

## Research article

# Spleen volume is independently associated with non-alcoholic fatty liver disease, liver volume and liver fibrosis

Samuel Helgesson<sup>a</sup>, Sambit Tarai<sup>a,b</sup>, Taro Langner<sup>b</sup>, Håkan Ahlström<sup>a,b</sup>,  
Lars Johansson<sup>b</sup>, Joel Kullberg<sup>a,b</sup>, Elin Lundström<sup>a,\*</sup>

<sup>a</sup> Radiology, Department of Surgical Sciences, Uppsala University, Sweden

<sup>b</sup> Antaros Medical AB, BioVenture Hub, Sweden



## ABSTRACT

Non-alcoholic fatty liver disease (NAFLD) can lead to irreversible liver damage manifesting in systemic effects (e.g., elevated portal vein pressure and splenomegaly) with increased risk of deadly outcomes. However, the association of spleen volume with NAFLD and related type 2-diabetes (T2D) is not fully understood. The UK Biobank contains comprehensive health-data of 500,000 participants, including clinical data and MR images of >40,000 individuals. The present study estimated the spleen volume of 37,066 participants through automated deep learning-based image segmentation of neck-to-knee MR images. The aim was to investigate the associations of spleen volume with NAFLD, T2D and liver fibrosis, while adjusting for natural confounders. The recent redefinition and new designation of NAFLD to metabolic dysfunction-associated steatotic liver disease (MASLD), promoted by major organisations of studies on liver disease, was not employed as introduced after the conduct of this study. The results showed that spleen volume decreased with age, correlated positively with body size and was smaller in females compared to males. Larger spleens were observed in subjects with NAFLD and T2D compared to controls. Spleen volume was also positively and independently associated with liver fat fraction, liver volume and the fibrosis-4 score, with notable volumetric increases already at low liver fat fractions and volumes, but not independently associated with T2D. These results suggest a link between spleen volume and NAFLD already at an early stage of the disease, potentially due to initial rise in portal vein pressure.

## 1. Introduction

The current pandemic of obesity, associated with ectopic fat, has increased the incidence of steatotic liver and related diseases. The gold-standard method for detecting steatohepatitis and quantifying liver fat content is histological examination of biopsies. Radiological examinations with ultrasound, computed tomography (CT) and magnetic resonance imaging (MRI) are non-invasive alternatives, providing image-based assessments with large liver coverage. The MRI-derived proton density fat fraction (PDFF) [1,2] is a biomarker for fat content, which is commonly used for studies on non-alcoholic fatty liver disease (NAFLD) [3]. The definition of NAFLD encompasses multiple conditions, ranging from liver steatosis to non-alcoholic steatohepatitis (NASH), and is based on exclusionary criteria (e.g., non-excessive alcohol consumption). Due to the inability of the term *non-alcoholic* to fully capture the aetiology of these conditions, and the expression *fatty* as potentially stigmatising, a recent Delphi consensus on a redefinition and new nomenclature was reached by several associations of liver disease [4]. The classification introduced metabolic dysfunction-associated steatotic liver disease (MASLD) as a substitute for NAFLD. The new definition MASLD includes cardiometabolic criteria for emphasising the association with metabolism [4], in contrast to NAFLD, typically defined only by a liver PDFF >5.56% not related to excessive alcohol consumption [5]. The new classification was introduced after the conduct of this study. Therefore, the old definition

\* Corresponding author.

E-mail address: [elin.lundstrom@uu.se](mailto:elin.lundstrom@uu.se) (E. Lundström).

and terminology of NAFLD was employed in this work. In studies of NAFLD, complementary information to the radiological examinations can be provided by other non-invasive tests [6–10], such as the blood sample-derived fibrosis-4 (FIB-4) score that reflects the likelihood of liver fibrosis [10].

