2)	223.5 (159; 292.5)	230 (167; 311)	250.5 (190; 346)	215.5 (160; 291)	246.5 (168; 314)

Patients' pre-existing conditions, along with clinical, laboratory and chest CT variables at ER presentation, including body composition parameters in the population as a whole, in hospitalized patients, in patients who underwent mechanical ventilation, in those who underwent mechanical ventilation or died, and in those who died. Continuous variables are presented as median (IQR); categorical variables are presented as frequencies (%). Column percentages are reported for all patients and row percentages are reported for subpopulations with each different outcome. Calendar period is expressed in weeks since the beginning of the outbreak. Cardiovascular diseases group heart failure, ischemic cardiopathy, and vascular diseases. COPD, chronic obstructive pulmonary disease; LDH, lactate dehydrogenase; SO2, oxygen saturation level; L/S, liver to spleen; TAT, total adipose tissue area; VAT, visceral adipose tissue area; SAT, subcutaneous adipose tissue area; IMAT, intermuscular adipose tissue area.

Body composition parameter selection

Association between CT fat distribution parameters and BMI

Association of CT fat distribution parameters with BMI was estimated only for patients with an available BMI measured within six months previous to ER presentation (n=88). Of the CT parameters describing fat distribution, the strongest association with BMI was for TAT (r=0.706, p<0.001), which we chose over SAT as a measure of general adiposity. TAT was strongly associated with SAT (r=0.959, p <0.001) while the associations between IMAT and VAT and both BMI and TAT were weaker (S1 Table).

Distribution of body composition parameters according to outcome

In a preliminary analysis (S2 Table), we evaluated the association between body composition parameters expressed in quartiles and outcomes, observing a linear relationship of all parameters with hospitalization and MV or death. For death alone, pectoral muscle density and VAT were linearly associated and almost no association was observed for TAT, while the relationship with IMAT was better described by a model including IMAT quartiles. As no association was found with the three outcomes, pectoral muscle area and liver-to-spleen ratio were dropped in subsequent analyses.

Intrareader agreement

Intrareader agreement was excellent for pectoral muscle area and density and for fat compartment areas (Spearman rho between 0.96 and 1.00, p<0.001) and moderate for liver-to-spleen ratio (Spearman rho=0.78, p=0.001).

Associations between body composition parameters and patient outcomes

After correcting for age, sex and calendar period, increased muscle density showed a protective effect on hospitalization (OR for one HU increase =0.967; 95%CI=0.935-1.000), death (OR for one HU increase =0.962; 95%CI=0.922-1.004) and MV or death (OR for one HU increase =0.964; 95%CI=0.934-0.996) (Fig 2). Increased TAT was a risk factor for hospitalization and for MV or death (OR for one cm^2 increase =1.005; 95%CI=1.002-1.008 and OR for one cm^2 increase =1.005; 95%CI=1.002-1.009, respectively), but only a small excess was appreciable for the risk of death (OR for one cm^2 increase =1.002; 95%CI=0.998-1.007).

Increased VAT and IMAT were significantly associated with hospitalization (OR for one cm^2 increase =1.028, 95%CI=1.008-1.049 and OR for one cm^2 increase =1.028, 95%CI=1.006-1.050, respectively) and MV or death (OR for one cm^2 increase =1.026, 95%CI=1.008-1.043 and OR for one cm^2 increase =1.024, 95%CI=1.005-1.043, respectively). Considering VAT and IMAT as risk factors for death alone, the associations were weaker and the excesses were possibly due to random

fluctuations (OR for one cm^2 increase of VAT =1.017, 95%CI=0.997-1.038, OR for the last quartile of IMAT vs. first quartile =1.615, 95%CI=0.431-6.053).

Except for VAT, the effect of body composition parameters on outcomes decreased or disappeared when excluding all comorbidities, cardiovascular diseases only, and diabetes only, but not when excluding previous cancer diagnosis only (S3 Table).

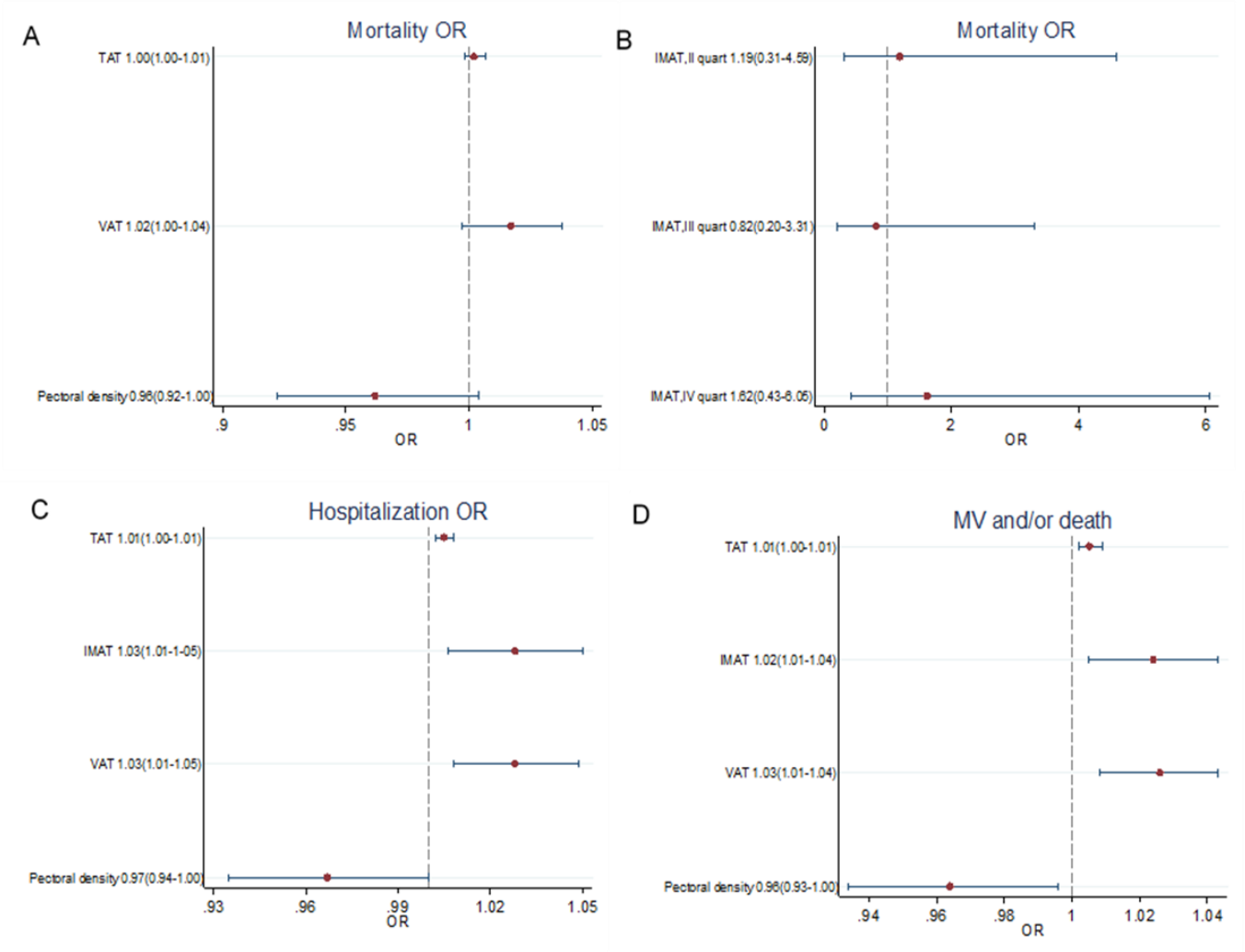


Figure 2. **Multivariate logistic models adjusted for sex, age, and calendar period (weeks since the beginning of the outbreak).** A) Mortality OR for unit increase with 95% CI for unit increase of pectoral muscle density (HU), VAT (cm^2), and TAT (cm^2). B) Mortality OR with 95% CI for IMAT quartiles (cm^2). C) Hospitalization OR with 95% CI for unit increase of pectoral muscle density (HU), VAT (cm^2), IMAT (cm^2), and TAT (cm^2). D) Mechanical ventilation and/or death OR with 95% CI for unit increase of pectoral muscle density (HU), VAT (cm^2), IMAT (cm^2), and TAT (cm^2). OR, Odds Ratio; CI, Confidence Interval; TAT, total adipose tissue area; VAT, visceral adipose tissue area; IMAT, intermuscular adipose tissue area.

Associations between body composition and disease severity at ER presentation

TAT, VAT, and IMAT in our sample were linearly associated with all secondary outcomes (CRP, SO2, and CT disease extension at ER presentation). Instead, pectoral muscle density was linearly associated only with CT disease extension and SO2 but not with CRP (S2 and S3 Figs). Consequently, in multivariate linear regression models corrected for age, sex and calendar period, all body composition parameters were used as continuous variables, with the exception of pectoral muscle density, which was used in quartiles in the model for CRP.

As reported in Table 2, in multivariable models a decreasing pectoral muscle density was linearly associated with increasing lung involvement (increasing CT disease extension and decreasing SO2), while the second and the fourth pectoral muscle density quartiles were inversely associated with CRP as an indicator of systemic inflammation. Increasing TAT, VAT, and IMAT were associated with increasing CT disease extension and decreasing SO2, while the association with CRP was higher for VAT (R squared 0.12 for VAT and 0.09 for TAT).

Table 2. Association of body composition parameters with biomarkers of disease progression at ER presentation.

Variables	CRP		SO2		CT disease extension	
	β	**95% CI**	***β***	**95% CI**	***β***	**95% CI**
Pectoral density (quart1: 3-27]	0					
(quart2: 28-34]	-3.648	-5.760; -1.535				
(quart3: 35-41]	-2.518	-4.733; -.304				
(quart4: 41.1-63]	-4.820	-7.238; -2.403				
Pectoral density[a]			0.058	0.000; 0.116	-0.485	-0.719; -0.252
TAT[a]	0.008	0.000; 0.016	-0.006	-0.011; -0.001	0.046	0 .025; 0.068
VAT[a]	0.064	0.017; 0.111	-0.033	-0.064; -0.003	0.258	0.136; 0.381
IMAT[a]	0.038	-0.016; 0.092	-0.036	-0.071; 0.000	0.245	0.106; 0.384

Multivariate linear regression models adjusted for sex, age and calendar period depicting the associations between pectoral density, TAT, VAT, and IMAT with disease severity at ER presentation described by CRP, SO2, and CT extension. TAT, total adipose tissue area; VAT, visceral adipose tissue area; IMAT, intermuscular adipose tissue area tissue area; ER, Emergency room; CRP, C-reactive protein; SO2, oxygen saturation level.

[a]for unit increase

Relationship between CT body composition parameters and age and mediation analysis

As part of the mediation analysis, to determine whether the effect of age on Covid-19 prognosis was at least partially mediated by a worse body composition, we analyzed the association between CT body composition parameters expressed in quartiles and age. Pectoralis muscle density linearly decreased with age. VAT and IMAT increased with age, while the relationship between TAT and age was more complex, without a clear association (S4 Table).

Consequently, a possible mediation effect on death was evaluated for VAT, IMAT, and pectoralis muscle density, after correcting for sex and calendar period. No mediation effect was found for VAT and IMAT, even if they were associated with both age and death. Instead, the effect of age on death decreased when adding pectoralis muscle density to the model. This analysis suggests that approximately 3% of the effect of age on death was mediated by decreased muscle density (Fig 3).

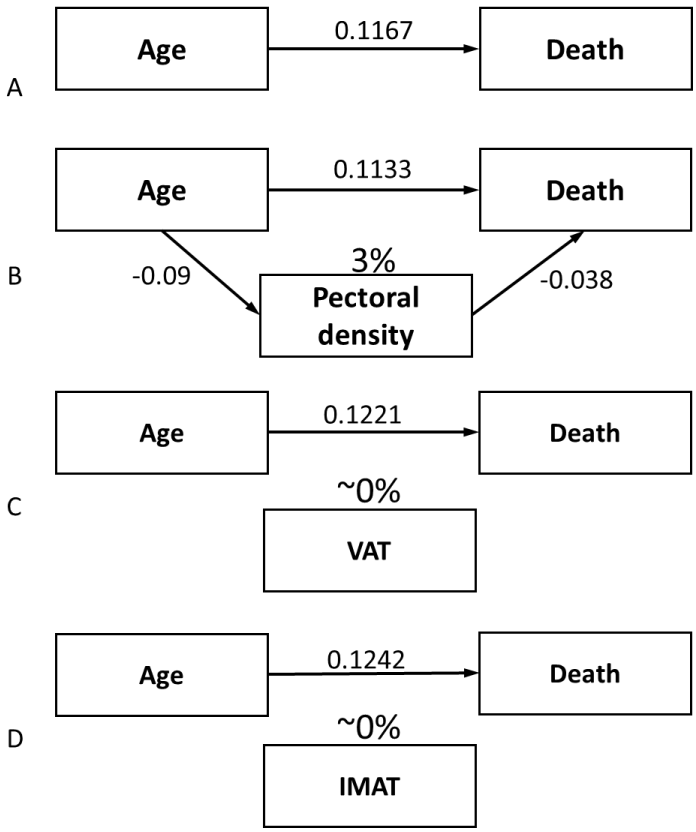


Figure 3. **Mediation analysis.** A) β coefficient of the relationship between age and the logit of death, after correcting for sex and calendar period. B) The coefficient decreases when adding pectoral muscle density to the model, indicating that about 3% of the effect of age on death is mediated by pectoral muscle quality. Vice versa, the coefficient does not decrease when adding VAT (C) or IMAT (D) to the model, suggesting that a mediation effect does not exist for ectopic fat on the relationship between age and death. VAT, visceral adipose tissue area; IMAT, intermuscular adipose tissue area.

DISCUSSION

This observational study showed an association of chest CT measures of fat distribution and muscle quality with a continuum of outcomes representing COVID-19 progression. Chest CT scans routinely performed in symptomatic COVID-19 patients at ER presentation were used to generate body composition measures. Increasing thoracic TAT as a measure of general adiposity as well as VAT and IMAT, representing ectopic fat compartments, were associated with increased risk of hospitalization, the composite of death and MV, and, to a lesser degree, death alone. A higher pectoral density, representing better muscle quality, exhibited a protective effect on the same outcomes. Pectoral muscle area, and liver-to-spleen ratio as a measure of liver steatosis, were not associated with these outcomes.

A few studies have validated thoracic CT to assess ectopic fat areas [38], and pectoral muscle area and density have been used to study the effect of muscle wasting on the outcomes of different diseases [30-32].

Our data are consistent with recently published studies on COVID-19 patients. In a small observational study of 51 patients, a predictive model for hospitalization including VAT and SAT measured on abdominal CT along with clinical variables, performed better than the model that included clinical variables only [39], while, in a study of 165 patients, increasing abdominal VAT was associated with MV or death [23]. In another small cohort of hospitalized patients, increasing upper abdominal VAT on chest CT was associated with higher risk of intensive care admission or MV [19], while in a study of 150 patients who performed chest CT at the ER, upper abdominal VAT was independently associated with the need for intensive care [20]. Higher VAT and lower lumbar skeletal muscle density on abdominal CT of 143 hospitalized COVID-19 patients were independently associated with critical illness [5]. Finally, in a small study of 58 patients, an increasing ratio between waist circumference (as a measure of fat) and paraspinal muscle circumference (as a measure of muscle), measured at T12 level, was associated with a higher probability of MV [40]. In comparison with the present investigation, all cited studies were conducted on smaller cohorts, for the most part including hospitalized patients only, and with a restricted spectrum of intermediate and final outcomes. Furthermore, in some of these studies body composition parameters were measured on patients undergoing abdominal CT scans for specific indications (e.g., abdominal pain), thus in a selected population with specific clinical characteristics [5, 39]. Besides the impact of body composition parameters on COVID-19 progression, we investigated their association with biomarkers of disease severity at ER presentation, trying to distinguish between the two main courses of COVID-19 progression: lung involvement and inflammatory response. Decreasing pectoral muscle density and increasing TAT, VAT, and IMAT were all associated with lung parenchyma involvement

reflected by SO2 and CT extension, while higher VAT showed the strongest association with CRP, a proxy of the systemic inflammatory response.

While the impact of other CT body composition parameters on COVID-19 outcomes was much less evident when excluding patients with comorbidities, the effect of VAT remained substantially similar. These sensitivity analyses suggest that body composition and comorbidities, which are linked in a complex interplay, may be on the same causal sequence in determining COVID-19 outcomes, with the exception of VAT, which may also act through different pathways.

Overall, our results confirm the association between adipose tissue, especially ectopic fat, and the inflammatory state driving disease severity and progression in COVID-19. VAT is known to be an endocrine organ with pro-inflammatory characteristics [21, 13] and different studies have measured higher levels of circulating inflammatory cytokines in people with visceral adiposity compared with lean individuals [13], leading to the hypothesis that they are susceptible to developing a more powerful cytokine storm during COVID-19 progression [41]. Moreover, abdominal obesity can profoundly alter pulmonary function by diminishing exercise capacity and augmenting airway resistance, resulting in increased respiratory fatigue [42]. Also, pectoral muscle density is a measure of respiratory muscle capacity, which is of central importance in COVID-19 patients undergoing MV. In fact, in these patients, death is frequently the consequence of muscle fatigue.

The risk of muscular insufficiency increases with age since sarcopenia is one of the main hallmarks of ageing [15, 43]. Accordingly, pectoral muscle density in our study decreased with age, while VAT and IMAT increased. Even if associated with muscle deterioration and function [30-31], IMAT is still a measure of ectopic fat deposition rather than a measure of the quality of the muscle itself. This association with age, along with the association between these parameters and COVID-19 outcomes, justified our choice to explore the possibility of a mediating effect of body composition on the strong relationship between age and COVID-19 outcome. We found that approximately 3% of the effect of age on death was mediated by decreased pectoral muscle density, while no mediation effect was found for VAT or IMAT, leading to the hypothesis that ectopic fat may belong to another causal pathway than that linking age, muscle quality, and death. This opens up the way for new study hypotheses on the pathogenetic mechanisms in COVID-19 disease progression.

This study has several limitations. Body composition was collected at ER admission, i.e., 2 to 10 days after symptom onset. Therefore, we cannot exclude that body composition was already altered by disease progression, leading to an inversion of the cause-effect interpretation. In fact, patients with more severe forms of COVID-19 may experience loss of muscle mass [44-45]. Nevertheless, the median time from symptom onset and ER visit in our study was 7 days, a very short time to see important changes in CT-measured body composition.

As data on patient height were mostly lacking, it was not possible to calculate the skeletal muscle index, a marker of muscle quantity more reliable than the skeletal muscle area. For this reason, we may argue that pectoral muscle quality seems to be more meaningful than pectoral muscle quantity. However, more studies are needed, especially because a recent study showed that lower pectoral muscle area and index were associated with COVID-19 outcomes [22]. Due to the lack of data on height, BMI was available only in a subset of patients. However, TAT allowed us to have a reliable measure of general adiposity, and ectopic fat depots, particularly VAT, are generally stronger outcome predictors than BMI when studying cardiometabolic risk and associated systemic inflammatory state [46].

We had to exclude 170 COVID-19 patients for whom CT was not available, mostly because chest X-rays and clinical presentation did not suggest pneumonia. Consequently, some potentially eligible patients were missed. This may have occurred among the less severe cases, who were referred to household isolation. This may have introduced a selection bias, especially if we envisage a possible role of body composition and above all obesity as a known risk factor for severe disease, in the decision to perform CT scans. This bias may have led to an underestimation of the association between body composition and patient outcomes.

Finally, the described associations may not be strong enough to be used as prognostic biomarkers in guiding clinical decision making.

Conclusion

In conclusion, we confirmed the association between low muscle quality and ectopic fat accumulation with COVID-19 severity and outcomes. VAT was particularly associated with inflammatory reaction in COVID-19, while all indices including pectoral muscle density were associated with parenchymal involvement. Low muscle quality appears to be one of the mechanisms for the extremely strong effect of age on COVID-19 mortality.

Despite the limits of this observational study, the consistency of results observed on different outcomes and indicators, including disease severity markers and medium-term outcomes, together with the results of previous smaller studies, make a causal relation plausible.