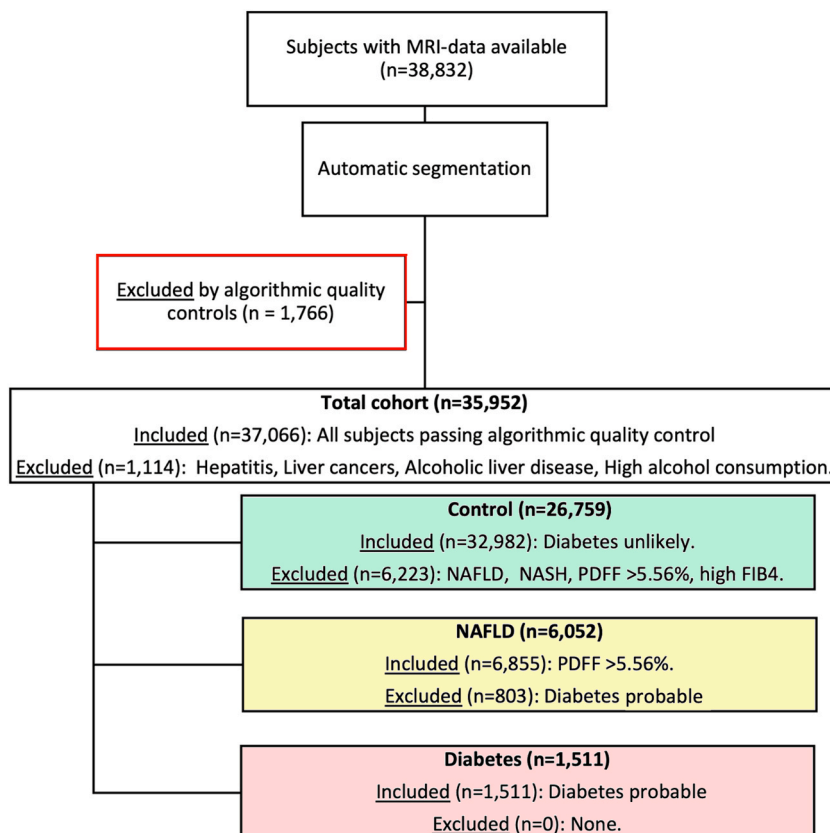
Clinically significant portal hypertension is caused by an increased liver resistance to perfusion in the presence of advanced chronic liver disease (advanced CLD), specifically that of liver stiffening due to fibrosis. This leads to a compensatory increase in portal vein pressure, propagating backwards to the interconnected splenic vein, which results in spleen enlargement. Morphological changes of the spleen can be observed in patients with advanced CLD, whereof splenomegaly is found in 60–65% [11,12]. Spleen size has been studied as a predictor of outcome for CLD [13–15]. Spleen and liver volume, normalized by body mass index (BMI), have been observed to correlate positively with hepatic venous pressure gradient (HVPG) non-cirrhotic portal hypertension [16].

The upper limit of normal spleen length is 12 cm as measured by ultrasound [17,18], the typical method for spleen size assessment in clinical practice. Measurement of spleen volume requires tomographic techniques, such as CT or MRI, and is not routinely assessed. Chow et al. suggested that the reference values of spleen length and volume should be updated, by implementing corrections for sex and height as an attempt to adjust for body size [19]. The weight and volume of many abdominal lean tissue organs are known to be positively associated with body size [20–22]. However, the correlation between spleen volume and such measurements varies between studies, in part likely due to differences in methodologies and study populations [20–22].

The UK Biobank (UKBB) is a large-scale biomedical database, with over 500,000 participants [23], available for research. Individuals of age 40–69 years were recruited by mail from the National Health Service between 2006 and 2010. Participation involves collection of blood, urine, saliva, and lifestyle questionnaires. In addition, medical records are available. The UKBB imaging study is a subcohort of UKBB, aiming for MRI of 100,000 participants [24].

Data from the UKBB imaging cohort allows for large-scale investigations of body composition and organ/tissue morphology in relation to various diseases, facilitated by the recent advances in automatic image analysis, such as image segmentation [25–27].

In context of the global rise in NAFLD, research targeting increased knowledge of disease progression (from liver steatosis, related obesity, and type 2-diabetes (T2D), towards steatohepatitis, portal hypertension, fibrosis, and cirrhosis), and non-invasive biomarkers for this progression, are warranted. Considering both the anatomical link between the spleen and liver, and that splenomegaly manifests in liver cirrhosis, potential relationships between spleen volume and NAFLD already at early-stage disease are of large



**Fig. 1.** Schematic overview of the subject groupings into cohorts. The sum of subjects in the three cohorts is less ( $n = 1630$ ) than the number of subjects in the total cohort. This is partly due to some subjects not being classified as either probable or unlikely to have diabetes ( $n = 1459$ ), as well as some subjects having incomplete data (e.g., PDFF values). Abbreviations: NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; PDFF, proton density fat fraction; FIB-4, fibrosis-4 score.

research interest and relatively unexplored. The UKBB provides an excellent opportunity to investigate such relationships due to the large number of subjects and comprehensive data available per individual.

The aim of this study was to investigate the associations of spleen volume with NAFLD, T2D and liver fibrosis, considering sex, age, and body size as possible confounders.

## 2. Materials and methods

### 2.1. Ethical considerations

Written informed consent for participation and publication was obtained from each participant by the UKBB at time of recruitment. The list of included participants was continuously updated according to information on participant withdrawal. This study was conducted within the confines of a research project on UKBB (application number 14237). Approval from the Swedish Ethical Review Authority was obtained prior to the start of this study (JUN 05, 2019, Dnr 2019-03073) and all procedures were carried out in accordance with the relevant guidelines and regulations.

### 2.2. Subjects

All UKBB participants of the imaging cohort, for which neck-to-knee body MRI data were available, were considered for this study. An initial quality control, by visual inspection of two-dimensional (2D) mean intensity projection representations of the three-dimensional (3D) data, was performed for excluding subjects with inadequate image quality (see flowchart in Fig. 1). Exclusion criteria included artifacts, such as fat-water signal swaps, intrusive amounts of background noise, metal artifacts, improper subject positioning during imaging and corrupt data. The number of available subjects after initial quality control was  $n = 38,832$ .

### 2.3. Imaging data

The neck-to-knee body MRI data were acquired at three sites, in five separate and overlapping stations, with parts of the abdomen present both in the second and third station. As a result, the spleen could be fully contained in either of these two stations or split between them. To ensure full coverage of the spleen, both stations were considered during image segmentation.

### 2.4. Segmentation of the spleen

A set of reference spleen segmentations, used for training of a deep neural network to automatically segment the remaining images of the UKBB imaging cohort, was created. Manual segmentations of the spleen were performed by a single operator for a total of 115 subjects. The operator was a medical student who received guidance from a professor in radiology. Segmentation of the whole spleen volume was conducted using SmartPaint 1.5.1 [28], a freely distributed software developed in-house, by delineation of 2D regions of interest (ROIs) in axial images.

Reference segmentations for network training consisted of two sets, A and B. Dataset A contained 85 subjects randomly sampled from the imaging cohort, as stratified by age, sex, and weight.

Dataset B contained 30 additional subjects, which represented challenging cases for the network when only dataset A was used as training data. The subjects of dataset B were determined by visual inspection of approximately 2000 coronal mean intensity projections. The types of errors occurring in dataset B were non-connected spleen volumes, failure to delineate large portions of the spleen or the whole spleen, and failure to capture the shape of the spleen.

The automatic segmentation method was based on a previously published kidney segmentation pipeline by Langner et al. [25], modified to predict a single whole volume instead of two. The spleen volume was calculated by multiplying the voxel volume by the number of voxels contained within the segmentation and expressed in litres (l). For more details, see Supplementary material.

### 2.5. Evaluation of the automatic segmentation

Repeated manual segmentations were performed on five subjects from dataset A to determine the intra-operator variability, representing a benchmark against which the performance of the automatic method could be compared. Subjects were selected by sampling with respect to normalized height and BMI, ensuring selection of various body sizes.

The similarity between the reference segmentation and the repeated segmentation was quantified using the Sorensen-Dice coefficient (Dice-score) [29], calculated according to  $Dice = (2|X \cap Y|) / (|X| + |Y|)$ , where X and Y were the two sets compared.

Eightfold cross-validation was performed in four different sets of the training data to evaluate segmentation performance of the automatic method compared to the manual delineation, by means of the Dice-score.

Following the automatic segmentation, an algorithmic quality control was used to flag cases with potential substandard automatic segmentations [25]. Segmentations with the worst algorithmic quality ratings were considered as potential failure cases and excluded from this study.

## 2.6. Additional image data and clinical data

Other MRI and clinical data relevant for this study were extracted from the UKBB. The data used in the analysis were obtained from the initial assessment visit (visit 1), at which participants were recruited and consent given (year 2006–2010), and from the imaging visit (visit 2), when the neck-to-knee body MRI examinations were performed (year 2014–2020). Detailed information about the UKBB, including data fields, can be found through the publicly available UKBB data showcase online [30].

Subject age was calculated as the difference between the date of birth (field 33) and the date of the MRI examination. Sex was obtained from the initial assessment visit (field 31).

Five different measurements were used for estimating the body size of the subject: height, weight, BMI, body surface area (BSA) and volumes of total lean tissue (TLT) and total adipose tissue (TAT). Height (field 12144) and weight (field 21002) were obtained from the imaging visit (year 2014–2020). BMI (field 21001) was calculated from the height and weight at the imaging visit (year 2014–2020):  $BMI = \text{weight} / \text{height}^2$  [kg/m<sup>2</sup>]. BSA (field 22427) was also calculated from the height and weight at the imaging visit (year 2014–2020):  $BSA = 0.20247 \cdot \text{weight}^{0.425} \cdot \text{height}^{0.725}$  [m<sup>2</sup>].