References

1. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020;382: 1708–1720. doi:10.1056/NEJMoa2002032

2. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA. 2020;323: 1775–1776. doi:10.1001/jama.2020.4683

3. WHO Coronavirus Disease (COVID-19) Dashboard. [cited 15 Feb 2021]. Available: https://covid19.who.int

4. Poston JT, Patel BK, Davis AM. Management of Critically Ill Adults With COVID-19. JAMA. 2020;323: 1839–1841. doi:10.1001/jama.2020.4914

5. Yang Y, Ding L, Zou X, Shen Y, Hu D, Hu X, et al. Visceral Adiposity and High Intramuscular Fat Deposition Independently Predict Critical Illness in Patients with Sars-COV-2. Obesity. n/a. doi:10.1002/oby.22971

6. Bhatraju PK, Ghassemieh BJ, Nichols M, Kim R, Jerome KR, Nalla AK, et al. Covid-19 in Critically Ill Patients in the Seattle Region - Case Series. N Engl J Med. 2020;382: 2012–2022. doi:10.1056/NEJMoa2004500

7. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 2020;8: 475–481. doi:10.1016/S2213-2600(20)30079-5

8. Watanabe M, Risi R, Tuccinardi D, Baquero CJ, Manfrini S, Gnessi L. Obesity and SARS-CoV-2: a population to safeguard. Diabetes Metab Res Rev. 2020; e3325. doi:10.1002/dmrr.3325

9. Lighter J, Phillips M, Hochman S, Sterling S, Johnson D, Francois F, et al. Obesity in patients younger than 60 years is a risk factor for Covid-19 hospital admission. Clin Infect Dis Off Publ Infect Dis Soc Am. 2020. doi:10.1093/cid/ciaa415

10. Valerio A, Nisoli E, Rossi AP, Pellegrini M, Todesco T, Ghoch ME. Obesity and Higher Risk for Severe Complications of Covid-19: What to do when the two pandemics meet. J Popul Ther Clin Pharmacol J Ther Popul Pharmacol Clin. 2020;27: e31–e36. doi:10.15586/jptcp.v27iSP1.708

11. Fezeu L, Julia C, Henegar A, Bitu J, Hu FB, Grobbee DE, et al. Obesity is associated with higher risk of intensive care unit admission and death in influenza A (H1N1) patients: a systematic review and meta-analysis. Obes Rev. 2011;12: 653–659. doi:10.1111/j.1467-789X.2011.00864.x

12. Simonnet A, Chetboun M, Poissy J, Raverdy V, Noulette J, Duhamel A, et al. High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. Obesity. n/a. doi:10.1002/oby.22831

13. Chait A, den Hartigh LJ. Adipose Tissue Distribution, Inflammation and Its Metabolic Consequences, Including Diabetes and Cardiovascular Disease. Front Cardiovasc Med. 2020;7. doi:10.3389/fcvm.2020.00022

14. Zoico E, Corzato F, Bambace C, Rossi AP, Micciolo R, Cinti S, et al. Myosteatosis and Myofibrosis: relationship with aging, inflammation and insulin resistance. Arch Gerontol Geriatr. 2013;57: 411–416. doi:10.1016/j.archger.2013.06.001

15. Correa-de-Araujo R, Addison O, Miljkovic I, Goodpaster BH, Bergman BC, Clark RV, et al. Myosteatosis in the Context of Skeletal Muscle Function Deficit: An Interdisciplinary Workshop at the National Institute on Aging. Front Physiol. 2020;11. doi:10.3389/fphys.2020.00963

16. Zamboni M, Gattazzo S, Rossi AP. Myosteatosis: a relevant, yet poorly explored element of sarcopenia. Eur Geriatr Med. 2019;10: 5–6. doi:10.1007/s41999-018-0134-3

17. Han A, Bokshan SL, Marcaccio SE, DePasse JM, Daniels AH. Diagnostic Criteria and Clinical Outcomes in Sarcopenia Research: A Literature Review. J Clin Med. 2018;7. doi:10.3390/jcm7040070

18. Mourtzakis M, Prado CMM, Lieffers JR, Reiman T, McCargar LJ, Baracos VE. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. Appl Physiol Nutr Metab Physiol Appl Nutr Metab. 2008;33: 997–1006. doi:10.1139/H08-075

19. Petersen A, Bressem K, Albrecht J, Thieß H-M, Vahldiek J, Hamm B, et al. The role of visceral adiposity in the severity of COVID-19: Highlights from a unicenter cross-sectional pilot study in Germany. Metabolism. 2020;110: 154317. doi:10.1016/j.metabol.2020.154317

20. Watanabe M, Caruso D, Tuccinardi D, Risi R, Zerunian M, Polici M, et al. Visceral fat shows the strongest association with the need of intensive care in patients with COVID-19. Metabolism. 2020;111: 154319. doi:10.1016/j.metabol.2020.154319

21. Ghazarian M, Luck H, Revelo XS, Winer S, Winer DA. Immunopathology of adipose tissue during metabolic syndrome. Turk Patoloji Derg. 2015;31: 172–180. doi:10.5146/tjpath.2015.01323

22. Ufuk F, Demirci M, Sagtas E, Akbudak IH, Ugurlu E, Sari T. The prognostic value of pneumonia severity score and pectoralis muscle Area on chest CT in adult COVID-19 patients. Eur J Radiol. 2020;131: 109271. doi:10.1016/j.ejrad.2020.109271

23. Favre G, Legueult K, Pradier C, Raffaelli C, Ichai C, Iannelli A, et al. Visceral fat is associated to the severity of COVID-19. Metab - Clin Exp. 2021;115. doi:10.1016/j.metabol.2020.154440

24. Iacobellis G, Malavazos AE, Ferreira T. COVID-19 rise in Younger adults with Obesity: Visceral Adiposity can predict the Risk. Obes Silver Spring Md. 2020 [cited 23 Dec 2020]. doi:10.1002/oby.22951

25. Földi M, Farkas N, Kiss S, Dembrovszky F, Szakács Z, Balaskó M, et al. Visceral adiposity elevates the risk of critical condition in COVID-19: A systematic review and meta-analysis. Obes Silver Spring Md. 2020. doi:10.1002/oby.23096

26. Besutti G, Giorgi Rossi P, Iotti V, Spaggiari L, Bonacini R, Nitrosi A, et al. Accuracy of CT in a cohort of symptomatic patients with suspected COVID-19 pneumonia during the outbreak peak in Italy. Eur Radiol. 2020; 1–10. doi:10.1007/s00330-020-07050-x

27. Tong Y, Udupa JK, Torigian DA, Odhner D, Wu C, Pednekar G, et al. Chest Fat Quantification via CT Based on Standardized Anatomy Space in Adult Lung Transplant Candidates. PLOS ONE. 2017;12: e0168932. doi:10.1371/journal.pone.0168932

28. Kim YS, Kim EY, Kang SM, Ahn HK, Kim HS. Single cross-sectional area of pectoralis muscle by computed tomography - correlation with bioelectrical impedance based skeletal muscle mass in healthy subjects. Clin Physiol Funct Imaging. 2017;37: 507–511. doi:10.1111/cpf.12333

29. Maurovich-Horvat P, Kallianos K, Engel L-C, Szymonifka J, Schlett CL, Koenig W, et al. Relationship of Thoracic Adipose Tissue Depots with Coronary Atherosclerosis and Circulating Inflammatory Biomarkers. Obes Silver Spring Md. 2015;23: 1178–1184. doi:10.1002/oby.21080

30. McDonald M-LN, Diaz AA, Ross JC, San Jose Estepar R, Zhou L, Regan EA, et al. Quantitative Computed Tomography Measures of Pectoralis Muscle Area and Disease Severity in Chronic Obstructive Pulmonary Disease. A Cross-Sectional Study. Ann Am Thorac Soc. 2014;11: 326–334. doi:10.1513/AnnalsATS.201307-229OC

31. Bak SH, Kwon SO, Han S-S, Kim WJ. Computed tomography-derived area and density of pectoralis muscle associated disease severity and longitudinal changes in chronic obstructive pulmonary disease: a case control study. Respir Res. 2019;20: 226. doi:10.1186/s12931-019-1191-y

32. Looijaard WGPM, Dekker IM, Stapel SN, Girbes ARJ, Twisk JWR, Oudemans-van Straaten HM, et al. Skeletal muscle quality as assessed by CT-derived skeletal muscle density is associated with 6-month mortality in mechanically ventilated critically ill patients. Crit Care. 2016;20. doi:10.1186/s13054-016-1563-3

33. Rossi PG, Ferroni E, Alegiani SS, Leoni O, Pitter G, Cereda D, et al. Survival of hospitalized COVID-19 patients in Northern Italy: a population-based cohort study by the ITA-COVID19 Network. medRxiv. 2020; 2020.05.15.20103119. doi:10.1101/2020.05.15.20103119

34. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40: 373–383. doi:10.1016/0021-9681(87)90171-8

35. Ballotari P, Ranieri SC, Vicentini M, Caroli S, Gardini A, Rodolfi R, et al. Building a population-based diabetes register: An Italian experience. Diabetes Res Clin Pract. 2014;103: 79–87. doi:10.1016/j.diabres.2013.11.020

36. Teigen LM, John R, Kuchnia AJ, Nagel EM, Earthman CP, Kealhofer J, et al. Preoperative Pectoralis Muscle Quantity and Attenuation by Computed Tomography Are Novel and Powerful Predictors of Mortality After Left Ventricular Assist Device Implantation. Circ Heart Fail. 2017;10. doi:10.1161/CIRCHEARTFAILURE.117.004069

37. Zhang Z, Zheng C, Kim C, Van Poucke S, Lin S, Lan P. Causal mediation analysis in the context of clinical research. Ann Transl Med. 2016;4. doi:10.21037/atm.2016.11.11

38. Diaz AA, Martinez CH, Harmouche R, Young TP, McDonald M-L, Ross JC, et al. Pectoralis muscle area and mortality in smokers without airflow obstruction. Respir Res. 2018;19: 62. doi:10.1186/s12931-018-0771-6

39. Chandarana H, Dane B, Mikheev A, Taffel MT, Feng Y, Rusinek H. Visceral adipose tissue in patients with COVID-19: risk stratification for severity. Abdom Radiol. 2020 [cited 29 Aug 2020]. doi:10.1007/s00261-020-02693-2

40. Kottlors J, Zopfs D, Fervers P, Bremm J, Abdullayev N, Maintz D, et al. Body composition on low dose chest CT is a significant predictor of poor clinical outcome in COVID-19 disease - A multicenter feasibility study. Eur J Radiol. 2020;132: 109274. doi:10.1016/j.ejrad.2020.109274

41. Finelli C. Obesity, COVID-19 and immunotherapy: the complex relationship! Immunotherapy. 2020 [cited 23 Sep 2020]. doi:10.2217/imt-2020-0178

42. Dietz W, Santos-Burgoa C. Obesity and its Implications for COVID-19 Mortality. Obesity. 2020;28: 1005–1005. doi:10.1002/oby.22818

43. Walston JD. Sarcopenia in older adults. Curr Opin Rheumatol. 2012;24: 623–627. doi:10.1097/BOR.0b013e328358d59b

44. Gualtieri P, Falcone C, Romano L, Macheda S, Correale P, Arciello P, et al. Body Composition Findings by Computed Tomography in SARS-CoV-2 Patients: Increased Risk of Muscle Wasting in Obesity. Int J Mol Sci. 2020;21. doi:10.3390/ijms21134670

45. Hosey MM, Needham DM. Survivorship after COVID-19 ICU stay. Nat Rev Dis Primer. 2020;6: 1–2. doi:10.1038/s41572-020-0201-1

46. Lumeng CN, Saltiel AR. Inflammatory links between obesity and metabolic disease. J Clin Invest. 2011;121: 2111–2117. doi:10.1172/JCI57132

7. CONCLUSIONS

The increasing prevalence of risk factors and disease burden associated with body composition, and the ability of imaging techniques to provide information on fat quantity and distribution as well as lean mass quantity and quality, result in a high research interest in imaging biomarkers of body composition. On one hand, these biomarkers may help in the diagnosis and staging of metabolic conditions, above all NAFLD (as reviewed in Chapter 5.1), and on the other, they may be used as risk and prognostic factor in a vast range of diseases, even not directly related to metabolic dysfunction, such as infectious diseases and cancer (as reviewed in Chapter 5.2 and 6). In these diseases, they can be a complement to classical prognostic factors, or they can give new insights into disease pathogenesis.

As observed in the systematic review exposed in Chapter 5.1.1, main guidelines on NAFLD recommend a case-finding approach of patients at high risk for advance disease among subjects with metabolic risk factors, even if the rationale is weak given the absence of available treatments for pre-clinical disease. Proposed algorithms to stratify patients are rarely implemented into clinical practice, since they generate too high referral rates to the hepatologist and to liver biopsy, especially in high risk patients' groups, such as HIV-infected patients or patients with diabetes (Chapter 5.1.3 and Chapter 5.1.4). Moreover, adherence to referral is low, as reported in the pilot study described in Chapter 5.1.4, at least partly because NAFLD is frequently overlooked both by primary care physicians and by citizens. In this setting, imaging biomarkers may be useful in two ways. First, largely available tests such as US with elastographic tools can be implemented into the screening algorithm, especially if combined with serum or molecular biomarkers, to allow a better selection of patients and to reduce referral rates. Second, imaging tests may be useful to avoid liver biopsy if able to correctly diagnose advanced disease. In both cases, the goal to select and diagnose patients with significant fibrosis is within reach, but if the goal is broadened also to NASH, as actually recommended by main guidelines, imaging biomarkers are in an early stage of development, lacking robust validation and harmonization of positivity thresholds (Chapter 5.1.2).

When body composition biomarkers are proposed as prognostic factors, they are generally obtained from imaging examinations performed for other reasons. This is an opportunity, because it means that we can obtain information without subjecting patients to adjunctive tests (frequently involving ionizing radiations). On the other hand, some operative and interpretative weaknesses may actually derive from the use of tests performed for other reasons. First, it can generate limitations in study design that can result in biases and comparability issues leading to inappropriate interpretation of results, as for example when we consider steatosis as a possible risk factor for liver metastasis

occurrence (Chapter 5.2.1 and 5.2.2). Second, it results in a delay in the harmonization and standardization of techniques as well as validation of biomarkers and their positivity thresholds. Not always we can use the most accurate biomarker, as for instance when we evaluate liver steatosis on CT because MRI is not available (Chapter 5.2.2), or when we evaluate fat distribution and muscle quantity and quality as prognostic factors in diseases that do not imply the execution of abdominal but only chest CT scan, for instance in COVID-19 (Chapter 6.2). Nonetheless, some hints on the role of imaging biomarkers of body fat distribution and lean mass quality (probably more than quantity) in different kinds of diseases may be gathered from our observational studies (Chapter 6.1 and 6.2).

The global picture that derives from our observations is that imaging biomarkers of body composition may have a diagnostic and prognostic value, but there is a lack of large multicenter studies with predefined procedures and thresholds prospectively testing their clinical utility. In the absence of this foreground evidence, the scientific community should probably abstain from translating results of suggestive observational studies in real changes of clinical practice.

8. APPENDICES

Appendix A

3.2.2. Accuracy of Imaging Methods for Steatohepatitis Diagnosis in Non-alcoholic Fatty Liver Disease Patients: A Systematic Review

SUPPLEMENTARY MATERIAL

Supplementary Material

Table of contents

1. Supplementary Methods section

Search algorithm for MEDLINE:

Vibration controlled transient elastography OR imaging OR "Magnetic Resonance Imaging, Interventional"[Mesh] OR "Magnetic Resonance Spectroscopy"[Mesh] OR "Ultrasonography"[Mesh] OR MRI AND Non Alcoholic Steatohepatitis OR NASH. Limits: ti/ab; humans

Test-related consequences:

As no systematic review was conducted for test-related consequences, they were considered only for those techniques with contrast media administration or radiation exposure.

2. Supplementary Results section:

Test-related consequences: No study mentioned test-related outcomes. Nevertheless, contrast-enhanced US, gadoxetic acid-enhanced MR and SPIO/USPIO-enhanced MR need contrast media administration, while CT and scintigraphy use significant doses of ionizing radiation.

Accuracy Results Synthesis

Elastography techniques

Four studies explored the accuracy of transient elastography (TE) in distinguishing between simple steatosis and NASH, including 47,[26] 127,[18] 183,[27] and 104[5] patients, respectively. Their results were not directly comparable since the various authors chose different pathological definitions of NASH. However, AUROC ranged from 0.35 (0.22-0.49)[5] to 0.82 (0.70-0.94).[26] Three studies also reported TE accuracy in diagnosing definite NASH (NAS≥5), with AUROCs from 0.65 (0.54-0.77) to 0.75 (0.68-0.82).[18,26,27] Only one study identified a liver stiffness (LS) optimal cut-off value (>7 KPa) for the differentiation between simple steatosis and NASH (NAS≥5), resulting in 86% sensitivity and 58% specificity.[27] Also, Eddowes et al. analysed different LS cut-off values, not for NASH diagnosis but for the identification of high-risk patients (NASH or fibrosis>1 based on the system described by Kleiner et al.), with sensitivities ranging from 62% (9 KPa) to 89% (5.8 KPa) and specificities from 30% (5.8 KPa) to 90% (9 KPa).[26]

Accuracy of acoustic radiation force impulse (ARFI) for NASH diagnosis was evaluated in two studies, on 64[28] and 32[19] patients, respectively. For the purpose of evaluating the accuracy in the differentiation between simple steatosis and NASH, both studies had a strong limitation: borderline patients were excluded from the study by Fierbinteanu Braticevici et al.,[28] while Guzman-Aroca et al. grouped together inflammation and fibrosis for the analysis.[19] The AUROCs were similar (0.87 and 0.90, respectively). In the first study a cut-off >1.1 m/s produced a sensitivity of 77% and a specificity of 72%, while in the second the optimal cut-off was >1.3 m/s, with 85% sensitivity and 83% specificity. Only 2/32 patients in the study by Guzman-Aroca et al. had ARFI velocity between 1.1 and 1.3 m/s, so it is conceivable that no significant changes in accuracy results would occur by applying the first cut-off also in the second study.

As opposed to these studies, others failed to find a significant association between TE or shear wave-based liver stiffness or velocity and the histological features of NASH (inflammation and/or ballooning).[s1,s2,s3,s4,s5]

Six studies reported on the accuracy of MRE in the differentiation between simple steatosis and NASH. Again, histopathological definitions of NASH were not homogeneous. AUROCs ranged from 0.70 to 0.76 in studies considering borderline and definite NASH as a single group in 117,[30] 100,[31] and 104[6] patients, respectively, with similar values for 2D and 3D MRE.[31] Loomba et al. identified different optimal cut-off values based on the technique used and reported a sensitivity of 42% and a specificity of 92% with a cut-off >3.26 KPa for 2D MRE. Slightly higher AUROCs were found in two studies defining NASH as NAS≥5 (0.77 - 0.79) in 49[29] and 142[18] patients, with sensitivity and specificity of 72% and 87% using an optimal cut-off value of >3.24 KPa.[29] These results were similar to those obtained in the same study by Costa-Silva et al. in a subset of patients

(n=21) without any degree of fibrosis (AUROC=0.78, cut-off >3.22 KPa, sensitivity 69%, specificity 87%). The highest AUROC (0.93) was reported by Chen et al. in a study including 58 patients, with sensitivity/specificity of 94%/73% for a cut-off >2.74 KPa and 83%/82% for a cut-off >2.90 KPa.[23] However, they grouped together patients with inflammation without fibrosis and NAFLD patients with fibrosis, the latter being the majority of NASH patients (29/36). Of note, in a recent study by Dillman et al. no association was found between MRE and histopathological features of NASH in a population of 69 children and young adults up to 21 years old.[41] Finally, some studies found that MRE liver stiffness was associated with inflammation independently from fibrosis,[29] while others found that the association decreased after adjusting for fibrosis in multivariable analysis, both for TE and for MRE.[18,s4,s6,s7]

US non-elastographic techniques

The US fatty score (FS) evaluated by Liang et al. in 101 bariatric patients showed an AUROC of 0.79, with sensitivity 81% and specificity 66% (optimal cut-off=7). In the same study the authors also introduced a modified fatty score (US-MFS) based on the total FS with an additional supplementary role of parenchymal echogenicity and gallbladder wall blurring. This latter score presented an AUROC of 0.82, with a cut-off of 3, producing sensitivity and specificity for NASH diagnosis of 72% and 86%, respectively. It should be mentioned that in this study NASH was defined as the presence of fibrosis (≥1) or acinar zone 3 hepatocellular injury with ballooning degeneration (≥2).[20] In 2012, Ballestri et al. assessed the accuracy of the US-fatty liver indicator (FLI) in the diagnosis of NASH and severe NASH (NAS≥5) in 53 NAFLD patients, obtaining AUROCs of 0.76 and 0.80, respectively. With a cut-off of 4, US-FLI presented 100% sensitivity and 46% specificity for the diagnosis of severe NASH.[32] The score US-fatty liver by Petrick et al. demonstrated, in a cohort of 513 patients undergoing bariatric surgery, a sensitivity of 89% and a specificity of 45% for the diagnosis of steatohepatitis (lobular inflammation) when considering mild or higher degrees of US-fatty liver.[22]

A different score comprehending also splenic diameter was tested by Zardi and coll. in 2011 on 94 patients, obtaining a sensitivity of 74% and a specificity of 66% (cut-off 5) for differentiating between simple steatosis and NASH; if only echo attenuation and focal fat sparing were considered (cut-off 1), the sensitivity and specificity increased to 75% and 92%.[35] Splenic longitudinal diameter (SLD) per se has also been tested for the diagnosis of NASH, obtaining an AUROC of 0.920, with sensitivity 88% and specificity 95% using a SLD cut-off of 116 mm in 83 patients.[34]

The last morphological US index which was tested for NASH diagnosis is US perihepatic adipose tissue thickness (PATT). Lirussi et al. showed that in 33 biopsy-proven NAFLD patients US-PATT

sensitivity and specificity, with a cut-off value of 11.8 mm, were 100% and 50% for differentiating between simple steatosis and NASH (including borderline patients), and 80% and 50% to predict necro-inflammatory activity grading.[33]

Only one study evaluated the accuracy of contrast-enhanced US (CEUS) in NASH diagnosis (66 patients), showing sensitivity and specificity of 100% and 95%, and both of 100%, respectively, for a signal intensity at 5 minutes lower than 137.8 and for a signal intensity at 20 minutes lower than 43.6. This study was limited by partial verification since liver biopsy was performed only in 31/66 NAFLD patients, while in the remaining 35/66 patients NASH was ruled out through clinical criteria.[36]

Finally, some Doppler US indices have been evaluated in NAFLD, without finding any association between them and histological features of NASH.[34,s8]

MR non-elastographic techniques

In 2014, Abrigo et al. reported in a cohort of 132 patients that the 31P-MRS-derived ratio between nucleotide triphosphate (α peak) levels and triphosphate levels (αNTP/TP) differentiated NASH from simple ateatosis, with an AUROC of 0.71 and sensitivity/specificity ranging from 28%/91% and 91%/16% for cut-off values of ≤10.57% and ≤16.36%, respectively.[21] By using a long-echo time 1H-MRS on 26 patients, Kim et al. in 2017 found that alanine concentration had an AUROC of 1.00 for the diagnosis of definite NASH (NAS≥5), with a 100% sensitivity and specificity for a cut-off >16.04%, whereas lactate + triglyceride concentration with a cut-off >360.8% demonstrated a sensitivity of 82% and a specificity of 67% (AUROC 0.78).[37] Other 31P-MRS results found to be associated with NASH histopathological features are adenosine triphosphate (ATP),[s9] phosphomonoester (PME)/total phosphorus (TP) ratio, nicotine adenine dinucleotide phosphate (NADPH)/(PME+ phosphodiester, PDE) ratio, NADPH/γ-ATP, NADPH/TP, inorganic phosphate (Pi), k values (chemical exchange rate constant of the Pi-to-ATP reaction) and unidirectional forward exchange flux (FATP),[s10,s11] all of these tested in a relatively small sample of patients (up to 30 patients) and none with available accuracy data.