TLT (field 22416) and TAT (field 22415) were obtained using data from the imaging visit (year 2014–2020), according to the following two approaches: (1) The UKBB provided TLT and TAT estimates for approximately 10,000 subjects, derived from imaging data by Linge et al. [31]. (2) Estimates for the rest of the imaging cohort were obtained by deep regression analysis, using the subjects with readily available estimates as training data (method previously presented by Langner et al. [32]). A corresponding approach was used for assessing liver PDFF.

Liver volume was estimated using a method previously described by Mora et al. [33], by applying deep learning-based automatic segmentation on the UKBB data from the imaging visit (year 2014–2020).

Diagnoses were obtained by means of codes for international classification of diseases (ICD), tenth revision (*Diagnoses – main ICD-10*, field 41203, and *Diagnoses – secondary ICD-10*, field 41205). These data-fields contained a summary of all primary and secondary ICD-codes provided in the hospital inpatient records. No distinction between primary and secondary diagnoses were considered in this study.

Subjects were labelled with respect to T2D status, as either *unlikely*, *possible* or *probable*, according to a classification algorithm published by Eastwood et al. [34].

Alcohol consumption was derived from transforming drinking frequency (field 1558) into an average number of days per week of drinking and multiplying it with the number of units per occasion (field 20403). Men with >21 units per week and women with >14 units per week were defined to have high alcohol consumption [35], in addition to subjects reporting a frequency of consuming >6 units of alcohol (field 20416) daily or almost daily. These data were obtained from the initial assessment visit (year 2006–2010).

Blood and plasma data were retrieved from samples collected during the initial assessment visit (year 2006–2010), as only a limited amount of corresponding data from the imaging visit was available at the time of data analysis. The following data fields were used to assess the FIB-4 score: alanine aminotransferase (ALT) [U/l] (field 30620), aspartate aminotransferase (AST) [U/l] (field 30650) and platelet count [10<sup>9</sup> cells/l] (field 30080). FIB-4 was calculated according to  $FIB-4 = (\text{Age} \cdot \text{AST}) / (\text{Platelet count} \cdot \sqrt{\text{ALT}})$ , with a score >3.25 defined as high [10].

## 2.7. Cohort definitions

The total study cohort and its subcohorts are presented in the schematic overview in Fig. 1. The total cohort was created by including all subjects with an inferred spleen volume and by excluding subjects with high alcohol consumption (as defined above), ICD-codes for chronic viral hepatitis (B18.0, B18.1, B18.2, B18.8, B18.9), alcoholic liver disease (K70.0, K70.9) and liver cancers (C22.0, C22.1, C22.2, C22.3, C22.7, C22.9, D37.6, C78.7). No diseases apart from the abovementioned were considered, despite their potential relevance for spleen volume, such as autoimmune hepatitis, haematological diseases [36], celiac disease [37], Wilson disease [38] and steatogenic drug usage.

From this total cohort, three mutually exclusive cohorts were created based on NAFLD and T2D status. The *NAFLD cohort* included all subjects with liver PDFF >5.56%, and excluded subjects flagged as *probable diabetic*. The *Diabetes cohort* included all subjects flagged as *probable diabetic* irrespective of liver PDFF. The control cohort included all subjects flagged as *unlikely diabetic* and excluded subjects with any of the following: PDFF >5.56%, high FIB-4 score and ICD-codes for NAFLD (K76.0, K76.9) and NASH (K75.8, K75.9).

To investigate the specific relevance of NAFLD and T2D for spleen volume, two subcohorts of the T2D cohort were created: *T2D with coexisting NAFLD* and *T2D without NAFLD*.

## 2.8. Statistics

All statistics were conducted using the Statsmodels 0.13.1 [39] and SciPy 1.7.3 [40] libraries for Python. Two-sided Student's t-test of independent variables with non-equal variance (Welch's t-test) was used to investigate potential differences in spleen volumes between cohorts. P-values <0.05 were considered statistically significant.

Moving weighted average plots (based on a Gaussian mean, calculated using a sliding window) were created for visualizing, and evaluating differences in spleen volume as a function of age and other potentially associated variables, between different cohorts and between the sexes. Different ranges of x-values (i.e., sliding windows) were used for different x-axis variables and determined empirically. 95% confidence intervals (CIs), calculated from the same sliding windows, were plotted with a lighter colour around the weighted average curves.

Simple and multiple linear regression with ordinary least squares (OLS) were performed on the control cohort, with spleen volume as the dependent variable and different body size measurements as explanatory variables (sex and age as additional explanatory variables), for determining the body size measurements most strongly associated with spleen volume. These measurements were subsequently used for normalizing the spleen volume with body size, by means of their ratio: spleen volume / body size. The single most prominent body size measurement was finally used in linear regression with multiple explanatory variables related to NAFLD, T2D, liver fibrosis and basic characteristics (i.e., liver PDFF, liver volume, diabetes status, FIB-4, sex, age, body size) on the total cohort for determining the vast association between spleen volume and potentially related features. Spleen and liver volumes were subtracted from TLT before regression, to avoid including the same estimate in several variables. In all regressions, the dependent variable was transformed by the natural logarithm, i.e.,  $\ln(\text{spleen volume})$ , to fulfil conditions of linearity, normality of errors and homoscedasticity of errors. No correction for multiple comparisons was performed.

### 3. Results

Details from the evaluation of the automatic spleen segmentation performance are provided in the Supplementary Material. Results from association analysis of spleen volume, NAFLD, T2D and basic subject characteristics are presented in the following sections.

#### 3.1. Subject characteristics

Basic features for the total, control, NAFLD and T2D cohorts are presented in Table 1. Within the T2D cohort, 53.1% of the subjects showed coexisting NAFLD [ $n = 534$  (52.8%) of T2D males,  $n = 269$  (53.9%) of T2D females] whereas 46.9% did not [ $n = 478$  (47.2%) of T2D males,  $n = 230$  (46.1%) of T2D females]. Among the three cohorts of NAFLD, T2D with coexisting NAFLD and T2D without NAFLD, the largest spleen volume ( $210.72 \pm 90.78$  ml) was obtained for T2D with coexisting NAFLD ( $p < 0.001$  compared to remaining two groups). The spleen volume did not differ significantly between individuals with NAFLD versus T2D without NAFLD ( $191.72 \pm 78.00$  ml and  $185.28 \pm 93.00$  ml, respectively;  $p = 0.077$ ).

#### 3.2. Association between spleen volume, body size and age

Results from simple and multiple linear regression of spleen volume and different body size measurements are presented in Table 2, with sex and age as additional explanatory variables in the multiple regression. All explanatory variables showed weak to moderate significant associations with spleen volume. In simple linear regression, the highest correlation ( $R^2$ ) was obtained for BSA, followed by TLT and weight. In multiple linear regression, weight produced the highest correlation, with BSA slightly below. Sex and age were significant independent variables in all regressions. Weight, BSA and TLT were determined as the most prominent body size variables and therefore used for spleen volume normalization in subsequent association analyses. As an alternative approach to using weight as a total body size estimate and TLT as a lean body size estimate, TAT and TLT were together used as complementary explanatory variables for total body size in the regression. They provided an  $R^2 = 0.300$  ( $p < 0.001$  for both variables), being only marginally lower than corresponding estimated model fits from weight and BSA.

#### 3.3. Associations between spleen volume and liver fat content and volume

Moving average curves of non-normalized and normalized spleen volumes vs liver PDFF and liver volume for the whole cohort are shown in Fig. 2. Males generally showed larger non-normalized spleen volumes compared to females. Increasing spleen volume was