Among MRI techniques without contrast media administration, quantitative susceptibility imaging proposed by Leporq et al. in a sample of 32 patients had the highest AUROC (0.91) for NASH diagnosis.[39] Multiparametric MRI (mpMRI) in two studies on 50 and 71 patients demonstrated AUROCs of 0.69, 0.74 and 0.80 in the differentiation between NASH and simple steatosis when considering corrected T1 (cT1) as index test, in the differentiation between NAS<5 and ≥5 for the same index test and finally in the diagnosis of NASH by using Liver Inflammation and Fibrosis

(LIF) score, respectively.[26,38] An optimal cut-off for LIF was identified (1.4), with sensitivity 91% and specificity 52%. In the same studies, the authors also explored the accuracy of mpMRI to distinguish between low- and high-risk (NASH or fibrosis>1) patients, identifying AUROCs of 0.73 for cT1 (with sensitivity/specificity of 97%/50% for a cut-off value of 875 ms) and 0.89 for LIF (with sensitivity/specificity of 90%/77% for a cut-off value of 1.4). Even if in slightly different samples (3/50 patients were excluded from TE accuracy analysis due to unreliable measurement), the AUROCs reported by Eddowes et al. for cT1 were similar or lower than those relative to TE liver stiffness. However, negative predictive values, indicating those patients for whom biopsy could potentially be avoided, were substantially higher for cT1 (80%-83%) compared to TE (39%-43%). A single study on 66 type 2 diabetes patients evaluated the diagnostic accuracy of intravoxel incoherent motion (IVIM) diffusion-weighted MRI in the differentiation between simple steatosis and NASH, showing AUROCs of 0.74, 0.68 and 0.61, respectively, for D, D* and f fractions. With different cut-off values, the three IVIM parameters presented sensitivities and specificities ranging from 49% to 69% and from 66% to 71%. respectively.[40] Conversely, two studies failed to demonstrate any independent association between DWI parameters and NASH features in 64 paediatric patients and 89 adults.[s12,s13] However, these two studies used only 3 b-values as opposed to 10 in the study by Parente et al.[40] Two included studies assessed the accuracy of morphological MRI findings, namely liver volume and preperitoneal fat area, demonstrating an AUROC of 0.74 for liver volume in 69 children and young adults,[41] and sensitivity/specificity of 93%/55% for a preperitoneal fat area >5 cm^2 in 59 diabetes patients.[42] In 2016, Gallego-Duran et al. obtained through MRI optical analysis a large set of "estimators" and the optimal combination of parameters predicting NASH (NASHMRI score), which produced, in a validation cohort of 87 patients, an AUROC of 0.83 with sensitivity of 87% and specificity of 60% (cut-off>0.5).[24] Finally, liver fat fraction, a known biomarker of steatosis that can be estimated more accurately by means of 1H-MRS or multi-echo MRI (both ensuring the measurement of proton density fat fraction), was evaluated as a potential indicator of necro-inflammatory activity, yielding inconsistent results in different studies mostly conducted in small cohorts of patients. [41,39,18,s10,s11,s14,s15]

Contrast-enhanced MRI techniques were used for NASH diagnosis in three studies. The relative gadoxetic acid enhancement in hepatobiliary phase presented a sensitivity of 97% and a specificity of 63% with a cut-off ≤1.24 (AUROC=0.85) in a retrospective study on 81 patients.[43] In 2008, Tomita et al. showed that SPIO-enhanced MRI in a small sample of 19 patients had AUROCs of 0.79 and 0.83, respectively, for T (time constant) and %T2 (relative decrease in T2), the first with 67% specificity and 100% sensitivity (cut-off 42.8) and the second with 73% sensitivity and 87% specificity (cut-off 32.5), both for the diagnosis of definite NASH (NAS≥5).[45] Finally, USPIO-

enhanced MRI-derived ΔR2* (difference in R2* between contrast-enhanced and baseline imaging) had an AUROC of 0.87 with sensitivity/specificity of 77%/85% and 91%/73% with different cut-off values, again to differentiate between NAS<5 and ≥5 in a study including 25 NAFLD patients.[44]

Other techniques

Few reports are available on associations between CT features and NASH,[s16,s17] only one of which including accuracy as an outcome; Naganawa and coll. in 2018 evaluated two predictive models for NASH diagnosis (NAS≥3) obtained from combinations of non-enhanced CT texture features, one for patients without and one for patients with high suspicion of fibrosis. In the validation dataset (n=35 patients), the first predictive model had an AUROC of 0.94, with sensitivity of 100% and specificity of 92% (cut-off 0.45), while the second produced an AUROC of 0.60, a sensitivity of 100% and a specificity of 31% (cut-off 0.81).[25]

The accuracy of TC99m-phytate colloid scintigraphy was reported in a single study on 37 NAFLD patients, which reported an AUROC of 0.82 for the liver-to-spleen (L/S) uptake ratio in the diagnosis of definite NASH, with sensitivity/specificity of 100%/75% by using a cut-off value of 2.93.[46] The same technique had previously been evaluated, with no association between scintigraphy results and inflammation found.[s18] On the other hand, another kind of liver scintigraphy, with 99mTc-MIBI as radiotracer, showed significant reduction of liver-to-heart (L/H) ratio and intrahepatic uptake in NASH vs simple steatosis patients.[s19]

3. Supplementary Tables

Excluded studies	**Reasons for exclusion**
Al-Busafi 2012	No outcomes considered. It explores associations of index test with steatosis but not with necroinflammatory activity or NASH
Allen 2017	Reported as abstract only
Ballestri 2015	Comment on a study from Bril et al. in which the authors introduce a re-analysis of their previously published data (Ballestri 2012). These few re-analyses have been included in the data extraction sheet of Ballestri 2012.
Ballestri 2015	Review
Cocciolillo 2013	Reported as abstract only
Cocciolillo 2014	No reference standard (no liver biopsy)

Cui 2016	No outcomes considered. It explores accuracy for fibrosis assessment and associations of index test with fibrosis but not with necroinflammatory activity or NASH
Dowman 2011	Systematic review. No metanalysis. A cross-check of the references has been carried out.
Doycheva 2016	No reference standard (no liver biopsy)
Ebinuma 2011	No outcomes considered. Necroinflammatory activity is considered only for its potential influence on the association between index test and fibrosis.
Eddowes 2016	Reported as abstract only
Festi 2013	Systematic review mainly about NAFLD, steatosis and fibrosis diagnosis. No metanalysis. A cross-check of the references has been carried out.
Forsgren 2014	No reference standard (no liver biopsy); no sub-analysis for different diffuse liver diseases.
Gameel 2017	Reported as abstract only
Garcia-Monzon 2015	Reported as abstract only
García-Monzón 2015	Included population is not with or at high risk for NAFLD (patients with gallstones).
Harrison 2016	Reported as abstract only
Kim 2016	Reported as abstract only
Kim 2016	Reported as abstract only
Kobayashi 2009	No outcomes considered. It explores associations of index test with steatosis/fibrosis but not with necroinflammatory activity or NASH
Koreeda 2010	Not found, not possible to contact authors
Kwok 2014	Metanalysis only about fibrosis or clinical-serum markers of NASH. A cross-check of the references has been carried out.
Lin 2015	Reported as abstract only
Lin 2017	Among histopathological parameters only steatosis and fibrosis are considered (no NASH diagnosis nor ballooning or inflammation)

Maric 2015	Reported as abstract only
Neuman 2014	Systematic review about clinical and serum biomarkers of NASH. No imaging technique considered.
Onishi 2006	No outcomes considered.
Paredes 2016	Reported as abstract only
Permutt 2012	No outcomes considered. It explores associations of index test with steatosis/fibrosis but not with necroinflammatory activity or NASH
Razmpour 2017	Systematic review of methods for the diagnosis of NAFLD in children. With respect to imaging techniques, only methods for steatosis and fibrosis diagnosis are mentioned.
Roberts 2016	Reported as abstract only
Sharma 2009	No reference standard (no liver biopsy)
Skamarauskas 2014	Study about NASH is only preclinical (animal model), included humans are 3 healthy volunteers for a feasibility study.
Smits 2014	Reported as abstract only
Tovo 2015	Review
Traussnigg 2013	Reported as abstract only
Yilmaz 2013	Probably systematic review; no answer from the authors.

Supplementary Table 1: Excluded studies and respective reasons for exclusion.

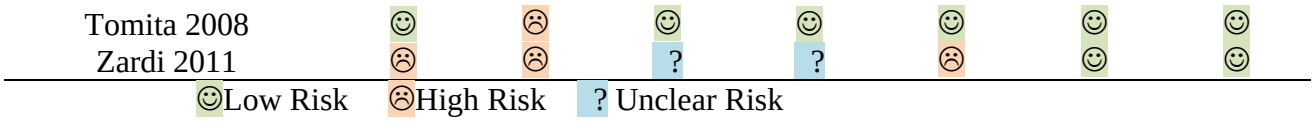
Index Tests and Study Outcomes

Index test	Studies evaluating associations between imaging and pathology	Studies evaluating accuracy in terms of AUROC	Studies identifying optimal cut-off values
Elastography techniques			
Transient Elastography (TE)	Attia2016 Cassinotto2016 Lee2017 Lupsor2010 Yilmaz2014 Yoneda2008 Yoneda2010	Imajo2016 † Eddowes2018	Lee2016 Park2017
Shear wave-based Elastography	Attia2016 Cassinotto2016 Lee2017 Osaki2010 Palmeri2011 Praveenraj2017 Yoneda2010		Fierbinteanu Braticevici2013 Guzman-Aroca2012
Magnetic Resonance Elastography (MRE)	Dillman2018	Imajo2016 ‡	Chen2011 Costa-Silva2018 Loomba2014 Loomba2016 Park2017
Magnetic Resonance techniques (other than MRE)			
MR Fat Fraction (by means of MRI or ^{1}H-MRS)	Chen2011 Cortez-Pinto1999 Dillman2018 Kalra2009 Leporq2017 Saadeh2002 Sevastianova2010 Traussnig2017 Valkovic2014	Imajo2016 §	
^{1}H-MRS and/or ^{31P}-MRS metabolites	Cortez-Pinto1999 Sevastianova 2010 Traussnig2017 Valkovic2014		Kim2017 Abrigo2014
Diffusion-Weighted (DW) and IntraVoxel Incoherent Motion (IVIM) MR imaging	Manning2017 Murphy2015		Parente2015
Multiparametric MRI: corrected T1 (cT1), Liver Inflammation Fibrosis (LIF) score		Eddowes2018	Pavlides2017
Gadoxetic Acid-enhanced MRI	Feier2013 Wu2013		Bastati2014

SPIO/USPIO-enhanced MRI	Asanuma2010 Tonan2012		Smits2015 Tomita2008
MRI liver volume		Dillman2018	
MRI early patchy enhancement	Elias2009		
MRI optical analysis			Gallego-Duran2016
Quantitative susceptibility MRI		Leporq2017	
MRI preperitoneal fat area			Parente2018
Ultrasonographic techniques (non-elastographic methods)			
US hyperechogenicity and/or US scores	Ataseven2005 Bril2015 Saadeh2002 Ballestri2017		Ballestri2012 Petrick2015 Liang2007 Zardi2011
Doppler US indices	Goncalves2016 Tarantino 2009		
Contrast-enhanced US (CEUS)	Moriyasu2005		Iijima2007
US Perihepatic Adipose Tissue Thickness (PATT)			Lirussi2009
US splenic longitudinal diameter			Tarantino2009
Other techniques			
CT liver-to-spleen ratio and other CT features	Ataseven2005 Saadeh2002 Oliva2006		
CT texture analysis			Naganawa2018
Scintigraphy	Duman2006 Masuda2012		Kikuchi2009

Supplementary Table 2: Included studies grouped based on index test technique and on type of outcome considered. † in combination with CAP; ‡ in combination with PDFF; § in combination with MRE.

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Abrigo 2014	?	☹	☺	☺	☺	☺	☺
Ballestri 2012	☹	☹	?	☺	☺	☺	☺
Bastati 2014	☹	☹	☺	☺	☹	☺	☺
Chen 2011	☹	☹	☺	☺	☺	☺	☹
Costa-Silva 2018	☺	☹	☺	☹	☺	☺	☺
Dillman 2018	☹	☹	?	☺	☹	☺	☺
Eddowes 2018	☺	☹	☺	☺	☺	☺	☺
Fierbinteanu Braticevici 2013	☹	☹	☺	☺	☹	☺	☺
Gallego-Duran 2016	☺	☺	☺	☺	☺	☺	☺
Guzman-Aroca 2012	☺	☹	☺	☺	☹	☺	☹
Iijima 2007	?	☹	☹	☹	☺	☺	☺
Imajo 2016	?	☹	?	☺	☺	☺	☹
Kikuchi 2009	?	☹	☺	?	☺	☺	☺
Kim 2017	☺	☹	☺	☺	☺	☺	☺
Lee 2016	☹	☹	☺	☺	☺	☺	☺
Leporq 2017	☹	☹	☺	☺	☹	☺	☺
Liang 2007	☺	☹	☺	☺	☹	☺	☹
Lirussi 2009	?	☹	?	?	☺	☺	☺
Loomba 2014	☹	☹	☺	☺	☺	☺	☺
Loomba 2016	☹	☹	☺	☺	☺	☺	☺
Naganawa 2018	☹	?	☺	?	☹	☺	☺
Parente 2015	☺	☹	?	☺	☺	☺	☺
Parente 2018	☺	☹	?	?	☺	☺	☺
Park 2017	☺	☹	☺	?	☺	☺	☺
Pavlides 2017	☺	☹	☺	☺	☺	☺	☺
Petrick 2015	☹	☹	?	☹	☹	☺	☹
Smits 2015	☹	☹	☺	☹	☹	☺	☺
Tarantino 2009	☹	☹	?	☺	☹	☺	☺



Supplementary Table 3: risk of bias analysis according to the QUADAS-2 tool for included studies with accuracy results.

Study	**Reasons for high/unclear risk of bias or high/unclear applicability concerns**
Abrigo 2014	Unclear if consecutive sample of patients. No prespecified threshold for index test.
Ballestri 2012	Non-consecutive sample of patients. No prespecified threshold for index test. Unclear if blinding in interpretation of reference standard (performed after index test).
Bastati 2014	Retrospective study on patients with previously biopsy-proven NAFLD and a complete gadoxetic acid-enhanced MR imaging protocol (the clinical indication for MR examination is not reported), leading to both high risk of bias and high concerns on applicability. No prespecified threshold for index test.
Chen 2011	Retrospective study on patients with previous MRE and liver biopsy for NAFLD, presumably non-consecutive and introducing inappropriate exclusion (no data about NAFLD patients with previous biopsy and without MRE). No concerns about applicability (patients match the review question). No prespecified threshold for index test. The reference standard is free of bias, but the target condition is defined differently from that specified in the review question (inflammation without fibrosis and fibrosis stage 1-4 grouped together as NASH)
Costa-Silva 2018	No prespecified threshold for index test. Large time interval between index test and reference standard (mean interval of 3.7 ± 3.4 months; range, 0–11 months), acknowledged as a limitation.

Dillman 2018	Retrospective study including patients who had undergone a quantitative liver MRI examination within 6 months of liver biopsy (n=69) among a prospectively recruited consecutive cohort of 160 children and young adults up to 21 years old with biopsy-confirmed NAFLD. This introduces inappropriate exclusion since no data are reported on patients who did not undergo MRI within 6 months of liver biopsy (91/160). We also have applicability concerns since the population include only children and young adults. No prespecified threshold for index test. Unclear if blinding in interpretation of reference standard (mostly performed after index test).
Eddowes 2018	No prespecified threshold for index test.
Fierbinteanu Braticevici 2013	Inappropriate exclusion of borderline patients ("difficult to diagnose" cases), limitation acknowledged by the authors, also introducing applicability concerns. No prespecified threshold for index test.
Gallego-Duran 2016	
Guzman-Aroca 2012	No prespecified threshold for index test. All included patients are morbidly obese, undergoing bariatric surgery (applicability concern). The reference standard is free of bias, but the target condition is defined differently from that specified in the review question: comparisons are between group A (simple steatosis) and groups B-C (inflammation and/or fibrosis); moreover, all patients in group B (inflammation) were in Matteoni group 2 (with lobular inflammation but without hepatocellular ballooning).
Iijima 2007	Unclear if consecutive sample of patients. No prespecified threshold for index test (and unclear blinding). Dixon criteria are not likely to correctly classify the target condition; moreover, blinding in reference standard interpretation is unclear. Not all patients received the same reference standard: in 31/66 patients a liver biopsy was performed, while in the remaining 35/66 NASH was clinically excluded through Dixon criteria. Moreover, the time interval between index test and reference standard is unclear.
Imajo 2016	Unclear if consecutive sample of patients. No prespecified threshold for index test. Unclear blinding in the interpretation of reference standard (not clear if performed before or after index tests).

	The definition of the target condition does not perfectly match the review question (NASH defined as steatosis, inflammation, ballooned hepatocytes, AND pericellular/perisinusoidal fibrosis).
Kikuchi 2009	Unclear if consecutive sample of patients. No prespecified threshold for index test (and unclear blinding). Unclear (not reported) time interval between index test and reference standard.
Kim 2017	No prespecified threshold for index test (and unclear blinding).
Lee 2016	Retrospective study including NAFLD patients who underwent liver biopsy with concomitant TE in the same day: inappropriate exclusion since no data are reported about patients who did not undergo both the diagnostic procedures in the same day. No prespecified threshold for index test (and unclear blinding).
Leporq 2017	Retrospective feasibility study on patients with previously biopsy-proven NAFLD (SS and NASH) and who undergone MR (the clinical indication for MR examination is not clearly stated), leading to both high risk of bias and high concerns on applicability. No prespecified threshold for index test (and unclear blinding).
Liang 2007	No prespecified threshold for index test. All included patients are morbidly obese, undergoing bariatric surgery (applicability concern). The reference standard is free of bias, but the target condition is defined differently from that specified in the review question: NASH defined as the presence of fibrosis (grade 1 or higher) or acinar zone 3 hepatocellular injury with ballooning degeneration (grade 2 or higher).
Lirussi 2009	Unclear if consecutive sample of patients. No prespecified threshold for index test. Unclear if blinding in interpretation of reference standard (performed after index test). Time interval between index test and reference standard is not reported.
Loomba 2014	The study includes patients with biopsy-proven NAFLD who also underwent MRE. The authors showed that, compared to patients who had an MRE, those who did not have an MRE (excluded) were less likely to have NASH and features of advanced disease on biopsy. So it is to be considered inappropriate exclusion. No concerns about applicability (patients match the review question).

	No prespecified threshold for index test.
Loomba 2016	The study includes patients with biopsy-proven NAFLD who also underwent 2D and/or 3D MRE (inappropriate exclusion, no data about patients who did not undergo MRE). No concerns about applicability (patients match the review question). No prespecified threshold for index test (and unclear blinding).
Naganawa 2018	Retrospective study on patients with previous liver biopsy for NASH and previous Non-enhanced CT, presumably non-consecutive and introducing inappropriate exclusion (no data about patients with previous biopsy and without CT). The clinical indication for CT is not clearly stated, leading also to concerns on applicability. Unclear if blinding in interpretation of index test. Unclear if time interval is appropriate (only mean time interval is reported, 3 months).
Parente 2015	No prespecified threshold for index test (and unclear blinding). Unclear if blinding in interpretation of reference standard (not clear if performed before or after biopsy).
Parente 2018	No prespecified threshold for index test. Unclear if blinding in interpretation of reference standard (not clear if performed before or after biopsy). Unclear if appropriate time interval between index test and reference standard.
Park 2017	No prespecified threshold for index test. Unclear if all patients were included in analysis (in tables histological features are reported for 100/104 patients).
Pavlides 2017	No prespecified threshold for index test.
Petrick 2015	Inappropriate exclusion of patients without US (no data about these patients); unclear if consecutive sample of patients. No prespecified threshold for index test. Unclear if blinding in interpretation of reference standard (performed after index test). Inappropriate time interval between index test and reference standard. All included patients are morbidly obese, undergoing bariatric surgery (applicability concern). The reference standard is free of bias, but the target condition is defined differently from that specified in the review question: steatohepatitis defined and graded based on lobular inflammation

	(no mention of ballooning) and NASH defined as the presence of steatohepatitis, fibrosis or cirrhosis.
Smits 2015	Different inclusion criteria for NASH and simple steatosis groups: NASH patients included if NAS≥5, leading to inappropriate exclusion of borderline patients (also introducing applicability concerns). Unclear if consecutive sample of patients. No prespecified threshold for index test (and unclear blinding). Not all patients received the same reference standard: in 6/11 NASH was clinically excluded because affected by heterozygous familial hypobetalipoproteinemia. Inappropriate time interval between index test and reference standard.
Tarantino 2009	Inappropriate exclusion of patients without visceral adiposity or with BMI>40, leading also to applicability concerns. No prespecified threshold for index test. Unclear if blinding in interpretation of reference standard (performed after index test).
Tomita 2008	No prespecified threshold for index test (and unclear blinding).
Zardi 2011	Retrospective study including patients who underwent US and liver biopsy with diagnosis of steatosis or NASH (the clinical indication for US and biopsy is not reported), leading to high risk of bias and high applicability concern. No prespecified threshold for index test. Unclear if blinding in interpretation of reference standard (performed after index test). Unclear if appropriate time interval between index test and reference standard.