**Table 1**  
Subject characteristics for the total, control, NAFLD and T2D cohort.

	Total	Male	Female	Control	NAFLD	T2D
<b>Number of subjects (n)</b>	35,952	16,882	19,070	26,759	6052	1511
<b>Sex (M/F)</b>	M: 19,070 F: 16,882	M: 16,882 F: 0	M: 0 F: 19,070	M: 11,551 F: 15,208	M: 3548 F: 2504	M: 1012 F: 499
<b>Age (years)</b>	$64.29 \pm 7.57$	$64.99 \pm 7.66$	$63.68 \pm 7.43$	$64.00 \pm 7.57$	$64.18 \pm 7.30$	$66.95 \pm 7.05$
<b>Height (cm)</b>	$170.00 \pm 9.36$	$177.17 \pm 6.59$	$163.66 \pm 6.39$	$169.63 \pm 9.30$	$171.19 \pm 9.53$	$171.67 \pm 8.99$
<b>Weight (kg)</b>	$75.88 \pm 14.66$	$83.58 \pm 12.78$	$69.05 \pm 12.71$	$72.89 \pm 13.08$	$86.93 \pm 14.74$	$86.75 \pm 16.09$
<b>BMI (kg/m<sup>2</sup>)</b>	$26.49 \pm 4.24$	$26.96 \pm 3.72$	$26.08 \pm 4.61$	$25.56 \pm 3.66$	$29.98 \pm 4.27$	$29.79 \pm 5.13$
<b>BSA (m<sup>2</sup>)</b>	$1.86 \pm 0.20$	$2.00 \pm 0.16$	$1.73 \pm 0.15$	$1.82 \pm 0.19$	$1.98 \pm 0.20$	$1.97 \pm 0.20$
<b>TLT (l)</b>	$24.24 \pm 4.80$	$28.36 \pm 3.28$	$20.64 \pm 2.46$	$23.73 \pm 4.67$	$26.09 \pm 4.92$	$26.40 \pm 4.50$
<b>TAT (l)</b>	$20.94 \pm 6.87$	$19.92 \pm 6.37$	$21.84 \pm 7.16$	$19.68 \pm 6.21$	$25.99 \pm 6.93$	$24.81 \pm 7.68$
<b>Spleen volume (ml)</b>	$164.17 \pm 73.34$	$193.91 \pm 79.00$	$137.84 \pm 56.03$	$156.33 \pm 68.74$	$191.72 \pm 78.00$	$198.80 \pm 92.67$
<b>Liver PDFF (%)</b>	$3.88 \pm 4.19$	$4.47 \pm 4.33$	$3.36 \pm 3.98$	$2.24 \pm 1.03$	$10.88 \pm 5.16$	$7.78 \pm 6.14$
<b>Liver volume (l)</b>	$1.40 \pm 0.29$	$1.53 \pm 0.29$	$1.29 \pm 0.25$	$1.34 \pm 0.24$	$1.65 \pm 0.32$	$1.70 \pm 0.40$

Subject characteristics for the subcohorts: total, control, NAFLD and T2D.

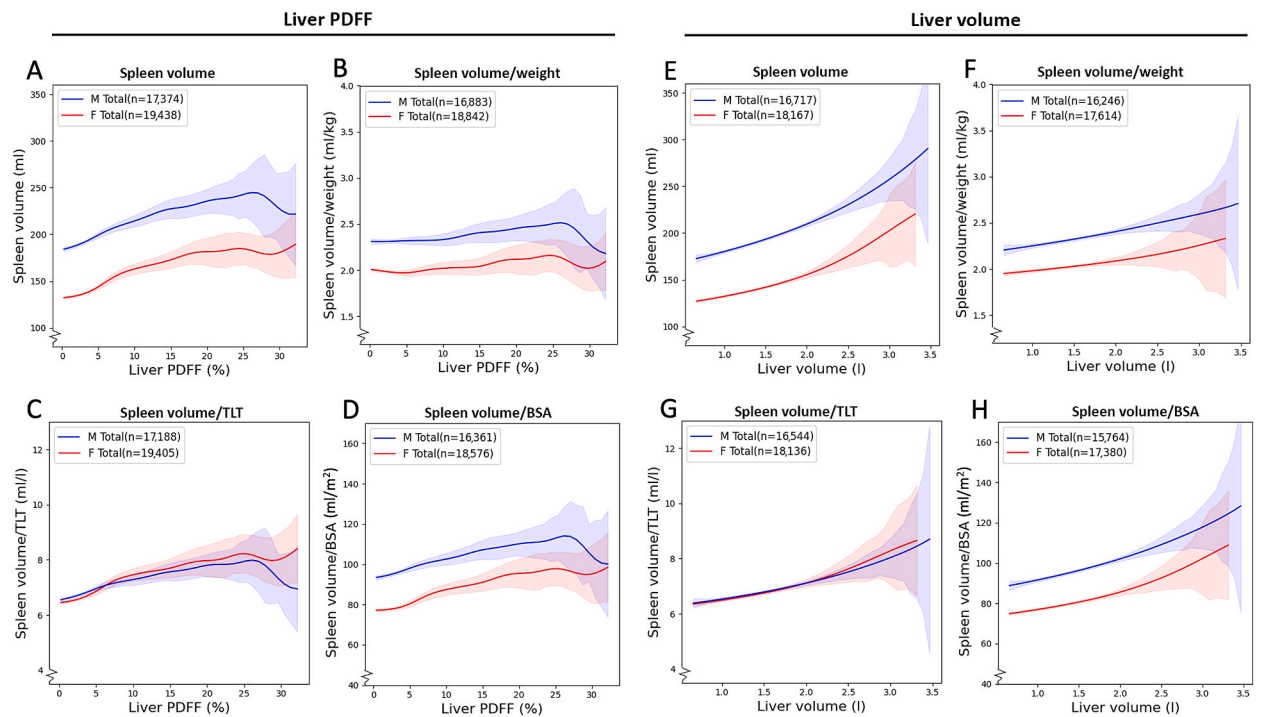
Values given as mean  $\pm$  standard deviation (except for number of subjects). Subject characteristics of males and females in the total cohort are presented separately in the second and third column, respectively. Abbreviations: M, males; F, females; BMI, body mass index; BSA, body surface area; TLT, total lean tissue volume; TAT, total adipose tissue volume; PDFF, proton density fat fraction.