Supplementary Table 4: Reasons for attribution of intermediate/high risk of bias or intermediate/high applicability concern. It is to be clarified that, since considered index tests are mostly not used in clinical practice, a domain on uninterpretable results was added in the flow and timing section of the QUADAS-2 tool, considering them as introducing a risk if excluded a priori and not quantified.

Elastographic techniques					
Study	**Study design and index test**	**Population and NASH prevalence**	**NASH definition**	**Accuracy simple steatosis vs NASH**	**Associations and Reproducibility**
Transient Elastography (TE) – liver stiffness (LS) Coefficient attenuation parameter (CAP)					
Attia 2016[s5]	Prospective LS	97 overweight or obese patients 48(60%) with NASH	NAS≥5		LS not significantly different between those with and without NASH (p=0.12 in overweight and 0.93 on obese patients respectively).
Cassinotto 2016[s4]	Prospective LS	291 patients, NASH prevalence not reported	NASH Clinical Research Network criteria and NAS, no further NASH definition		TE LS slightly associated with NAS score (Spearman's correlation coefficient of 0.28) in the univariate analyses, but not in multivariate analysis after correction for fibrosis stage.
Eddowes 2018[26]	Prospective LS	50 patients; 38(76%) with NASH, 47 with reliable TE	Steatosis, lobular inflammation and ballooning	AUROC=0.82(0.70-0.94) AUROC for NAS≥5=0.74(0.59-0.89)	LS significantly different between simple steatosis and NASH (p<0.001) LS significantly associated with: hepatocyte ballooning (p=0.003), NAS (p=0.026) and lobular inflammation (p=0.035)
Imajo 2016[18]	Prospective LS; CAP	142 patients; 108(76%) with NASH, 127 with reliable TE	Steatosis, inflammation, ballooning, and pericellular/perisinusoidal fibrosis	AUROC=0.80(0.73–0.88) † AUROC for NAS≥5=0.65(0.54-0.77) †	LS significantly associated with grade of hepatic inflammatory activity (p=0.025); association not maintained in multivariable analysis after correction for fibrosis. CAP not associated with inflammation.
Lee 2016[27]	Retrospective LS; CAP	183 patients 94(51.4%) with NASH	Steatosis, inflammation and ballooning; NAS≥5	LS>7 kPa: AUROC=0.751(0.677–0.824); sensitivity=86.2%, specificity=58.4% CAP>250 dB/m: AUROC=0.743(0.669–0.816), sensitivity=96%, specificity=49% Score based on LS, CAP and ALT: AUROC=0.812(0.724–0.880)	In multivariable analyses CAP > 250 dB/m (OR 4.133, 95% CI 1.037–16.470), LS > 7 kPa (OR 3.399, 95% CI 1.295–8.291), and ALT level > 60 IU/L (OR 7.557, 95% CI 2.997–19.059) independently predicted NASH

Lee 2017[s20]	Prospective LS	94 patients; 30(32%) with NASH	NAS≥5		LS associated with lobular inflammation (p=0.009). No association between LS and NAS, ballooning or NASH presence.
Park 2017[5]	Prospective LS	104 patients 76(76%) with NASH ‡	NAS≥2	AUROC=0.35(0.22-0.49) Cut-off>5.6 KPa: sensitivity=61.1%, specificity=59.1%, PPV=83%, NPV=31.7%	
Lupsor 2010[s6]	Prospective LS	72 patients; NASH prevalence not reported	Brunt 1999 criteria. No further NASH definition		LS associated with fibrosis (r=0.661, p<0.0001), ballooning (r=0.385, p=0.001), lobular inflammation (r=0.364, p=0.002) and steatosis (r= 0.435, p<0.0001). -in multivariable analysis for LS including fibrosis, ballooning, lobular inflammation and steatosis: fibrosis is the only independent factor (p<0.001).
Yilmaz 2014[s21]	Prospective LS	235 patients 135(57%) with definite NASH	NAS≥5		LS >7.9 KPa in 34/135 definite NASH vs 12/100 simple steatosis+borderline NASH (p<0.01); multivariable logistic regressions: definite NASH independently associated with LS >7.9 KPa (OR 2.11; 95%CI 1.08-4.09) -prediction model with diabetes, LS >7.9 KPa and positive cytokeratin 18 M30 fragment determined an 81% probability of distinguishing definite NASH from other forms of NAFLD.
Yoneda 2008[s7]	Prospective LS	102 patients, 97 with reliable TE, NASH prevalence not reported	Steatosis, lobular inflammation and either ballooning or zone 3 fibrosis		Significant association between inflammatory activity grade and LS (p=0.0063); in multivariable analysis for LS including steatosis, inflammation and fibrosis, only fibrosis remains independently associated with LS
Yoneda 2010[s22]	Prospective LS	54 patients, NASH prevalenc	Steatosis, lobular inflammation and either ballooning or		stepwise increase in LS with increasing inflammatory activity (p =0.016)

		e not reported	zone 3 fibrosis		
Acoustic Radiation Force Impulse (ARFI) and Supersonic Shear Imaging (SSI) – shear wave velocity (SWV) and liver stiffness					
Attia 2016[s5]	Prospective ARFI velocity	97 overweight or obese patients 48(60%) with NASH	NAS≥5		ARFI velocity not associated with the presence of NASH in all patients (p= 0.96), overweight patients (p=0.53) and obese patients (p=0.83)
Cassinotto 2016[s4]	Prospective SSI and ARFI liver stiffness	291 patients, NASH prevalence not reported	NASH Clinical Research Network criteria and NAS, no further NASH definition		ARFI and SSI stiffness not associated with NAS in univariate analyses.
Fierbinteanu Braticevici 2013[28]	Prospective SWV	64 patients 43(67%) with NASH	Brunt 1999/Kleiner 2005 criteria. Patients divided into simple steatosis and NASH, *borderline patients excluded*.	AUROC=0.87 Cut-off>1.10 m/s: sensitivity=77%, specificity=72%, PPV=85%, NPV=60%	Median SWV significantly higher in NASH than simple steatosis (p<0.001). SWV correlated with inflammation grade (r = 0.386, p<0.001).
Guzman-Aroca 2012[19]	Prospective SWV	32 bariatric patients 24(75%) with NASH/fibrosis (18 with inflammation and 6 with fibrosis)	Matteoni 1998 criteria. Patients categorized as simple steatosis, inflammation and fibrosis. *Comparisons between SS and NASH/fibrosis.*	NASH and/or fibrosis vs simple steatosis: AUROC=0.9 Cut-off 1.3 m/s: sensitivity=85%, specificity=83%, PPV=89%, NPV=77%	SWV significantly higher in inflammation than simple steatosis (p<0.001)
Lee 2017[s20]	Prospective SSI and ARFI SWV and liver stiffness	94 patients; 30(32%) with NASH	NAS≥5		ARFI liver stiffness but not SSI liver stiffness associated with lobular inflammation (p=0.006 and p=0.352 respectively). No association between liver stiffness and NAS.
Osaki 2010[s1]	Prospective SWV	23 patients; 100%	Brunt 1999 criteria. No further		No significant difference in SWV among inflammatory activity grades (p= 0.12);

		with NASH; 19 with reliable SWV	specified NASH definition.		good reproducibility for SWV repeated measures in the same lobe (average coefficient of variance 11.9% ± 8.0%); SWV largely varied between right and left lobe
Palmeri 2011[s2]	Mixed retrospective/ prospective ARFI shear stiffness	172 patients, 135 with reliable ARFI, NASH prevalence not reported	NASH Clinical Research Network criteria, no further NASH definition		hepatocyte ballooning and inflammation were not significantly associated with shear stiffness (p=0.11 for ballooning and p=0.69 for inflammation)
Praveenraj 2017[s3]	Retrospective SWV	28 bariatric patients, all with NAS 0 to 4	NASH Clinical Research Network criteria and NAS, no further NASH definition		Significant negative correlation between SWV and hepatic steatosis grade (r=-0.471; p=0.011). No significant correlation between SWV and inflammation or ballooning.
Yoneda 2010[s22]	Prospective ARFI median velocity	54 patients, NASH prevalence not reported	Steatosis, lobular inflammation and either ballooning or zone 3 fibrosis		ARFI median velocity significantly different among inflammatory activity groups (p=0.016), without a stepwise change.
Magnetic Resonance Elastography (MRE) – liver stiffness (LS)					
Chen 2011[23] (2D MRE)	Retrospective LS	58 patients 36(72%) with NASH/fibrosis (7 inflammation and 29 fibrosis)	Brunt 1999. Patients categorized as simple steatosis, inflammation without fibrosis, and NAFLD with fibrosis, the latter two classified as NASH.	AUROC=0.93 Cut-off>2.74 KPa: sensitivity=94%, specificity=73%, PPV=85%, NPV=89% Cut-off>2.90 KPa: sensitivity=83%, specificity=82%, PPV=88%, NPV=75%	LS greater in inflammation without fibrosis than in simple steatosis (p=0.028) When LS is the response variable while MRI relative fat fraction (RFF), inflammation grade and fibrosis stage are the effect variables, partial correlation coefficients are 0.095 (p=0.52) for RFF, 0.462 (p=0.0097) for grade, 0.651 (p<0.0001) for stage
Costa-Silva 2018[29] (2D MRE)	Prospective LS	49 patients 25(51%) with NASH	NAS≥5	AUROC=0.79 Cut-off 3.24 Kpa: sensitivity=72%, specificity=88%, PPV=86%, NPV=72%. in fibrosis=0 patients (n=21): AUROC=0.78	LS significantly different between simple steatosis and definite NASH (p=0.002) and between borderline NASH and definite NASH (p=0.011). No differences between simple steatosis and borderline NASH (p =0.14).

				Cut-off 3.22 kPa: sensitivity=69%, specificity=87%	Moderate correlation between LS and NAS (r=0.549, p<0.0001) in pts with fibrosis=0 (n=21): LS significantly different between simple steatosis/borderline NASH and definite NASH (p=0.041)
Imajo 2016[18] (2D MRE)	Prospective LS	142 patients; 108(76%) with NASH	Steatosis, inflammation, ballooning and pericellular/perisinusoidal fibrosis	AUROC=0.81 § AUROC for NAS≥5=0.77 §	LS significantly associated with grade of hepatic inflammatory activity (p=0.046); association not maintained in multivariable analysis after correction for fibrosis.
Loomba 2014[30] (2D MRE)	Prospective LS	117 patients 106(91%) with NASH	Kleiner 2005. Borderline with definite NASH.	AUROC=0.73 Cut-off 3.26 Kpa: sensitivity=42%; specificity=92%; PPV=95%; NPV=32%	
Loomba 2016[31] (2D and 3D MRE)	Prospective LS	100 patients 87(87%) with NASH	Kleiner 2005. Borderline with definite NASH.	2D MRE (60 Hz): AUROC=0.75; optimal cut-off=2.92 Kpa; 3D MRE (60 Hz): AUROC=0.76; optimal cut-off=2.42 Kpa; 3D MRE (40 Hz): AUROC=0.74; optimal cut-off=1.93 KPa	
Park 2017[5] (2D MRE)	Prospective LS	104 patients 76(76%) with NASH ‡	NAS≥2	AUROC=0.70 Cut-off>2.53 KPa: sensitivity=63.9%, specificity=68.2%, PPV=86.8%, NPV=36.6%	
Dillman 2018[41]	Retrospective LS	69 children and young adults ≤21 years old; 37(54%) with NASH	NAS≥5		LS not significantly associated with any histopathologic variable; not significantly different (p=0.09) among categories (simple steatosis, NASH without fibrosis, NASH with fibrosis)

Supplementary Table 5: Result synthesis for elastographic techniques including also secondary outcomes and studies not exploring accuracy. AUROC: area under the receiver operating characteristic curve. NAS, NAFLD Activity Score. † For the combination of liver stiffness and CAP; ‡ Histological data reported for 100/104 patients; § For the combination of MRE and Proton Density Fat Fraction.

US non-elastographic techniques					
Study	**Study design and index test**	**Population and NASH prevalence**	**NASH definition**	**Accuracy Simple steatosis vs NASH**	**Associations and Reproducibility**
US B-mode parameters and scores					
Ballestri 2012[32]	Prospective US-fatty liver indicator (US-FLI) (2-8): liver/kidney contrast (2–3), US posterior attenuation (0–1), vessel blurring (0–1), difficult visualization of gallbladder wall (0–1) or diaphragm (0–1), focal sparing (0–1)	53 patients; 35(66%) with NASH	Steatosis, lobular inflammation and ballooning; severe NASH for NAS≥ 5	AUROC=0.76 for NASH; 0.80 for severe NASH. US-FLI<4 ruled out severe NASH with NPV=94%; specificity=46%.	US-FLI significantly different in NASH vs simple steatosis (p=0.002) and NAS ≥5 vs <5 (p<0.001); associated with lobular inflammation, ballooning, NAS and Brunt grade (p<0.05). † After correction for age, gender, total cholesterol and metabolic syndrome, US-FLI remained associated with NASH (OR 2.236; p=0.007). In a different cohort of 31 fatty liver patients inter-observer agreement= 0.805-0.882 (p<0.001).
Ballestri 2017[s23]	Retrospective US-FLI	352 patients with different chronic liver diseases, 123 with NAFLD (70.7% of them with NASH).	Steatosis, lobular inflammation and ballooning		US-FLI associated with lobular inflammation in all patients (rho=0.380; p<0.001) and in NAFLD (rho=0.490; p<0.001), with ballooning in all patients (rho=0.619; p<0.001) and in NAFLD (rho=0.485; p<0.001), with Brunt grade in patients with NAFLD (rho=0.622; p<0.001).
Liang 2007[20]	Prospective US fatty score (FS) (0–8):	101 obese bariatric patients;	Fibrosis (≥grade 1) or acinar zone 3	FS: AUC=0.79; cut-off 7; sensitivity=81%;	US parameters and NASH:

	parenchymal echogenicity, far gain attenuation, gallbladder (GB) wall blurring, portal vein (PV) wall blurring, and hepatic vein (HV) blurring. Modified FS (MFS) (0–2): 0 for FS<7 and the sum of parenchymal echogenicity + gallbladder wall blurring <3; score 1 for FS≥7 or the latter ≥3; score 2 for FS≥7 and the latter ≥3	72(71%) with NASH	hepatocellular injury with ballooning (≥grade 2)	specificity=66%; accuracy=76%; PPV=85%; NPV=58% MFS: AUC=0.82; cut-off 3; sensitivity=72%; specificity=86%; accuracy=76%; PPV=93%; NPV=56%	-Parenchymal echogenicity: r=0.411, p<0.001; -Far gain attenuation: r=0.219, p=0.02; -GB wall blurring: r=0.430, p<0.001; -PV wall blurring: r=0.265, p=0.007; -FS: r=0.464; p<0.001; -MFS: r=0.557; p<0.001. Multivariable analysis: MFS>2 independent predictor of NASH (p=0.005, OR=10.47). Intra-observer agreements for parenchymal echogenecity, GB wall blurring, FS and MFS was perfect (κ=0.87, 0.82, 0.79 and 0.83), for far gain attenuation, PV blurring and HV blurring substantial (κ=0.76, 0.65 and 0.75). Inter-observer agreements for HV blurring and FS was moderate (κ=0.57 and 0.53), for the remaining parameters substantial (κ from 0.61 to 0.75).
Lirussi 2009[33]	Prospective US PATT (perihepatic adipose tissue thickness)	65 patients (33 with liver biopsy); 27(82%) with NASH	Brunt 1999. Borderline with definite NASH	Cut-off 11.8 mm: sensitivity=100%, Specificity=50%, AUROC=75%. To predict necro-inflammatory activity grading: sensitivity=80%, specificity=50%, AUROC=60%	
Petrick 2015[22]	Prospective US-Fatty liver (mild, moderate, or severe according to the fall in echo amplitude, extent	513 bariatric patients 146(28%) with steatohepatitis; 164(32%) with NASH.	Brunt 1999. Steatohepatitis defined as lobular inflammation;	For steatohepatitis: US fatty liver (mild+): sensitivity=89%; specificity=45%; PPV=39%;	US-Fatty liver significantly different between simple steatosis and steatohepatitis (p<0.0001)

	of liver/kidney discrepancy and of echo loss from portal vein)		NASH defined as steatohepatitis, fibrosis or cirrhosis	NPV=91%; Accuracy=58%	
Tarantino 2009[34]	Prospective Spleen longitudinal diameter (SLD)	83 patients; 43(52%) with NASH	Kleiner 2005. Lobular inflammation 0-3, no further specified NASH definition	AUROC=0.920 Cut-off 116 mm: sensitivity=88%, specificity=95%	SLD significantly higher in NASH vs simple steatosis, p=0.0001 after correction for age, sex, waist circumference and metabolic syndrome; SLD>120 mm more frequent in NASH than simple steatosis (p=0.0001); in NASH group SLD not associated with inflammation grading; averaged intraobserver error for SLD<6% interobserver error for SLD: concordance correlation coefficient high (Pearson's p=0.93)
Zardi 2011[35]	Retrospective US score (0-6): echo amplitude attenuation (0–2), focal fat sparing (0–1), splenic diameter (0–3).	94 patients; 74(79%) with NASH	Steatosis, lobular inflammation and ballooning.	Cut-off≥5: sensitivity=74%, specificity=66%; only echo attenuation and focal fat sparing (cut-off=1): sensitivity=92%, specificity=75%.	NASH predictors: echo attenuation (p<0.05; OR=3.43; CI=1.02-11.57), focal fat sparing (p<0.05; OR=3; CI=1.02-11.88), splenic diameter (p<0.05; OR=1.66; CI=1.04-3.26). US-score and a score with only focal fat sparing and splenic diameter both different between NASH and simple steatosis (p<0.05 and p=0.026).
Ataseven 2005[s16]	Prospective US grade of hyperechogenicity	22 NASH patients and 20 healthy controls	Steatosis, inflammation and ballooning		US hyperechogenicity grade not associated with histopathological grade (p=0.553).

Bril 2015[s24]	Prospective US score: sum of liver parenchymal echogenicity, far gain attenuation, gallbladder wall blurring, portal vein blurring, hepatic vein blurring (0-2 each)	146 patients, NASH prevalence not reported	Steatosis, inflammation and ballooning		No correlations between US parameters and histological inflammation, ballooning and fibrosis after adjustment for steatosis grade, apart from the association between far gain attenuation and inflammation grade (r=0.17 p<0.05). Inter-rater agreements for portal vein blurring=79% (j=0.45, p<0.001), far gain attenuation=70% (j=0.43, p<0.001), hepatic vein blurring=65% (j=0.36, p<0.001), parenchymal echogenicity=51% (j=0.13, p=0.04), gallbladder wall blurring=51% (j=0.11, p=0.07).
Saadeh 2002[s14]	Prospective US hyperechogenicity grade and pattern	25 patients, 17(68%) with NASH	Steatosis, inflammation and ballooning		US features did not distinguish NASH nor the presence of hepatocyte ballooning, Mallory's hyaline, or fibrosis. Intraobserver agreement for US pattern of steatosis: substantial (ƙ=0.77; 95%CI 0.34 –1.0); for US grade of steatosis: substantial (ƙ=0.63; 95%CI 0.36–0.90). Interobserver agreement for US pattern and grade of steatosis only fair to moderate (ƙ from 0.40 to 0.43).
Contrast-Enhanced Ultrasound (CEUS)					
Iijima 2007[36]	Prospective	66 patients (liver biopsy in	Brunt 1999. NASH for	Signal intensity 5 minutes	NASH patients had a decreased signal

	CEUS Signal intensities 5 and 20 minutes after Levovist administration	31 patients: 21 with NASH; *in the remaining 35 NASH was clinically excluded*) + 10 healthy volunteers ‡	presence of parenchymatitis independently of fibrosis	Cut-off=137.8: sensitivity=100%, specificity=95%, accuracy=80%. Signal intensity 20 minutes Cut-off=43.6: sensitivity, specificity and accuracy=100%.	intensity after Levovist and a more rapid decrease of signal intensity when compared to simple steatosis patients. Signal intensities at 5 and 20 minutes significantly lower in NASH vs simple steatosis
Moriyasu 2005[s25]	Prospective Time intensity curves after Levovist administration	60 patients with NAFLD; 15(25%) with NASH	Not clear NASH definition		NASH vs simple steatosis: time intensity curves showed lower peaks and more rapid disappearance of the microbubbles. all NASH cases vs almost no one in simple steatosis patients showed mild to severe decrease of uptake, not according to the degree of steatosis but to the degree of fibrosis
Doppler US indices					
Goncalves 2016[s8]	Prospective PVI-Portal Venous Pulsatility Index; HARI-Hepatic Artery Resistance Index; HAPI-Hepatic Artery Pulsatility Index; HVWP-Hepatic Vein Waveform Pattern	49 patients; 44 (90%) with NASH and 37 (75%) with definite NASH	Steatosis, lobular inflammation and ballooning; definite NASH for NAS≥ 5		No significant association between any Doppler index and inflammation
Tarantino 2009[34]	Prospective Splenic Artery Resistance Index (SARI)	83 patients; 43(52%) with NASH.	Kleiner 2005. Lobular inflammation 0-3, no further specified NASH definition		SARI not significantly different between NASH and simple steatosis. Interobserver error: SARI concordance correlation coefficient sufficiently high (Pearson's P=0.89).

Supplementary Table 6: Result synthesis for US non-elastographic techniques including also secondary outcomes and studies not exploring accuracy. AUROC: area under the receiver operating characteristic curve; CI: confidence interval; NAS: NAFLD Activity Score. † In a subsequent analysis, after controlling for steatosis grade, the association of US-FLI with lobular inflammation (ρ=0.389, p=0.004 adjusted) and NAS (ρ=0.342, p=0.013 adjusted) remained significant; while the correlations with ballooning and Brunt grading of necroinflammatory activity went lost.[s26] ‡ not clear if included in analysis.

MR non-elastographic techniques					
Study	**Study design and index test**	**Population and NASH prevalence**	**Definition of NASH**	**Accuracy simple steatosis vs NASH**	**Associations and Reproducibility**
^{1}H-MRS and/or ^{31}P-MRS metabolites					
Abrigo 2014[21] (*^{31}P-MRS)*	Prospective Nucleotide Triphosphate (α peak)/Triphosphate (αNTP/TP)	132 patients 95(72%) with NASH	Matteoni 1998. NASH for type 3 and 4 (fat accumulation and ballooning ± Mallory hyaline or fibrosis)	α-NTP/TP: AUROC=0.71 Cut-off≤10.57%: sensitivity=28%; specificity=91%; PPV=78%; NPV=43% Cut-off≤16.36%: sensitivity=91%; specificity=16%; PPV=65%; NPV=50%.	αNTP/TP significantly lower in NASH vs simple steatosis (p <0.001) αNTP/TP significant predictor of NASH after adjusting for diabetes, BMI and ALT. Ballooning mildly correlated with α-NTP (ρ=0.299, p<0.001). Repeated α-NTP measurements showed CV of 7.4% and ICC of 0.838
Kim 2017[37] (long echo time ^{1}H-MRS)	Prospective Alanine (Ala), lactate+trygliceride (Lac+TG)	26 patients; 11(42%) with NASH	NAS≥5	Ala: AUROC=1.00 Cut-off>16.04%: sensitivity=100%, specificity=100% Lac+TG: AUROC=0.78	Ala significantly higher in NASH vs simple steatosis (p<0.001); positively correlated with

				Cut-off>360.8%: sensitivity=82%, specificity=67%	lobular inflammation (r=0.513, p=0.007) and NAS (r=0.743, p<0.001) Lac+TG significantly higher in NASH vs simple steatosis (p<0.001); positively correlated with NAS (r=0.474, p=0.014). Averaged ICCs for all metabolites >0,9 (p<0.001)
Cortez-Pinto 1999[s9] (^{31}P-MRS before, during and after fructose administration)	Prospective Adenosine triphosphate (ATP) peak amplitude (^{31}P spectra before, at 12 minutes after, and at 60 minutes after fructose infusion)	8 NASH and 7 healthy controls	Not clear NASH definition		In the 2 groups liver ATP levels are similar before fructose; at 12 min after fructose a similar significant decline in liver ATP stores is registered (β-ATP peak decrease and phosphomonoester peak increase); at 60 min after fructose ATP levels are significantly lower in NASH pts than controls (slower ATP recovery) .
Sevastianova 2010[s27] (^{31}P-MRS)	Prospective Phosphomonoester (PME), phosphodiester (PDE), phosphocholine, inorganic phosphate, glycerophosphorylethanolamine (GPC), phosphoenolpyruvate, γ -NTP, α -NTP, nicotinamide adenine dinucleotide phosphate (NADPH), uridine diphosphoglucose, and β -NTP, all data expressed as ratios between single	22 NAFLD patients, 9(41%) with NASH, 12 healthy controls, 9 cirrhotic patients	Brunt 1999 criteria; no further specified NASH definition		NADPH/(PME + PDE) in NASH and cirrhosis pts significantly higher than in controls (p= 0.017) and increased as a function of the severity of liver disease (p=0.007). Other metabolites not significantly different

	resonances or over the sum of PME and PDE				between NASH patients and simple steatosis or controls.
Traussnigg 2017[s10] (^{1}H- and ^{31}P-MRS)	Prospective ^{1}H-MRS: lipid saturation (SI), unsaturation (UI), polyunsaturation (PUI) and monounsaturation (MUI) indices ^{31}P-MRS: PME= phosphoethanolamine (PE) + phosphocholine, PDE= glycerophosphocholine (GPC) + glycerophosphoethanolamine, uridine-diphosphoglucose, NADPH, inorganic phosphate (Pi), phosphatidylcholine, α- and γ-ATP, residual phosphocreatine (PCr) and total phosphorus (TP), chemical exchange rate constant k of the Pi-to-ATP reaction, and unidirectional forward exchange flux (FATP))	30 patients; 22(73%) with NASH	Steatosis, ballooning, lobular inflammation		No differences between simple steatosis and NASH in SI, PUI and MUI indices. Higher PME/TP in grade 2 ballooning than grade 1 (p=0.015); NADPH/(PME+PDE) (p=0.048), NADPH/γ-ATP (p=0.028) and NADPH/TP (p=0.026) ratios increased from grade 0 to 2 ballooning; k significantly lower in NASH vs simple steatosis (p=0.003), associated with lobular inflammation (p=0.007); FATP significantly lower in NASH vs simple steatosis (p<0.001),. associated with lobular inflammation (p<0.001).
Valkovic 2014[s11] (^{31}P-MRS with saturation transfer technique)	Prospective α- β- and γ-ATP, Pi, PME, PDE, nicotinamide adenosine dinucleotide (NADH), uridine diphosphoglucose (UDPG), PCr, k of the Pi-to-ATP reaction, FATP	16 patients, 10(63%) with NASH, 9 healthy controls	Brunt 1999 criteria; no further specified NASH definition		k significantly lower in NASH vs simple steatosis (p<0.01) and controls (p<0.01); Pi significantly lower in NASH vs simple steatosis (p<0.05); FATP significantly lower in NASH vs simple steatosis

					(p<0.01) and controls (p<0.01). Reproducibility on 6 healthy volunteers: -mean test-retest variability CV=9.0 %
Multiparametric MRI (Liver MultiScan)- corrected T1 (cT1), Liver Inflammation and Fibrosis (LIF) score					
Eddowes 2018[26]	Prospective T1 corrected for T2* (cT1)	50 patients 38(76%) with NASH	Lobular inflammation and ballooning	AUROC for NASH vs SS=0.69 AUROC for NAS≥5 vs <5=0.74	cT1 associated with ballooning (p=0.001) and NAS (p=0.013); different in simple steatosis vs NASH (p=0.004). mean coefficient of variance over 10 normal volunteers for cT1=1.8% †
Pavlides 2017[38]	Prospective LIF score (0-4) based on cT1 cut-offs.	71 patients 46(65%) with NASH	Steatosis, ballooning, lobular inflammation	AUROC=0.80 Cut-off 1.4: sensitivity=91%, specificity=52%	LIF associated with ballooning (p<0.0001); lobular inflammation p=0.024); SAF score (p<0.0001); different in simple steatosis vs NASH (p<0.0001). In a subgroup of 33 patients: mean of the differences between 2 observers=-0.106 (p=0.31); 95% levels of agreement from 1.061 to -1.258.
Diffusion weighted (DW) MRI and Intravoxel Incoherent Motion DW MRI (Apparent Diffusion Coefficient – ADC, D, D, f)*					
Parente 2015[40] (IVIM; 10 b-values)	Prospective Pure molecular-based (D), perfusion-related (D*), and vascular (f) Fractions	59 T2DM patients; 22(37%) with NASH	Steatosis, lobular inflammation and ballooning	-D: AUROC=0.742; cut-off 0.760: sensitivity=69% specificity=66%; -D*: AUROC=0.678; cut-off 41.45: sensitivity=68% specificity=71%;	NASH and simple steatosis groups significantly different in terms of D (p=0.002) and D* (0.023); no significant difference for f

				-f: AUROC=0.607; cut-off 34.23: sensitivity=49% specificity=70%.	
Manning 2017[s12] (DWI; 3 b-values)	Prospective ADC, D, F	64 children; 34(53%) with NASH	Kleiner 2005 criteria; borderline together with definite NASH		No significant associations between DWI parameters and ballooning, inflammation or NASH vs simple steatosis. Interobserver reproducibility (subgroup of 20%) was excellent (ICCs 0.984, 0.992, 0.976 for ADC, D, F); no systematic bias between readers ($p>0.05$).
Murphy 2015[s13] (DWI; 3 b-values)	Prospective ADC, D, F	89 patients; 79(89%) with NASH	Kleiner 2005 criteria; borderline together with definite NASH		In univariate model ADC decreased with inflammation and ballooning, and F decreased with inflammation, ballooning and NASH (all $p<0.05$). In multivariable models for DWI parameters: inflammation, ballooning and NASH had no independent effect after correction for steatosis/fibrosis.
Quantitative susceptibility MRI					
Leporq 2017[39]	Retrospective Susceptibility (ppm)	32 patients; 20(62.5%) with NASH	Steatosis, ballooning, lobular inflammation	AUROC=0.91	Susceptibility (ppm) was significantly decreased in NASH vs simple steatosis ($p<0.001$)
MRI optical analysis					
Gallego-Duran 2016[24]	Prospective	126 patients (estimatio	Kleiner 2005. Ballooning and inflammation.	NASHMRI score: -estimation cohort:	Estimator E3 (harmonic mean) from T2 Fast

	NASHMRI score obtained from most predicting estimators	n cohort n=39 and validation cohort n=87); 65(51%) with NASH		AUROC=0.88. Best cut-off>0.5: sensitivity=87%, specificity=74%, PPV=80%, NPV=82% -validation cohort: AUROC=0.83. Cut-off>0.5: sensitivity=87%, specificity=60%, PPV=71% and NPV=81%.	Spin Echo, estimator E57 (second order contrast) from dynamic T1, and estimator E73 (averaged mean curvature) from T2 with fat suppression, were independently associated with NASH (p=0.015). NASHMRI score to obtain the probability of NASH based on these estimators.
Gadoxetic Acid-enhanced MRI - Relative Enhancement in hepatobiliary phase					
Bastati 2014[43]	Retrospective Relative Enhancement (RE) in hepatobiliary phase	81 patients; 35(43%) with NASH	NASH for activity≥2 and steatosis≥1 with any fibrosis	AUROC=0.85 Cut-off≤1.24: sensitivity=97%; specificity=63%	RE significantly lower in NASH vs simple steatosis (p<0.001), negatively correlated with ballooning (r=-0.44, p<0.0001) and lobular inflammation (r=-0.59, p<0.0001). Multivariable analysis: after correction for steatosis and fibrosis, lobular inflammation and ballooning remained independently associated with RE (p=0.002 and 0.04).
Feier 2013[s28]	Retrospective Relative Enhancement (RE) in hepatobiliary phase	102 patients with chronic liver disease, 17(17%) with NASH	Not clear NASH definition		RE moderately associated with necro-inflammatory activity (p=0.002); in multivariable model, necro-inflammatory activity was not independently

					associated with RE after correction for fibrosis.
Wu 2013[s29]	Retrospective Relative Enhancement (RE) in hepatobiliary phase	25 patients; 12(48%) with NASH	Brunt 1999 criteria; no further specified NASH definition		RE significantly lower in NASH vs simple steatosis (p=0.03). No significant association between RE and necro-inflammatory grade (p=0.09).
SPIO/USPIO-enhanced MRI					
Smits 2015[44] (USPIO)	Prospective Difference (Δ) in R2* between contrast-enhanced and baseline	24 patients (6 simple steatosis patients not biopsy-proven) 13(54%) with NASH	NAS≥5 when steatosis, inflammation and ballooning present	AUROC=0.87 Cut-off<45.5 sec^{-1}: sensitivity=77%; specificity=91%. Cut-off<58.3 sec^{-1}: sensitivity=85%; specificity=73%.	mean hepatic ΔR2* lower in NASH than simple steatosis (p=0.006). Intraobserver reproducibility: ICC (95% CI) = 0.96 (0.93-0.98). Interobserver reproducibility: ICC (95% CI)= 0.88 (0.75-0.94).
Tomita 2008[45] (SPIO)	Prospective Relative decrease in T2 (%T2) and time constant (T)	19 patients; 10(53%) with NASH	NAS≥5	T: AUROC=0.79 Cut-off=42.8: specificity=67%, sensitivity=100%, PPV=77%, NPV=100%. %T2: AUROC=0.83 Cut-off=32.5: specificity=73%, sensitivity=88%, PPV=70%, NPV=89%.	NAS correlated with T (r= 0.66, p= 0.002) and %T2 (r= -0.58, p= 0.009). in pts with definite NASH vs not definite NASH T was significantly higher (p= 0.028) and %T2 significantly lower (p= 0.009)
Asanuma 2010[s30] (SPIO)	Prospective liver relative signal enhancement (RSE) calculated as the ratio between signal intensities pre and post contrast media administration	20 patients, 13(65%) with NASH	NAS≥5		SPIO-MRI RSE was not associated with NAS (r=0.0682, p=0.8072)
Tonan 2012[s31] (SPIO)	Retrospective Percent SPIO-related reduction in liver-to-muscle signal intensity ratio (reduction-%LMR)	32 patients; 24(75%) with NASH	Matteoni 1998. NASH for type 3 and 4 (fat accumulation and ballooning		significant difference in reduction-%LMR

			± Mallory hyaline or fibrosis)		between simple steatosis and NASH (p<0.001) and between necroinflammation 0 vs 2 (p<0.001); 0 vs 3 (p<0.05); and 1 vs 2 (p<0.05); reduction-%LMR decreased with increasing necroinflammation grade (p<0.001). Reproducibility (Bland–Altman analysis): mean differences 3.5% and standard deviation 10%. No correlation between the average and the difference in the reduction-%LMR calculated by two readers (r=-0.11, p=0.553). No proportional bias or fixed bias in the Bland–Altman plot.
Gadolinium-enhanced MRI					
Elias 2009[s32]	Retrospective Early patchy liver enhancement	30 patients; NASH prevalence not reported	Brunt 1999 criteria; no further specified NASH definition		No association between early patchy enhancement and necroinflammatory activity grade (p=0.28).
MRI Liver Volume					
Dillman 2018[41]	Retrospective Liver Volume	69 children and young adults ≤21 years old; 37(54%) with NASH	NAS≥5	AUC=0.741	MRI liver volume significantly associated with portal inflammation (ρ = 0.25; p = 0.04), and NAS (ρ = 0.35; p = 0.003); significantly different (p=0.0005) among simple steatosis, NASH

					without fibrosis, NASH with fibrosis. In logistic regression, liver volume was the only significant NASH MRI predictor (OR= 1.001; 95% CI=1.000–1.002; p=0.007)
MRI preperitoneal fat area					
Parente 2018[42]	Prospective Preperitoneal fat area (cm^2)	66 T2DM patients; 23(35%) with NASH	Steatosis, ballooning and lobular Inflammation	Cut-off=5: sensitivity=93%; specificity=55%	Preperitoneal fat area significantly higher in NASH vs simple steatosis (p=0.002). All abdominal fat areas were good predictors of NASH, but the preperitoneal fat area best explained model variability (R^2 = 0.379). ICCs for all fat compartment areas excellent (> 0.85).
Fat Fraction (FF) and Proton Density Fat Fraction (PDFF)					
Chen 2011[23] (two-point Dixon method)	Retrospective Relative fat fraction (RFF)	58 patients 36(72%) with NASH/fibrosis (7 inflammation and 29 fibrosis)	Brunt 1999. Patients categorized as simple steatosis, inflammation without fibrosis, and NAFLD with fibrosis, the latter two classified as NASH.		RFF not significantly different between inflammation/fibrosis and simple steatosis groups. When RFF is the response variable while MRE liver stiffness (LS), inflammation grade and fibrosis stage are the effect variables, partial correlation coefficients are 0.095 (p=0.52) for LS, 0.334 (p=0.132) for grade, 0.114 (p=0.960) for stage.

Cortez-Pinto 1999[s9] (^{1}H-MRS)	Prospective ^{1}H-MRS liver fat-water ratio	8 NASH and 7 healthy controls	Not clear NASH definition		^{1}H-MRS liver fat-water ratio significantly higher in NASH than in controls (p=0.009), with some overlap.
Dillman 2018[41] (chemical shift-encoded imaging or multiecho PDFF)	Retrospective MRI- liver fat fraction	69 children and young adults ≤21 years old; 37(54%) with NASH	NAS≥5		MRI liver fat fraction significantly associated with hepatocellular ballooning (ρ=0.31; p=0.02), and NAS (ρ=0.52; p<0.0001); significantly different (p=0.004) among categories (simple steatosis, NASH without fibrosis, NASH with fibrosis)
Imajo 2016[18] (MRI-PDFF)	Prospective MRI-PDFF	142 patients; 108(76%) with NASH	Steatosis, inflammation, ballooning and pericellular/perisinusoidal fibrosis		PDFF not associated with inflammation
Kalra 2009[s15] (two-point Dixon method)	Prospective MRI signal intensities in in and out of phase with relative signal loss respect to kidney	10 patients, 5(50%) with NASH	Matteoni 1998. NASH for type 3 and 4 (fat accumulation and ballooning ± Mallory hyaline or fibrosis)		MRI signal intensities not significantly different between NASH and simple steatosis patients.
Leporq 2017[39] (MRI fat fraction)	Retrospective Fat fraction	32 patients; 20(62.5%) with NASH	Steatosis, ballooning, lobular inflammation		Fat fraction not significantly different between NASH and simple steatosis
Saadeh 2002[s14] (two-point Dixon method)	Prospective MR signal loss in out of phase compared to in phase (with spleen as the reference)	25 patients, 17(68%) with NASH	Steatosis, inflammation and ballooning		MR signal loss did not distinguish NASH nor the presence of ballooning, Mallory's hyaline, or fibrosis. Intraobserver and interobserver agreement was

					excellent (ICC from 0.69 to 0.88).
Traussnigg 2017[s10] (^{1}H-MRS)	Prospective ^{1}H-MRS PDFF	30 patients; 22(73%) with NASH	Steatosis, ballooning, lobular inflammation		^{1}H-MRS PDFF associated with lobular inflammation (r=0.42, p=0.021).
Valkovic 2014[s11] (^{1}H-MRS)	Prospective hepatic fat content (%fat)	16 patients, 10(63%) with NASH, 9 healthy controls	Brunt 1999 criteria; no further specified NASH definition		%fat significantly higher in NASH than in simple steatosis patients (p<0.01) and controls (p<0.01).
Wu 2013[s29] (two-point Dixon method)	Retrospective Fat fraction	25 patients; 12(48%) with NASH	Brunt 1999 criteria; no further specified NASH definition		Fat fraction not significantly different between simple steatosis and NASH (p=0.24).