**Table 2**

Simple and multiple linear regression of spleen volume (dependent variable) and body size measurements, age, and sex (explanatory variables) on the normal cohort.

Variables	Simple linear regression				Multiple linear regression					
	Beta	95% CI	R <sup>2</sup>	P	Beta	95% CI	R <sup>2</sup>	P	Main	Sex / Age
<b>ln(Spleen volume) vs..</b>										
Age	-0.006	[-0.007, -0.006]	0.014	<0.001	-0.53	[0.343, 0.258]	0.186	<0.001	<0.001 / -	
Height	0.019	[0.019, 0.020]	0.192	<0.001	0.012	[0.011, 0.013]	0.224	<0.001	<0.001 / <0.001	
Weight	0.016	[0.016, 0.016]	0.268	<0.001	0.013	[0.012, 0.013]	0.306	<0.001	<0.001 / <0.001	
BMI	0.035	[0.033, 0.036]	0.097	<0.001	0.030	[0.029, 0.031]	0.260	<0.001	<0.001 / <0.001	
BSA	1.134	[1.111, 1.156]	0.288	<0.001	0.968	[0.939, 0.997]	0.302	<0.001	<0.001 / <0.001	
TLT	0.047	[0.046, 0.048]	0.270	<0.001	0.049	[0.047, 0.051]	0.276	<0.001	0.002 / <0.001	
TAT	0.011	[0.010, 0.012]	0.028	<0.001	0.017	[0.016, 0.017]	0.251	<0.001	<0.001 / <0.001	
TAT & TLT					0.011	[0.010, 0.012]	0.300	<0.001	<0.001 / <0.001	
TAT & TLT					0.040	[0.038, 0.042]		<0.001		

The natural logarithm of spleen volume was used in the analyses. R<sup>2</sup> = adjusted R<sup>2</sup>. Beta = unstandardized coefficient. - = not applicable. Abbreviations: CI, confidence interval; BMI, body mass index; BSA, body surface area; TLT, total lean tissue volume; TAT, total adipose tissue volume; PDFF, proton density fat fraction. Units: spleen volume, ml; height, cm; weight, kg; BMI, kg/m<sup>2</sup>; BSA, m<sup>2</sup>; TLT, l; TAT, l.



**Fig. 2.** Moving weighted average curves for spleen volume vs A) – D) liver proton density fat fraction (PDFF) and E) – H) liver volume, with different normalizations for body size: A) and E) no normalization and normalization with respect to B) and F) weight, C) and G) volume of total lean tissue (TLT), and D) and H) body surface area (BSA). Weighted arithmetic mean values and 95% confidence intervals were based on A) – D) a gaussian PDFF sliding window of 2% and E) – H) a gaussian liver volume sliding window of 0.5 l.