Supplementary Table 7: Result synthesis for MR non-elastographic techniques including also secondary outcomes and studies not exploring accuracy. AUROC: area under the receiver operating characteristic curve. T2DM: Type 2 Diabetes Mellitus; NAS: NAFLD Activity Score; ICC, intraclass correlation. † Reproducibility data derived from another study published by the same research group and conducted on a mixed population (not only NAFLD patients).[s33]

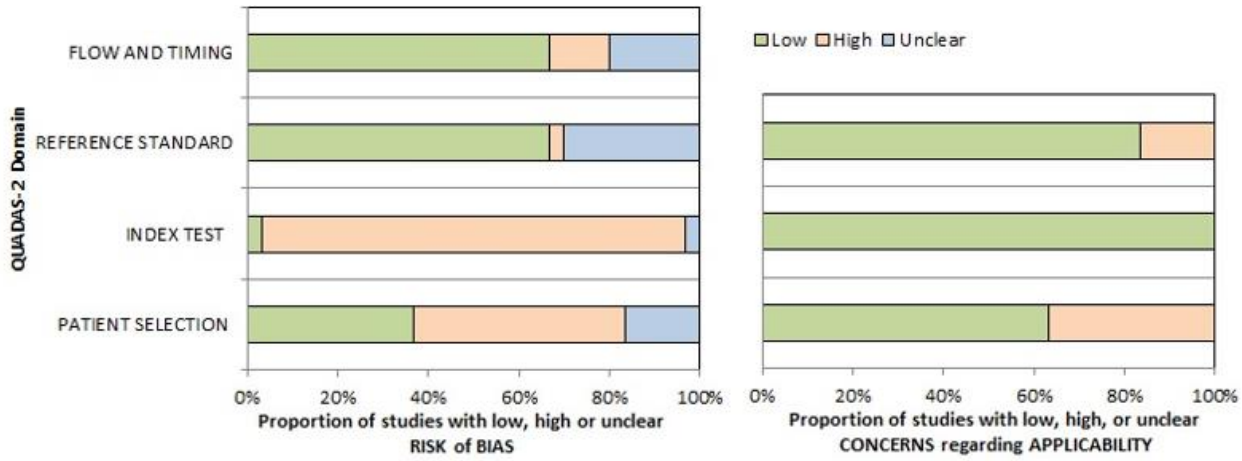
Other techniques					
Study	**Study design and index test definition**	**Population and NASH prevalence**	**Definition of NASH**	**Accuracy SS vs NASH**	**Associations and Reproducibility**
Computed Tomography (CT)					
Naganawa 2018[25]	Retrospective Non-Contrast-Enhanced CT texture features; logistic models for NASH from the most predictive features	88 patients (learning dataset=53 patients and validation dataset=35 patients). Prevalence of NASH not reported.	NAS≥3	Patients without high suspicion of fibrosis: NASH model based on mean0 and skewness2, with cut-off=0.45: AUROC=0.93 and 0.94 in learning and validation datasets; accuracy=94%, specificity=92%, sensitivity=100%, PPV=100%, NPV=80%. Patients with high suspicion of fibrosis: NASH model based on mean0 and kurtosis4, with cut-off=0.81: AUROC=0.81 and 0.60 in learning and validation datasets, accuracy=42%, specificity=31%, sensitivity=100%, PPV=100%, NPV=21%.	
Ataseven 2005[s16]	Prospective Liver CT density, liver to spleen CT density ratio	22 NASH patients and 20 healthy controls	Steatosis, inflammation and ballooning		Mean liver density decreased in NASH than controls ($p<0.05$), liver/spleen ratios decreased in NASH than controls ($p<0.05$) Mean liver density and liver/spleen density ratio significantly associated with histopathological grade ($r=-0.716$ $p<0.01$; $r=-0.633$ $p=0.001$)
Oliva 2006[s17]	Retrospective CT liver-to-spleen attenuation ratio, pattern	21 patients, 12(57%) with NASH	Brunt 1999, NASH when necroinflammatory activity was present		No significant differences in liver-to-spleen ratios, caudate-to-right-lobe ratios, preportal space,

	of steatosis, cranio-caudal liver span, caudate-to-right lobe ratio, preportal space distance, presence of porta hepatis lymph nodes and ascites				presence of porta hepatic lymph nodes between different degrees of necroinflammatory activity Significant differences in the cranio-caudal liver span ($p<0.05$), and the caudate-to-right-lobe ratio ($p<0.05$) between NASH and simple steatosis. No significant differences in liver-to-spleen attenuation ratios, preportal space, or the presence of porta hepatic lymph nodes.
Saadeh 2002[s14]	Prospective CT liver density and liver to spleen ratio, CT qualitative steatosis grade and pattern	25 patients, 17(68%) with NASH	Steatosis, inflammation and ballooning		CT features did not distinguish NASH nor the presence of ballooning, Mallory's hyaline, or fibrosis. Intraobserver agreement for CT pattern of steatosis moderate (ƙ=0.51; 95%CI 0.22–0.81); for CT grade of steatosis substantial (ƙ=0.61; 95%CI 0.35–0.88). Interobserver agreement for pattern and severity of steatosis only fair to moderate (ƙ from 0.40 to 0.43).
Liver Scintigraphy					
Kikuchi 2009[46] (Tc99m-phytate colloid scintigraphy)	Prospective Liver to spleen uptake ratio (L/S)	37 patients; 29(78%) with definite NASH.	Kleiner 2005. Definite NASH for NAS≥5 (no patient with borderline NASH)	AUC=0.82 Cut-off value=2.93: specificity=75%, sensitivity=100%, PPV=94%, NPV=100%	L/S significantly decreased in definite NASH vs simple steatosis patients ($p<0.01$). L/S independently predicted definite NASH (OR=0.078; 95% CI: 0.008-0.739; p=0.026) after correction for age, BMI, ferritin, diabetes and hypertension.
Duman 2006[s18] (Tc99m-phytate colloid scintigraphy)	Prospective Hepatic perfusion time (normal >6 s), blood pool clearance time (time to 50% of the peak activity,	22 NASH patients	NASH when steatosis (>10% of hepatocytes), lobular inflammation and ballooning degeneration, Mallory hyaline fibrosis or sinusoidal fibrosis.		In NASH patients: altered right/left lobe ratio, splenic shift of the tracer, and prolonged blood pool clearance time. No significant correlations between scintigraphy parameters

	normal <10 s), liver right/left lobe (normal >2.5), liver/spleen (normal >1.4) and liver/bone marrow radiocolloid uptake ratios.				and histological findings of inflammation
Masuda 2012[s19] (99mTc-MIBI liver scintigraphy)	Prospective. Uptake ratio of the liver to heart (liver/heart ratio)	26 patients, 11(42%) with borderline NASH, 11(42%) with definite NASH	Kleiner 2005. NAS≥5 for definite NASH; NAS=3-4 for borderline NASH.		Intrahepatic uptake of 99mTc-MIBI remarkably lower in NASH than in simple steatosis. Liver/heart ratio significantly lower in patients with NASH ($p<0.01$) and borderline NASH ($p<0.05$) than in patients with simple steatosis. Liver/heart ratio correlated with NAS ($r=-0.413$, $p<0.05$).

Supplementary Table 8: Result synthesis for techniques other than elastography, US and MR, including also secondary outcomes and studies not exploring accuracy. AUROC: area under the receiver operating characteristic curve. NAS: NAFLD Activity Score.

4. Supplementary Figure 1



Supplementary Fig. 1: Summary results of risk of the bias analysis performed by the QUADAS-2 tool for studies considering accuracy as an outcome.

5. Supplementary References

s1. Osaki A, Kubota T, Suda T, et al. Shear wave velocity is a useful marker for managing nonalcoholic steatohepatitis. World J Gastroenterol. 2010 Jun 21;16(23):2918-25.

s2. Palmeri ML, Wang MH, Rouze NC, et al. Noninvasive evaluation of hepatic fibrosis using acoustic

radiation force-based shear stiffness in patients with nonalcoholic fatty liver disease. J Hepatol. 2011 Sep;55(3):666-672.

s3. Praveenraj P, Gomes RM, Basuraju S, et al. Preliminary Evaluation of Acoustic Radiation Force Impulse Shear Wave Imaging to Detect Hepatic Fibrosis in Morbidly Obese Patients Before Bariatric Surgery. J Laparoendosc Adv Surg Tech A. 2016 Mar;26(3):192-5.

s4. Cassinotto C, Boursier J, de Lédinghen V, et al. Liver stiffness in nonalcoholic fatty liver disease: A comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy.

Hepatology. 2016 Jun;63(6):1817-27.

s5. Attia D, Bantel H, Lenzen H, et al. Liver stiffness measurement using acoustic radiation force impulse elastography in overweight and obese patients. Aliment Pharmacol Ther. 2016 Aug;44(4):366-79.

s6. Lupsor M, Badea R, Stefanescu H, et al. Performance of unidimensional transient elastography in staging non-alcoholic steatohepatitis. J Gastrointestin Liver Dis. 2010 Mar;19(1):53-60.

s7. Yoneda M, Yoneda M, Mawatari H, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with nonalcoholic fatty liver disease (NAFLD). Dig Liver Dis. 2008 May;40(5):371-8.

s8. Gonçalves Dos Reis Monteiro ML, Ferreira De Almeida E Borges V, Machado De Alcântara T, et al. Liver hemodynamic patterns in nonalcoholic steatosis: Doppler ultrasonography and histological evaluation. Minerva Gastroenterol Dietol. 2016 Mar;62(1):19-29.

s9. Cortez-Pinto H, Chatham J, Chacko VP, et al. Alterations in liver ATP homeostasis in human nonalcoholic steatohepatitis: a pilot study. JAMA. 1999 Nov 3;282(17):1659-64.

s10. Traussnigg S, Kienbacher C, Gajdošík M, et al. Ultra-high-field magnetic resonance spectroscopy in non-alcoholic fatty liver disease: Novel mechanistic and diagnostic insights of energy metabolism in non-alcoholic steatohepatitis and advanced fibrosis. Liver Int. 2017 Oct;37(10):1544-1553.

s11. Valkovič L, Gajdošík M, Traussnigg S, et al. Application of localized ^{31}P MRS saturation transfer at 7 T for measurement of ATP metabolism in the liver: reproducibility and initial clinical application in patients with non-alcoholic fatty liver disease. Eur Radiol. 2014 Jul;24(7):1602-9.

s12. Manning P, Murphy P, Wang K, et al. Liver histology and diffusion-weighted MRI in children with nonalcoholic fatty liver disease: A MAGNET study. J Magn Reson Imaging. 2017 Oct;46(4):1149-1158.

s13. Murphy P, Hooker J, Ang B, et al. Associations between histologic features of nonalcoholic fatty liver disease (NAFLD) and quantitative diffusion-weighted MRI measurements in adults. J Magn Reson Imaging. 2015 Jun;41(6):1629-38.

s14. Saadeh S, Younossi ZM, Remer EM, et al. The utility of radiological imaging in nonalcoholic fatty liver disease. Gastroenterology. 2002 Sep;123(3):745-50.

s15. Kalra N, Duseja A, Das A, et al. Chemical shift magnetic resonance imaging is helpful in detecting hepatic steatosis but not fibrosis in patients with nonalcoholic fatty liver disease (NAFLD). Ann Hepatol. 2009 Jan-Mar;8(1):21-5.

s16. Ataseven H, Yildirim MH, Yalniz M, et al. The value of ultrasonography and computerized tomography in estimating the histopathological severity of nonalcoholic steatohepatitis. Acta Gastroenterol Belg. 2005 Apr-Jun;68(2):221-5.

s17. Oliva MR, Mortele KJ, Segatto E, et al. Computed tomography features of nonalcoholic steatohepatitis with histopathologic correlation. J Comput Assist Tomogr. 2006 Jan-Feb;30(1):37-43.

s18. Duman DG, Dede F, Akin H, et al. Colloid scintigraphy in non-alcoholic steatohepatitis: a conventional diagnostic method for an emerging disease. Nucl Med Commun. 2006 Apr;27(4):387-93.

s19. Masuda K, Ono M, Fukumoto M, et al. Usefulness of Technetium-99 m-2-methoxy-isobutyl-isonitrile liver scintigraphy for evaluating disease activity of non-alcoholic fatty liver disease. Hepatol Res. 2012 Mar;42(3):273-9.

s20. Lee MS, Bae JM, Joo SK, et al. Prospective comparison among transient elastography, supersonic shear imaging, and ARFI imaging for predicting fibrosis in nonalcoholic fatty liver disease. PLoS One. 2017 Nov 27;12(11):e0188321.

s21. Yilmaz Y, Eren F. A Bayesian approach to an integrated multimodal noninvasive diagnosis of definitive nonalcoholic steatohepatitis in the spectrum of nonalcoholic fatty liver disease. Eur J Gastroenterol Hepatol. 2014 Nov;26(11):1292-5.

s22. Yoneda M, Suzuki K, Kato S, et al. Nonalcoholic fatty liver disease: US-based acoustic radiation force impulse elastography. Radiology. 2010 Aug;256(2):640-7.

s23. Ballestri S, Nascimbeni F, Baldelli E, et al. Ultrasonographic fatty liver indicator detects mild steatosis and correlates with metabolic/histological parameters in various liver diseases. Metabolism. 2017 Jul;72:57-65.

s24. Bril F, Ortiz-Lopez C, Lomonaco R, et al. Clinical value of liver ultrasound for the diagnosis of nonalcoholic fatty liver disease in overweight and obese patients. Liver Int. 2015 Sep;35(9):2139-46.

s25. Moriyasu F, Iijima H, Tsuchiya K, et al. Diagnosis of NASH using delayed parenchymal imaging of contrast ultrasound. Hepatol Res. 2005 Oct;33(2):97-9.

s26. Ballestri S, Nascimbeni F, Romagnoli D, et al. Do ultrasonographic semiquantitative indices predict histological changes in NASH irrespective of steatosis extent? Liver Int. 2015 Oct;35(10):2340-1.

s27. Sevastianova K, Hakkarainen A, Kotronen A, et al. Nonalcoholic fatty liver disease: detection of elevated nicotinamide adenine dinucleotide phosphate with in vivo 3.0-T 31P MR spectroscopy with proton decoupling. Radiology. 2010 Aug;256(2):466-73.

s28. Feier D, Balassy C, Bastati N, et al. Liver fibrosis: histopathologic and biochemical influences on diagnostic efficacy of hepatobiliary contrast-enhanced MR imaging in staging. Radiology. 2013 Nov;269(2):460-8.

s29. Wu Z, Matsui O, Kitao A, et al. Usefulness of Gd-EOB-DTPA-enhanced MR imaging in the evaluation of simple steatosis and nonalcoholic steatohepatitis. J Magn Reason Imaging. 2013 May;37(5):1137-43.

s30. Asanuma T, Ono M, Kubota K, et al. Super paramagnetic iron oxide MRI shows defective Kupffer cell uptake function in non-alcoholic fatty liver disease. Gut. 2010 Feb;59(2):258-66.

s31. Tonan T, Fujimoto K, Qayyum A, et al. CD14 expression and Kupffer cell dysfunction in non-alcoholic steatohepatitis: superparamagnetic iron oxide-magnetic resonance image and pathologic correlation. J Gastroenterol Hepatol. 2012 Apr;27(4):789-96.

s32. Elias J Jr, Altun E, Zacks S, et al. MRI findings in nonalcoholic steatohepatitis: correlation with histopathology and clinical staging. Magn Reson Imaging. 2009 Sep;27(7):976-87.

s33. Banerjee R, Pavlides M, Tunnicliffe EM, et al. Multiparametric magnetic resonance for the non-invasive diagnosis of liver disease. J Hepatol. 2014 Jan;60(1):69-77.

Appendix B

3.2.3. Application of guidelines for the management of nonalcoholic fatty liver disease in three prospective cohorts of HIV-monoinfected patients

Supplemental Table S1. Univariable analyses by presence of metabolic risk factors (diabetes and/or obesity; selected characteristics).

	Obesity and/or diabetes (n=344)	**No obesity and/or diabetes** (n=1270)
Age (years)	52 (47-58)**	48 (43-53)**
Male gender (%) 1168	242 (70.3)	926 (72.9)
Ethnicity (%)		
White/Caucasian	243 (70.6)**	1114 (87.7)**
Black non-Hispanic	78 (22.7)**	108 (8.5)**
Hypertension (%) 546	177 (51.5)**	369 (29.1)**
Active tobacco smoker (%**)**	83 (24.1)	475 (37.4)
MSM (%)	87 (25.3)**	490 (38.6)**
Active IDU (%) 254	44 (12.8)	210 (16.5)
HIV duration (years)	16.2 (8.9-21.3)	15.8 (8-21.8)
CD4 cell count (cells/μL)	617 (439-803)	638 (451-805)

Undetectable HIV viral load (≤50 copies) (%)	252 (73.3)*	1021 (80.4)*
Current ART regimen (%)		
NRTIs	305 (88.7)	1069 (84.2)
NNRTIs	163 (47.4)	549 (43.2)
PIs	201 (58.4)*	656 (51.7)*
Integrase inhibitors	130 (37.8)	449 (35.4)
Platelets (10^9/L)	203 (161-249)*	219 (177-260)*
Total cholesterol (mmol/L)	4.6 (3.9-5.4)*	4.8 (4.1-5.5)*
Triglycerides (mmol/L)	1.5 (1.1-2.2)	1.4 (1.0-2.2)
ALT (IU/L)	27 (19-41)	25 (18-37)
AST (IU/L)	25 (19-33)	23 (19-29)
NAFLD (%)	165 (48.0)**	376 (29.6)**
FIB-4 >2.67 (%)	30 (8.7)	77 (6.1)
Risk of progressive liver disease (%)	103 (29.9)*	295 (23.2)*

Legend: Risk of progressive liver disease was defined as elevated ALT or NAFLD with significant liver fibrosis. Continuous variables are expressed as median (IQR) and categorical variables as number (%). *p <0.05; **p <0.001. The p-values refer to t test or χ^2 test between patients with obesity and/or diabetes and those without neither of the two conditions.

Abbreviations: ALT, alanine aminotransferase; ART, antiretroviral therapy; AST, aspartate aminotransferase; BMI, body mass index; FIB-4, fibrosis-4 score; HIV, human immunodeficiency virus; IDU, injection drug

use; IU, international units; MSM, men having sex with men; NAFLD, non-alcoholic fatty liver disease; NNRTIs, non-nucleoside reverse transcriptase inhibitors; NRTIs, nucleoside reverse transcriptase inhibitors; PIs, Protease Inhibitors.

Appendix C

3.2.2 Baseline liver steatosis has no impact on liver metastases and overall survival in rectal cancer patients

Additional files

Supplementary Table S1: Clinical data, CT-defined liver steatosis, and outcomes (metastases, deaths) in patients subdivided according to stage

Table S1		**Stage II (n=77)**	**Stage III (n=125)**	**Stage IV (n=72)**	**Stage XM0 (n= 9)**
Age; mean (SD)		70.56 y (13.22)	65.40 y (14.07)	67.51 y (13.94)	80.00 y (8.57)
Sex; n (%)	Male	49 (63.64%)	68 (54.40%)	44 (61.11%)	6 (66.67%)
	Female	28 (36.36%)	57 (45.60%)	28 (38.89%)	3 (33.33%)
Median follow-up (IQR) for Overall Survival		56 (32-76)	53 (30-78)	16 (7-32.5)	7 (1-39)
T Stage; n (%)	1	0 (0.00%)	3 (2.40%)	0 (0.00%)	0 (0.00%)
	2	0 (0.00%)	16 (12.80%)	0 (0.00%)	0 (0.00%)
	3	65 (84.42%)	80 (64.00%)	0 (0.00%)	0 (0.00%)
	4	12 (15.58%)	26 (20.80%)	4 (5.56%)	0 (0.00%)
	X	0 (0.00%)	0 (0.00%)	0 (0.00%)	9 (100.00%)
	Missing	0 (0.00%)	0 (0.00%)	68 (94.44%)	0 (0.00%)
N Stage; n (%)	0	77 (100.0%)	6 (4.80%)	0 (0.00%)	0 (0.00%)
	1	0 (0.00%)	88 (70.40%)	1 (1.39%)	0 (0.00%)
	2	0 (0.00%)	31 (24.80%)	2 (2.78%)	0 (0.00%)
	X	0 (0.00%)	0 (0.00%)	1 (1.39%)	9 (100.00%)
	Missing	0 (0.00%)	0 (0.00%)	68 (94.44%)	0 (0.00%)
Grade; n (%)	Well differentiated	0 (0.00%)	1 (0.80%)	2 (2.78%)	0 (0.00%)
	Moderately differentiated	38 (49.35%)	48 (38.40%)	14 (19.44%)	1 (11.11%)
	Poorly differentiated	27 (35.06%)	50 (40.00%)	28 (38.89%)	3 (33.33%)
	Missing	12 (15.58%)	26 (20.80%)	28 (38.89%)	5 (55.56%)
Therapy:					
Neoadjuvant chemotherapy; n (%)		33 (42.86%)	73 (58.40%)	3 (4.17%)	0 (0.00%)

Neoadjuvant radiotherapy; n (%)	44 (57.14%)	88 (70.40%)	1 (1.39%)	0 (0.00%)
Surgery; n (%)	62 (80.52%)	105 (84.00%)	14 (19.44%)	0 (0.00%)
Adjuvant chemotherapy; n (%)	37 (48.05%)	70 (56.00%)	8 (11.11%)	0 (0.00%)
Adjuvant radiotherapy; n (%)	3 (3.90%)	8 (6.40%)	3 (4.17%)	0 (0.00%)
Palliative chemotherapy; n (%)	0 (0.00%)	2 (1.60%)	36 (50.00%)	0 (0.00%)
Palliative radiotherapy; n (%)	9 (11.69%)	10 (8.00%)	9 (12.50%)	1 (11.11%)
No therapy; n (%)	3 (3.90%)	6 (4.80%)	19 (26.39%)	7 (77.78%)
Missing; n (%)	2 (2.60%)	0 (0.00%)	2 (2.78%)	1 (11.11%)
Liver metastases:				
Synchronous; n (%)	-	-	42 (58.33%)	-
Metachronous; n (%)	6 (7.79%)	15 (12.00%)	5 (16.67%) *	0 (0.00%)
All-sites metastases:				
Synchronous; n (%)	-	-	72 (100.00%)	-
Metachronous; n (%)	24 (31.17 %)	37 (29.60%)	-	0 (0.00%)
Overall deaths; n (%) Other causes; n (%) **	32 (41.56%) 9 (28.13%)	49 (39.20%) 11 (22.45%)	64 (88.89%) 6 (9.38%);	7 (77.78%) 2 (28.57%)
Steatosis at baseline; n (%) Mild-to-moderate; n (%) Severe; n (%)	21 (27.27%) 15 (19.48%) 6 (7.79%)	39 (31.20%) 31 (24.80%) 8 (6.40%)	27 (37.50%) 22 (30.56%) 5 (6.94%)	3 (33.3%) 3 (33.3%) 0 (0.00%)

Stage XM0 was considered for patients with unknown local staging but exclusion of distant metastases with CT scan. TNM stage refers to clinical staging. IQR: Interquartile range. *the percentage was calculated on a total of 30 patients, after excluding patients with baseline liver metastases. ** missing values were 45, 76, 8, and 2 in stage II, III, IV, and XM0, respectively.

Supplementary Table S2: Follow-up and treatment characteristics in patients with and without liver steatosis.

Table S2		**No Liver Steatosis (n= 193; 68.2%)**	**Liver Steatosis (n= 90; 31.8%)**
Overall survival follow-up duration; mean (SD) (months)		43.11 (30.34)	48.97 (35.31)
Liver metastases follow-up duration; mean (SD) (months) (patients without synchronous liver metastases)		36.49 (27.64)	37.70 (27.80)
Neoadjuvant chemotherapy; n (%)		70 (36.3)	39 (43.3)
Neoadjuvant radiotherapy; n (%)		88 (45.6)	45 (50.0)
Surgery; n (%)		119 (61.7)	62 (68.9)
Adjuvant chemotherapy; n (%)		74 (38.3)	41 (45.6)
Adjuvant radiotherapy; n (%)		8 (4.1)	6 (6.7)
Palliative chemotherapy; n (%)		24 (12.4)	14 (15.6)
Palliative radiotherapy; n (%)		24 (12.4)	5 (5.6)
No therapy; n (%)		25 (13.0)	10 (11.1)
Missing; n (%)		4 (2.1)	1 (1.1)
All-sites metastases; n (%)	Synchronous	45 (23.3)	27 (30.0)
	Metachronous*	43 (29.1)	18 (28.6)

SD, standard deviation. * percentages were calculated after excluding stage IV patients (45 in no liver steatosis group and 27 in liver steatosis group).

Supplementary Table S3: Baseline characteristics of patients with moderate/severe steatosis.

Table S3		**Moderate/Severe Liver Steatosis (n= 19)**			
		Overall	**Synchronous liver metastases (n=2)**	**Metachronous liver metastases (n=2)**	**Deaths (n=9)**
Age; mean (SD)		64.7 (15.5)			
Sex; n (%)	Male	12 (63.2)	1	1	5
	Female	7 (36.8)	1	1	4
Stage; n (%)	II	6 (31.6)	-	-	-
	III	8 (42.1)	-	1	5
	IV	5 (26.3)	2	1	4
	XM0	-	-	-	-
Grade; n (%)	Well	-	-	-	-
	Moderately	11 (57.9)	2	1	3
	Poorly	5 (26.3)	-	-	4
	Missing	3 (15.8)	-	1	2

TNM stage refers to clinical staging. SD: Standard deviation.

Supplementary Table S4: CT characteristics and liver function test changes after neoadjuvant chemotherapy, defining the presence or absence of post-chemotherapy liver damage.

Table S4	**Patients without post-chemotherapy liver damage (n= 38, 61.3%)**	**Patients with post-chemotherapy liver damage (n= 24, 38.7%)**
Liver steatosis:		
appearance	0 (0.0%)	5 (20.1%)
Missing	33 (86.8%)	14 (58.3%)
Liver volume:		
Increase >10%	0 (0.00%)	8 (33.3%)
Decrease >10%	0 (0.00%)	3 (12.5%)
Missing	24 (58.5%)	9 (37.5%)
AST:		
Increase > 40 and doubled	0 (0.00%)	8 (33.3%)
Missing	2 (5.3%)	2 (8.3%)
ALT:		
Increase > 49 and doubled	0 (0.0%)	8 (33.3%)
Missing	1 (2.3%)	2 (8.3%)
GGT:		
Increase > 73 and doubled	0 (0.0%)	3 (12.5%)
Missing	8 (21.0%)	4 (16.7%)

Supplementary Table S5: Follow-up, treatment, and outcome measures in patients with and without post-chemotherapy liver damage.

Table S5	**Patients without post-chemotherapy liver damage (n= 38, 61.3%)**	**Patients with post-chemotherapy liver damage (n= 24, 38.7%)**
Follow-up overall survival (months) Median (IQR)	59.5 (38-84)	37 (26-47)
Follow-up liver metastases (months) Median (IQR)	50 (28.5-62)	32 (19-46)
Adjuvant therapy	30 (78.9%)	18 (75%)
Deaths during follow-up	5 (13.2%)	10 (41.7%)
Liver metastases during follow-up	2 (5.3%)	5 (20.1%)
All sites metastases during follow-up	14 (36.8%)	10 (41.7%)

IQR: Interquartile range

Supplementary Table S6: Demographic and clinical characteristics of patients with and without post-chemotherapy-induced liver damage, reported also in patients with events.

		Patients without post-chemotherapy liver damage (n= 38, 61.3%)			**Patients with post-chemotherapy liver damage (n= 24, 38.7%)**		
		Overall	**Metachronous liver metastases (n=2)**	**Deaths (n=5)**	**Overall**	**Metachronous liver metastases (n=5)**	**Deaths (n=10)**
Age at diagnosis; mean (SD)		59.1 (12.9)			60.7 (12.3)		
Sex; n (%)	Male	19 (50%)	2	4	11 (45.8%)	1	5
	Female	19 (50%)	0	1	13 (54.2%)	4	5
Stage; n (%)	II	12 (31.6%)	1	3	8 (33.3%)	1	4
	III	26 (68.4%)	1	2	16 (66.7%)	4	6
Grade; n (%)	Well	0 (0.0%)	0	0	0 (0.0%)	0	0
	Moderately	19 (50.0%)	1	3	10 (41.7%)	1	4
	Poorly	8 (21.1%)	1	1	8 (33.3%)	2	4
	Missing	11 (28.9%)	0	1	6 (25.0%)	2	2

TNM stage refers to clinical staging. SD: Standard deviation.

APPENDIX D

4.1 Prognostic impact of muscle quantity and quality and fat distribution in diffuse large B-cell lymphoma patients: an observational study

Supplementary materials

Supplementary Table 1. PET-CT indices of body composition of the patients at the time of diagnosis

PET-CT indices of body composition	**All patients (n=116)**
L3-SMA (cm^2)	139.9 ± 34.9
L3-SMI (cm^2/m^2)	50.2 ± 10
L3-SMD (HU)	34.7 ± 8.1
% of patients with low L3-SMI: - cut-off values 1 (L3-SMI)[a] - cut-off values 2 (L3-SMI-2)[b] - cut-off values 3 (L3-SMI-3)[c]	 29 (25.0%) 40 (34.5%) 30 (25.9%)
% of patients with low L3-SMD	61 (52.6%)
L3-Total adipose tissue area (cm^2)	317 (189.5-431.0)
L3-SAT (cm^2)	140.5 (90-211.3)
L3-VAT (cm^2)	125 (82.8-207.5)
L3- IMAT (cm^2)	18.5 (10.0-25.3)
VAT/SAT	0.9 (0.6-1.3)
PT-SMI (cm^2/m^2)	78.9 ± 19.1
PT-SMD (HU)	46.2 ± 6.5
PT-IMAT (cm^2)	19 (13-28)

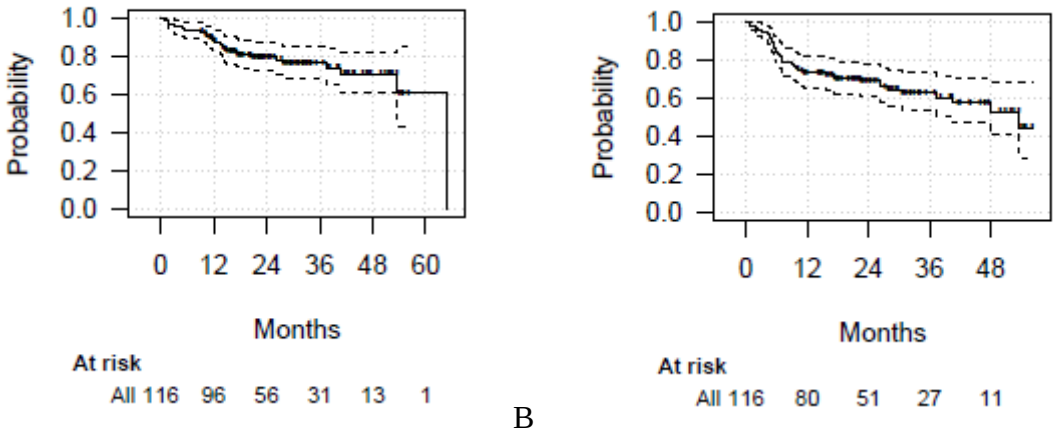
Data are reported as frequencies and percentage for categorical variables, means and standard deviations for normally distributed continuous variables and median and interquartile range for non-normally distributed continuous variables. PET-CT, positron emission tomography-computed tomography; L3, third lumbar vertebra; SMA, skeletal muscle area; PT, proximal thigh; SMI, skeletal muscle index; SMD, skeletal muscle density; VAT, visceral adipose tissue; SAT, subcutaneous adipose tissue; IMAT, inter-muscular adipose tissue. For SMI at the level of the third lumbar vertebra, three different sets of cut-off values were used. [a]L3-SMI: <43 cm^2/m^2 for men with

BMI <25, <53 cm^2/m^2 for men with BMI ≥25, and <41 cm^2/m^2 for women; [b]L3-SMI-2: <55.8 cm^2/m^2 for men and <38.9 cm^2/m^2 for women; [c]L3-SMI-3: <52.4 cm^2/m^2 for men and <38.5 cm^2/m^2 for women.

Supplementary Table 2. Correlations between different PET-CT indices of body composition and correlations between PET-CT indices with serological data

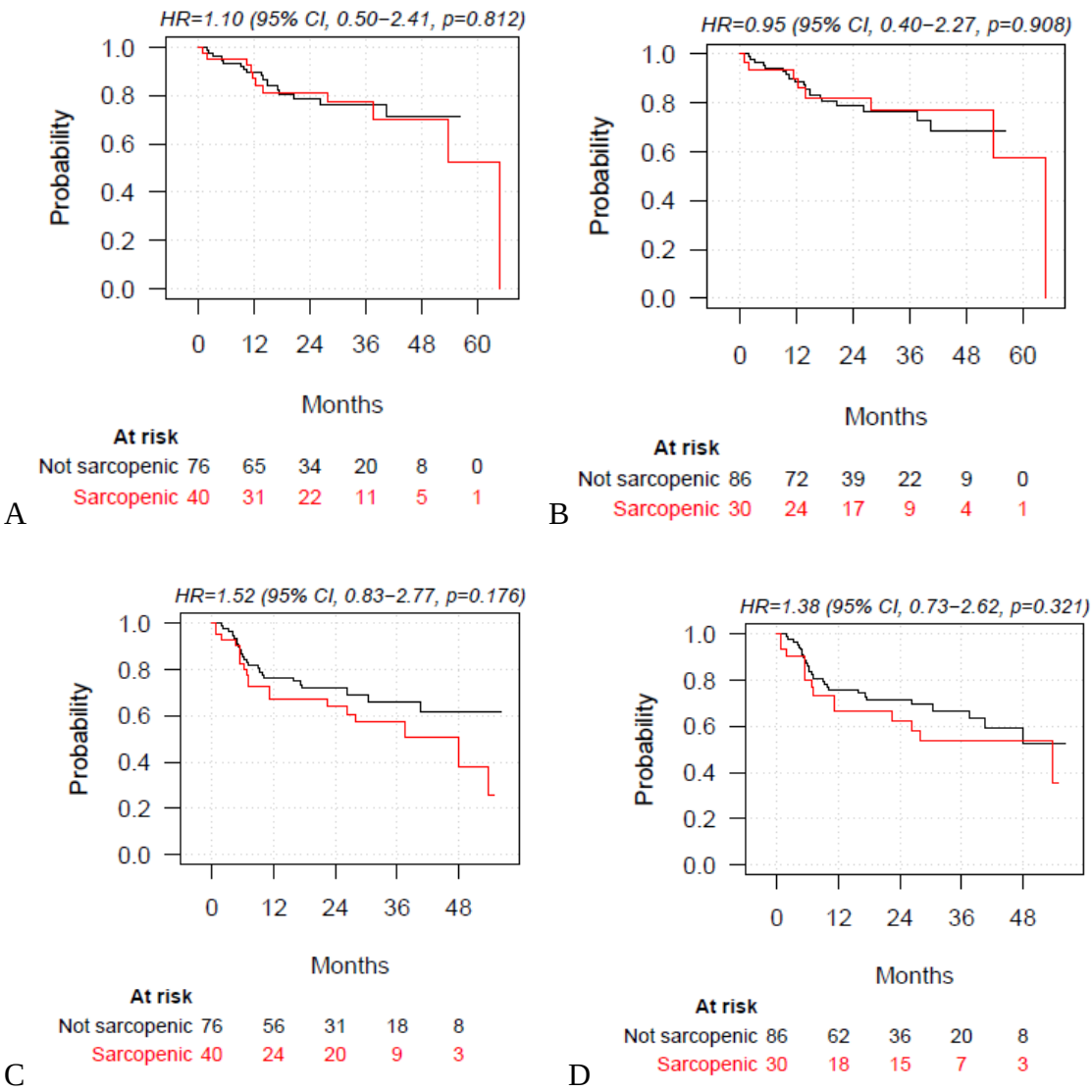
PET-CT indices of body composition	**Correlation coefficient (95%CI)**
L3-SMI and L3-SMD	0.36 (0.19 – 0.51)
L3-SMI and VAT	0.21 (0.03 – 0.38)
L3-SMD and VAT	-0.20 (-0.37 – -0.01)
L3-SMI and PT-SMI	0.65 (0.53 – 0.74)
L3-SMD and PT-SMD	0.86 (0.81 – 0.90)
L3-IMAT and PT-IMAT	0.63 (0.51 – 0.73)
Body-composition PET-CT indices and serological data	
Serum Albumin and L3-SMI	0.20 (0.02 – 0.37)
Serum Protein level and L3-SMI	0.10 (-0.09 – 0.28)
C-reactive protein and L3-SMI	-0.12 (-0.30 – 0.07)
Serum Glucose and L3-SMI	0.04 (-0.15 – 0.22)
Serum Vitamin D and L3-SMI	-0.04 (-0.25 – 0.17)
Serum Albumin and L3-SMD	0.38 (0.22 – 0.53)
Serum Protein level and L3-SMD	0.41 (0.24 – 0.55)
C-reactive protein and L3-SMD	-0.25 (-0.42 – -0.07)
Serum Glucose and L3-SMD	-0.07 (-0.25 – 0.11)
Serum Vitamin D and L3-SMD	0.03 (-0.18 – 0.24)
Serum Albumin and VAT	-0.00 (-0.18 – 0.18)
Serum Protein level and VAT	-0.08 (-0.26 – 0.11)
C-reactive protein and VAT	0.01 (-0.18 – 0.19)
Serum Glucose and VAT	0.35 (0.18 – 0.50)
Serum Vitamin D and VAT	-0.05 (-0.26 – 0.16)

PET-CT, positron emission tomography-computed tomography; CI, confidence interval; L3, third lumbar vertebra; PT, proximal thigh; SMI, skeletal muscle index; SMD, skeletal muscle density; VAT, visceral adipose tissue; IMAT, intramuscular adipose tissue.

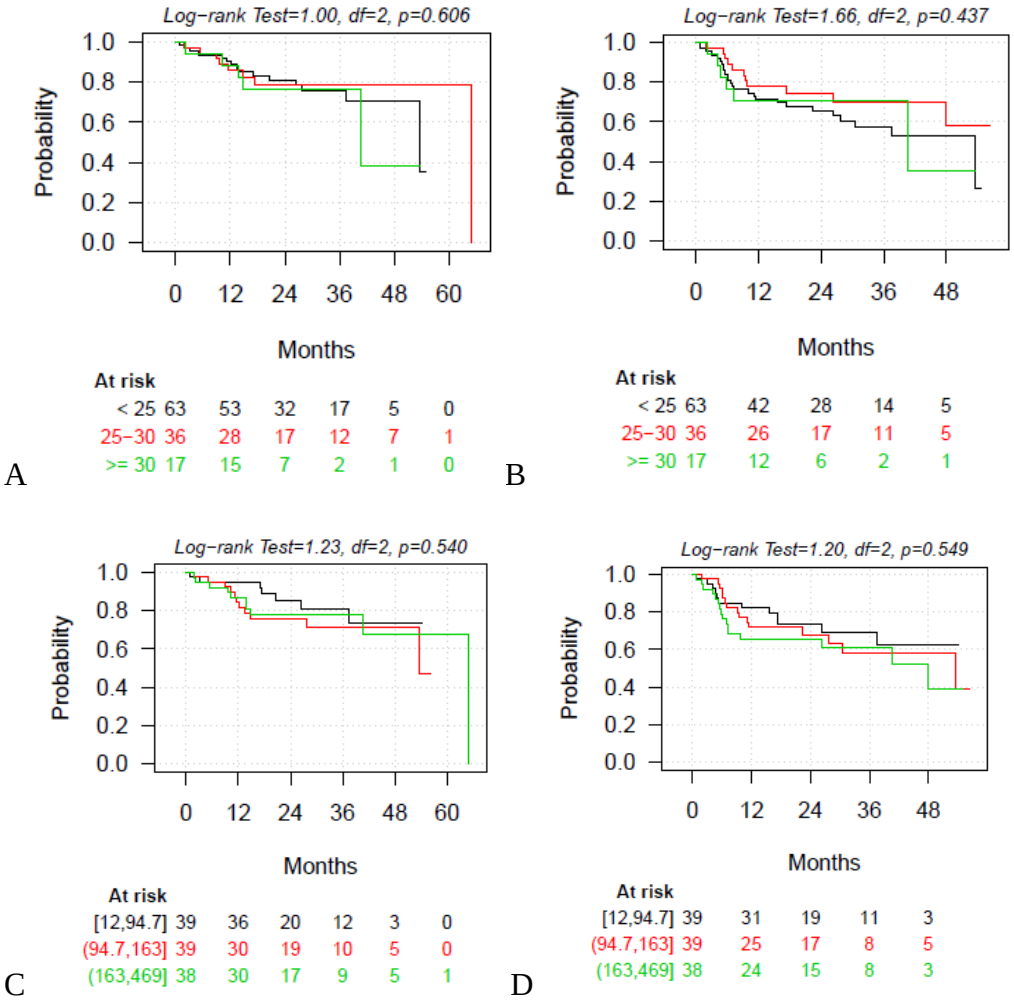


A B

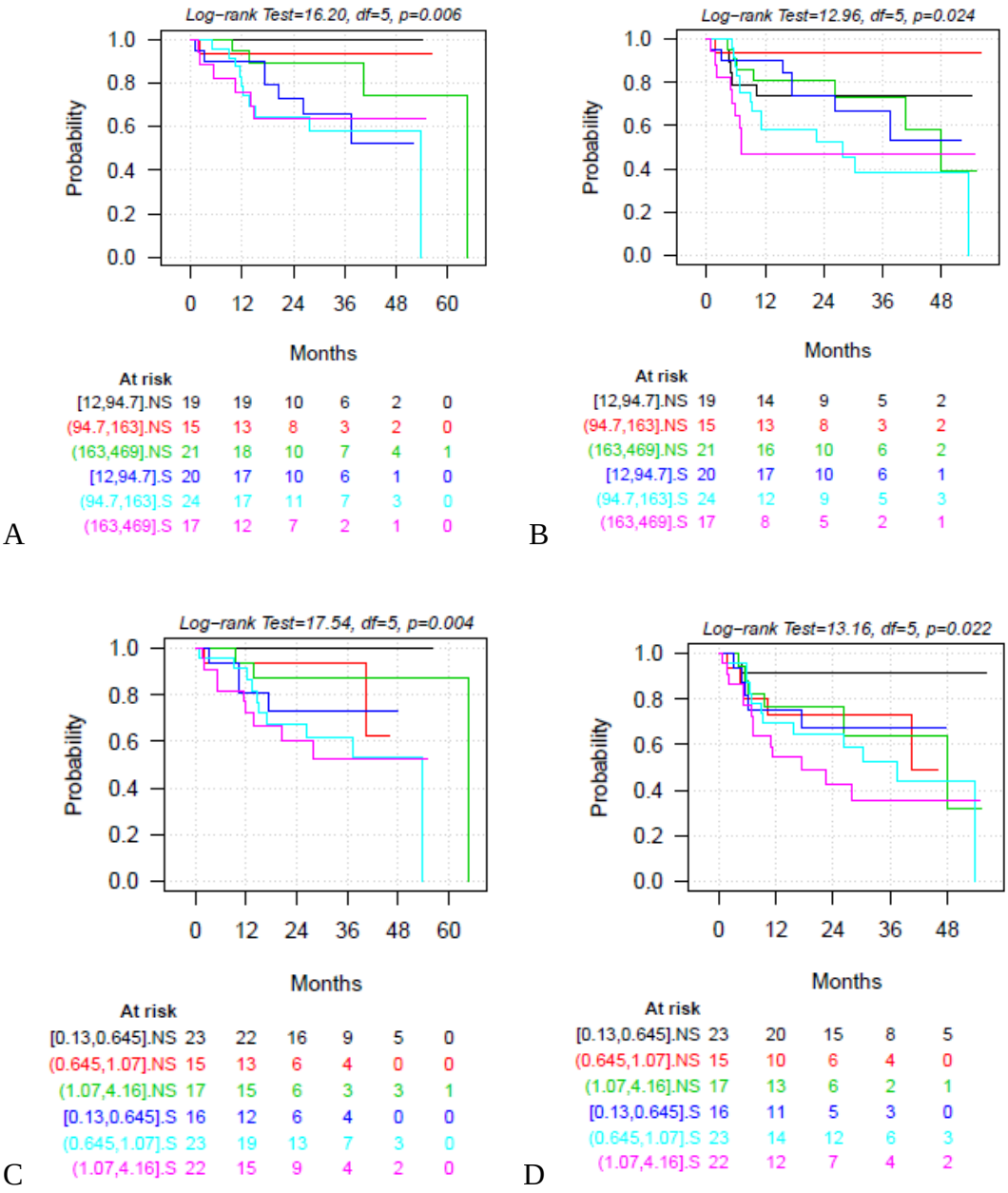
Supplementary Figure 1. Overall survival (**A**) and progression-free survival (**B**) with respective 95% confidence intervals in the whole population



Supplementary Figure 2. Overall survival by muscle depletion (sarcopenic) according to L3-SMI-2 cut-off values (<55.8 cm^2/m^2 for men and <38.9 cm^2/m^2 for women) (**A**), and L3-SMI-3 cut-off values (<52.4 cm^2/m^2 for men and <38.5 cm^2/m^2 for women) (**B**). Progression-free survival by muscle depletion (sarcopenic) according to L3-SMI-2 cut-off values (<55.8 cm^2/m^2 for men and <38.9 cm^2/m^2 for women) (**C**), and to L3-SMI-3 cut-off values (<52.4 cm^2/m^2 for men and <38.5 cm^2/m^2 for women) (**D**). L3, third lumbar vertebra; SMI, skeletal muscle index.



Supplementary Figure 3. Overall survival (**A**) and progression-free survival (**B**) by BMI groups (<25; 25-30; ≥30). Overall survival (**C**) and progression-free survival (**D**) by VAT tertiles (12-94.7; 94.7-163; 163-469). BMI, body mass index; VAT, visceral adipose tissue.



Supplementary Figure 4. Overall survival (**A**) and progression-free survival (**B**) stratified by poor muscle quality (according to L3-SMD cut-off values) and visceral obesity (defined based on VAT tertiles). Overall survival (**C**) and progression-free survival (**D**) stratified by poor muscle quality (according to L3-SMD cut-off values) and visceral obesity (defined based on tertiles of VAT/SAT ratio). L3, third lumbar vertebra; SMD, skeletal muscle density; VAT, visceral adipose tissue; SAT, subcutaneous adipose tissue; S, sarcopenic patients defined according to skeletal muscle density at the level of the third lumbar vertebra (L3-SMD); NS, not sarcopenic patients.

APPENDIX E

4.2 The impact of chest CT body composition parameters on clinical outcomes in COVID-19 patients

Supporting information

	BMI (n=88)	TAT	VAT	SAT	IMAT
BMI	1				
TAT	0.7059 (0.000)	1			
VAT	0.5053 (0.000)	0.5536 (0.000)	1		
SAT	0.6304 (0.000)	0.9590 (0.000)	0.3613 (0.001)	1	
IMAT	0.6119 (0.000)	0.6934 (0.000)	0.6452 (0.000)	0.5357 (0.000)	1

S1 Table. Correlations between BMI and CT fat distribution parameters. Spearman correlations between BMI and CT fat distribution parameters. CT: Computed Tomography; IMAT: intermuscular adipose tissue area; SAT: subcutaneous adipose tissue area; TAT: total adipose tissue area; VAT: visceral adipose tissue area.

Variables	Hospitalization		Mechanical ventilation or death		Death	
	β	95% CI	β	95% CI	β	95% CI
Pectoral area						
I quart [6-12]	0		0		0	
II quart [12.1-17]	0.288	-0.358;0.933	-0.240	-0.895;0.415	-0.014	-0.804;0.775
III quart [17.1-21]	-0.045	-0.710;0.620	-0.112	-0.798;0.573	0.176	-0.635;0.987
IV quart [21.1-50]	-0.316	-0.957;0.325	-0.370	-1.061;0.321	-0.584	-1.506;0.338
Pectoral density						
I quart [3-27]	0		0		0	
II quart [28-34]	-0.591	-1.288;0.105	-0.910	-1.592;-0.229	-0.587	-1.352;0.179
III quart [35-41]	-0.434	-1.135;0.267	-0.551	-1.195;0.094	-0.925	-1.744;-0.107
IV quart [41.1-63]	-1.639	-2.334;-0.945	-1.252	-1.99;-0.514	-1.406	-2.369;-0.443
L/S ratio						
I quart [0.15-0.920]	0		0		0	
II quart [0.920-1.059]	-0.199	-0.913;0.516	-0.055	-0.702;0.593	0.808	-0.002;1.619
III quart [1.059-1.76]	-1.033	-1.711;-0.355	-0.767	-1.462;-0.071	-0.029	-0.930;0.871
IV quart [1.76-2.25]	-0.892	-1.581;-0.203	-0.683	-1.382;0.017	0.149	-0.739;1.036
TAT						
I quart [20- 159]	1		0		0	
II quart [160-223]	0.620	-0.077;1.316	0.507	-0.245;1.260	0.098	-0.770;0.966
III quart [224-292]	0.488	-0.200;1.177	0.157	-0.620;0.934	0.000	-0.882; 0.882
IV quart [293-649]	0.758	0.051;1.464	0.761	0.019;1.502	0.000	-0.882; 0.882
VAT						
I quart [2-23]	0		0		0	
II quart [24-34]	0.831	0.202;1.461	1.619	0.757;2.481	1.517	0.366; 2.668
III quart [35-47]	1.001	0.348;1.653	1.312	0.425;2.199	1.341	0.161;2.520
IV quart [48-118]	1.941	1.187;2.695	2.185	1.329;3.042	2.148	1.031;3.266
IMAT						
I quart [0-18]	0		0		0	
II quart [19-27]	1.502	0.791;2.212	1.378	0.481;2.275	1.341	0.259; 2.423
III quart [28-37]	1.523	0.814;2.232	1.491	0.602;2.381	0.913	-0.216;2.041
IV quart [38-83]	2.100	1.318;2.882	2.140	1.263;3.017	1.891	0.847; 2.935

S2 Table. Preliminary analysis to verify the relationship and linearity between CT body composition parameters and the logit of the outcome. CT: Computed Tomography; IMAT: intermuscular adipose tissue area; L/S: liver to spleen ratio; TAT: total adipose tissue area; VAT: visceral adipose tissue area.

All patients	Hospitalization		Mechanical ventilation or death		Death	
	OR adj	95% CI	OR adj	95% CI	OR adj	95% CI
Pectoral density[a]	0.967	0.935-1.000	0.964	0.934-0.996	0.962	0.922-1.004
TAT[a]	1.005	1.002-1.008	1.005	1.002-1.009	1.002	0.998-1.007
VAT[a]	1.028	1.008-1.049	1.026	1.008-1.043	1.017	0.997-1.038
IMAT						
I quart [0-18]					1	
II quart [19-27]					1.190	0.308-4.588
III quart [28-37]					0.820	0.203-3.312
IV quart [38-83]					1.615	0.431-6.053
IMAT[a]	1.028	1.006-1.050	1.024	1.005-1.043		

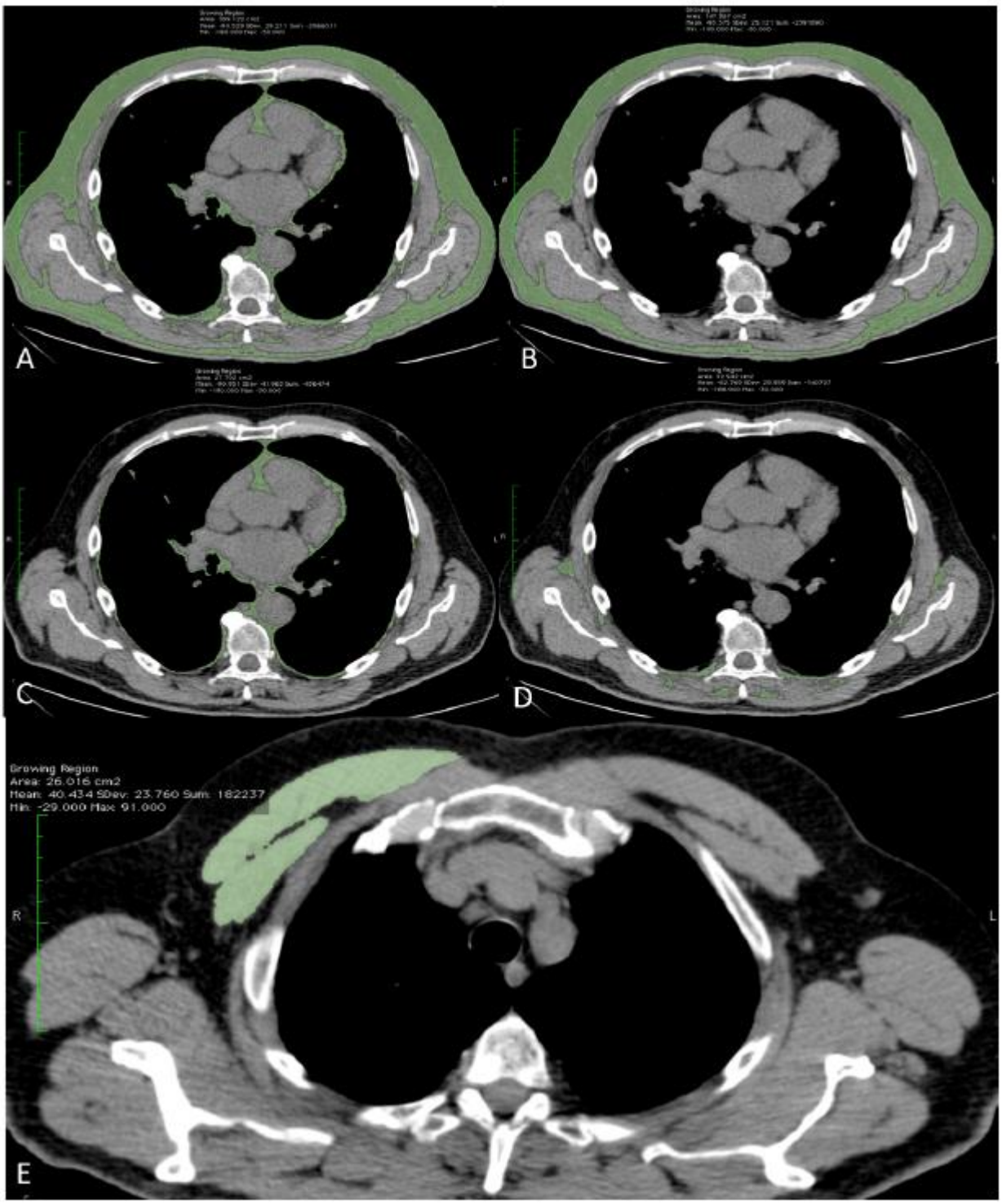
After excluding all patients with comorbidities	Hospitalization		Mechanical Ventilation or death		Death	
	OR adj	95% CI	OR adj	95% CI	OR adj	95% CI
Pectoral density[a]	0.984	0.948-1.021	0.991	.950-1.035	0.998	0.941-1.059
T7-T8 TAT[a]	1.004	1.000-1.008	1.002	0.997-1.006	0.998	0.992-1.005
T7-T8 VAT[a]	1.024	1.002-1.047	1.023	1.001-1.045	1.016	0.988-1.045
IMAT						
I quart [0-18]					1	
II quart [19-27]					1.877	0.333-10.592
III quart [28-37]					0.445	0.062-3.184
IV quart [38-83]					0.982	0.161-5.977
T7-T8 IMAT[a]	1.019	0.996-1.042	1.003	0.979-1.028		

S3 Table. Multivariate logistic model adjusted for sex, age, and calendar period (weeks since the beginning of the outbreak). Hospitalization, ventilation and/or death, and mortality OR with 95% CI are reported for unit increase of CT body composition parameters (HU for pectoral density and cm2 for VAT and TAT) and for IMAT quartiles. The same model is reported after excluding all patients with comorbidities, only patients with diabetes, only patients with cardiovascular comorbidities, including hypertension, and only patients with previous cancer diagnosis. OR adj: adjusted for age, sex, and calendar period. HU: Hounsfield Unit; IMAT: intermuscular adipose tissue area; TAT: total adipose tissue area; VAT: visceral adipose tissue area.

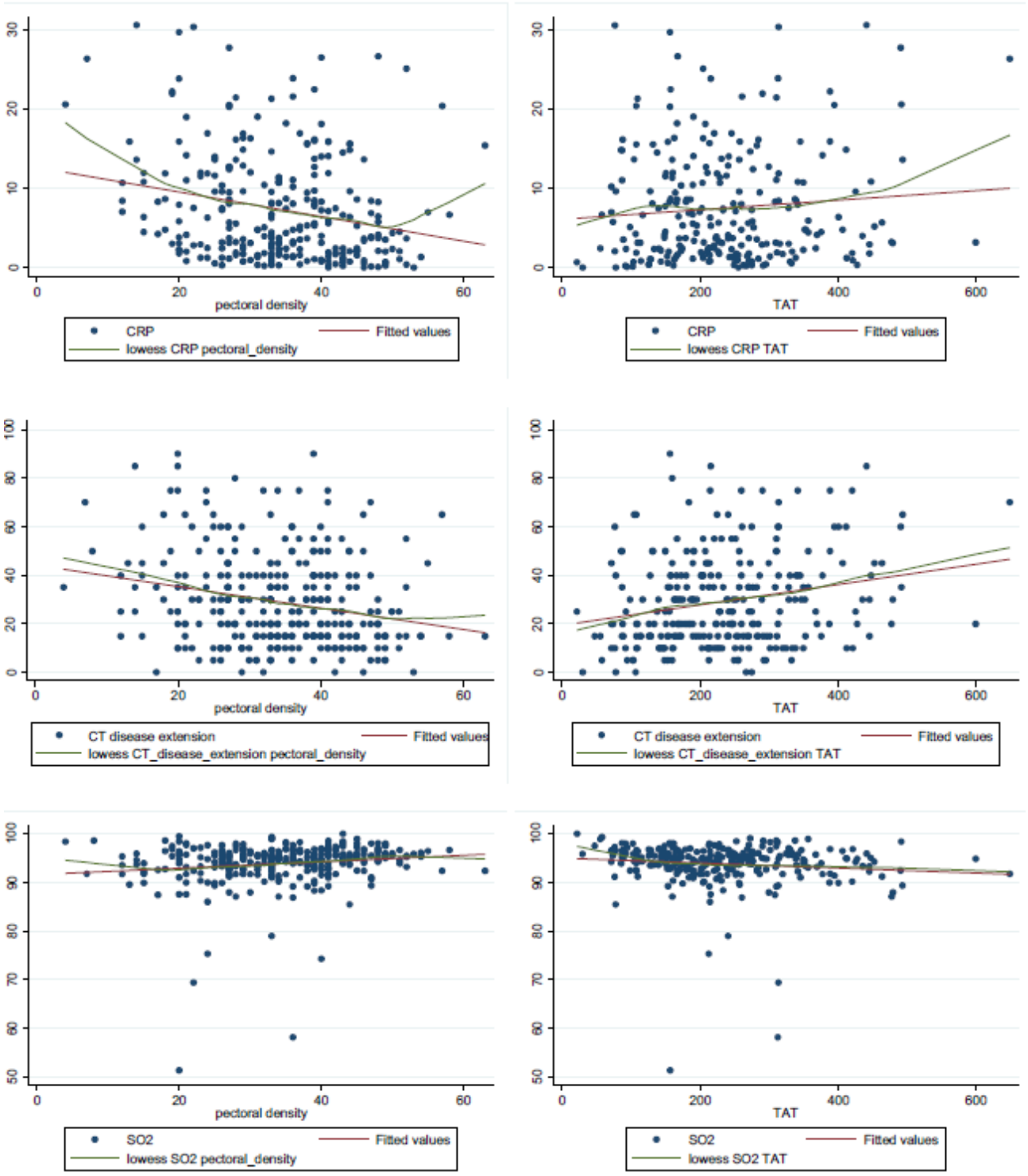
a as continuous variable (for one unit increase).

	Age quartiles						
	Total	[17-53)	[53-66)	[66-76)	[76-96]	P*	Spearman's ρ
Pectoral density [I quart: 3-27]	85 (27.3)	7 (8.8)	24 (30.8)	30 (38.0)	24 (32.4)	0.000	-0.308 (0.000)
[II quart: 28-34]	76 (24.4)	14 (17.5)	18 (23.1)	18 (22.8)	26 (35.1)		
[III quart: 35-41]	78 (25.1)	25 (31.3)	17 (21.8)	18 (22.8)	18 (24.3)		
[IV quart: 41.1-63]	72 (23.2)	34 (42.5)	19 (24.4)	13 (16.5)	6 (8.1)		
TAT [I quart: 20- 159]	68 (25.0)	25 (37.9)	13 (20.0)	11 (15.9)	19 (26.4)	0.017	0.050 (0.413)
[II quart: 160-234]	68 (25.0)	9 (13.6)	15 (23.1)	19 (27.5)	25 (34.7)		
[III quart: 224-292]	68 (25.0)	18 (27.3)	14 (21.5)	20 (29.0)	16 (22.2)		
[IV quart: 293-649]	68 (25.0)	14 (21.2)	23 (35.4)	19 (27.5)	12 (16.7)		
VAT [I quart: 2-23]	83 (26.5)	41 (51.3)	19 (24.4)	11 (14.1)	12 (15.6)	0.000	0.390 (0.000)
[II quart: 24-34]	80 (25.6)	22 (27.5)	20 (25.6)	16 (20.5)	22 (28.6)		
[III quart: 35-47]	74 (23.6)	12 (15.0)	21 (26.9)	26 (33.3)	15 (19.5)		
[IV quart: 48-118]	76 (24.3)	5 (6.3)	18 (23.1)	25 (32.1)	28 (36.4)		
IMAT [I quart: 0-18]	76 (27.6)	41 (62.1)	18 (26.9)	9 (13.0)	8 (11.0)	0.000	0.402 (0.000)
[II quart: 19-27]	66 (24.0)	9 (13.6)	16 (23.9)	20 (29.0)	21 (28.8)		
[III quart: 28-37]	67 (24.4)	10 (15.2)	21 (31.3)	18 (26.1)	18 (24.7)		
[IV quart: 38-83]	66 (24.0)	6 (9.1)	12 (17.9)	23 (31.9)	26 (35.6)		

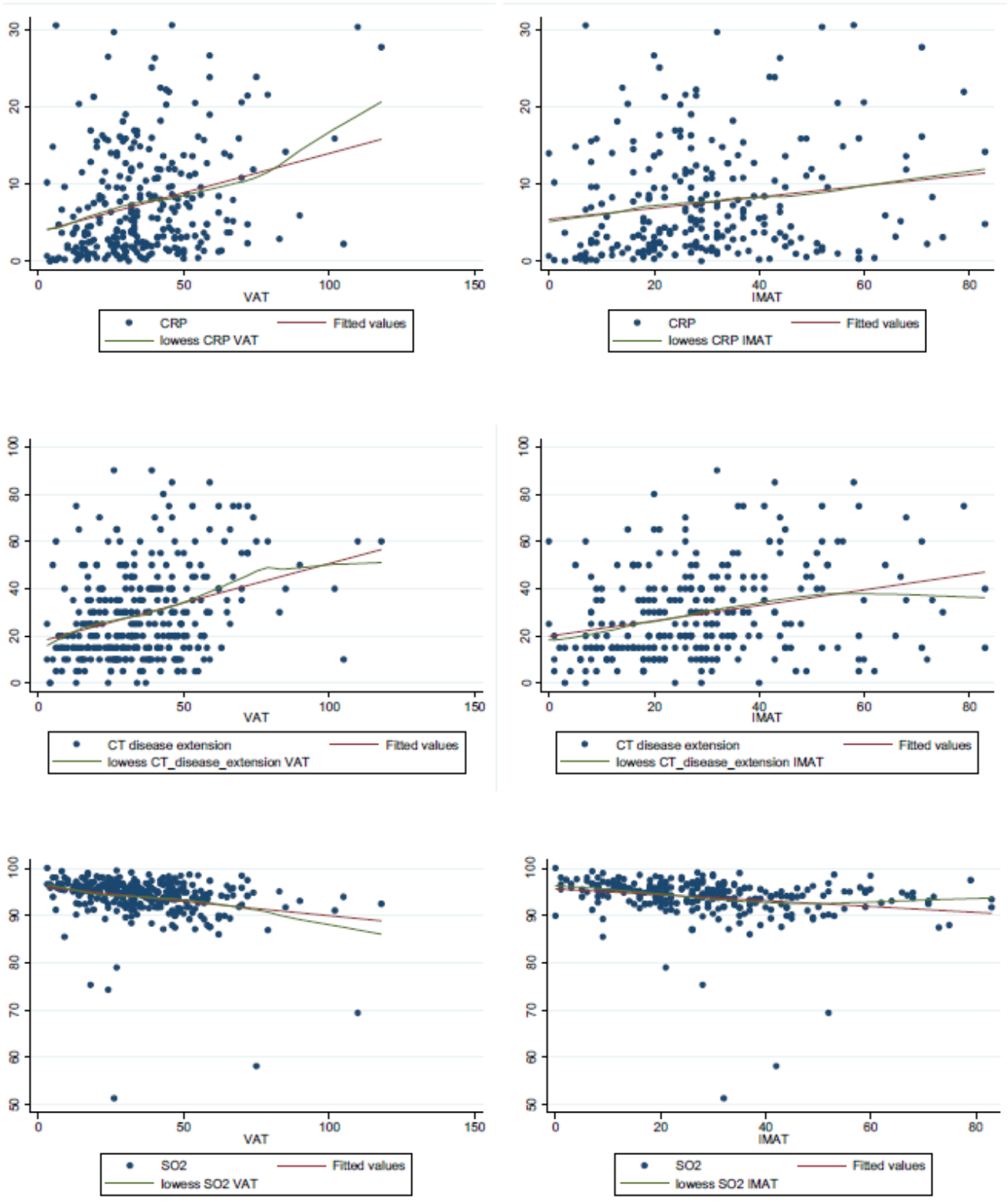
S4 Table. Distribution of CT body composition parameters expressed in quartiles in different age quartiles. P* Pearson's chi-squared test and p-value for the hypothesis of independence in the two-way table. IMAT: intermuscular adipose tissue area; TAT: total adipose tissue area; VAT: visceral adipose tissue area



S1 Fig. Representative images for different CT body composition parameters. Total adipose tissue, TAT (A), subcutaneous adipose tissue, SAT (B), visceral adipose tissue, VAT (C), intermuscular adipose tissue, IMAT (D), were all measured at the level of T7-T8 vertebrae. Pectoral muscle area and density were measured on the right side at a level immediately superior to the aortic arch (E).



S2 Fig. Figure S2. Regression line (red) and locally weighted scatterplot smoothing (lowess) smoother curve (green) superimposed on the scatter diagram for pectoral density and CRP, CT disease extension, SO2, and for TAT and CRP, CT disease extension, SO2. CRP: C-reactive protein; CT: Computed Tomography; SO2: oxygen saturation level; TAT: total adipose tissue area.



S3 Fig. Regression line (red) and lowess smoother curve (green) superimposed on the scatter diagram for VAT and CRP, CT disease extension, SO2, and for IMAT and CRP, CT disease extension, SO2. CRP: C-reactive protein; CT: Computed Tomography; IMAT: intermuscular adipose tissue area SO2: oxygen saturation level; VAT: visceral adipose tissue area.