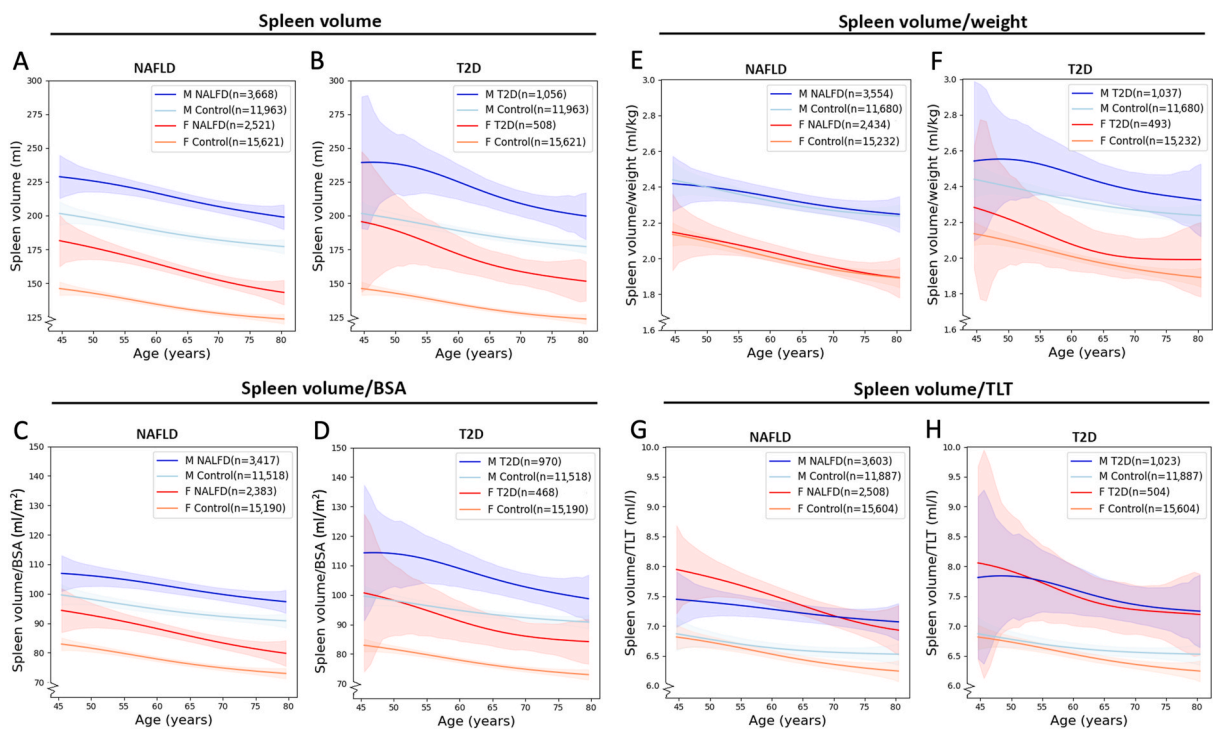
associated with increasing liver PDFF for both sexes already at low liver PDFF levels. For liver PDFF >25% ( $n = 105$  and  $64$  for females and males, respectively), the CIs were relatively large. Similar patterns were observed for normalized spleen volume vs liver PDFF, with the exception of no pronounced increase in spleen volume/weight with increasing liver PDFF (flat or possibly decreasing slope at PDFF <5%) and no significant difference in spleen volume/TLT between men and women. Increasing spleen volume was associated with increasing liver volume, for both sexes with and without normalization for body size, with no significant difference in spleen volume/TLT between men and women.

### 3.4. Association between spleen volume and age in sex-stratified NAFLD, T2D and controls

Moving average curves of non-normalized and normalized spleen volumes vs age (including 95% CI), in controls and subjects with NAFLD and T2D, are presented in Fig. 3. Both non-normalized and normalized spleen volumes were generally larger in males than females and in T2D and NAFLD compared to controls, for both sexes within the approximate age range 50–80 years. However, no general significant difference was observed between the NAFLD and T2D groups (overlapping CIs). Within each cohort, males and females showed a similar pattern of age-related decrease in spleen volume. After normalization of the spleen volume by weight, the difference between controls and each of NAFLD or T2D largely disappeared but the sex-related difference remained. Normalization of the spleen volume by BSA resulted in similar associations to those obtained by the non-normalized spleen volume. After normalization by TLT, the difference between the sexes largely diminished whereas the difference between the sex-stratified control, NAFLD and T2D cohorts remained. For the lowest and highest ages, and for the NAFLD and T2D cohorts in general, the measurement uncertainty (CI) was relatively large.

### 3.5. Association between spleen volume and multiple variables related to NAFLD, T2D and liver fibrosis

Results from multiple linear regression, with spleen volume as dependent variable and multiple explanatory variables related to NAFLD, T2D, liver fibrosis and basic characteristics, are presented in Table 3. Multiple linear regression with TLT and TAT as body size variables, instead of weight, is presented in Supplementary Table 2.



**Fig. 3.** Moving weighted average curves of **A) – B)** spleen volume and **C) – H)** normalized spleen volume vs age, in controls and subjects with non-alcoholic fatty liver disease (NAFLD) and type 2-diabetes (T2D). Normalization was conducted with respect to **C) – D)** body surface area (BSA), **E) – F)** weight, and **G) – H)** volume of total lean tissue (TLT). Weighted arithmetic mean values and 95% confidence intervals were based on a Gaussian age sliding window of 7 years.

**Table 3**  
Multiple linear regression of spleen volume (dependent variable) and variables related to NAFLD, T2D, liver fibrosis and basic characteristics (explanatory variables).

Variable	Beta	SE	t	P	95% CI
Constant	-2.9272	0.023	-127.061	0	[-2.972, -2.882]
Age	-0.0041	0.000	-14.741	$5.23 \cdot 10^{-49}$	[-0.005, -0.004]
Sex*	0.1314	0.005	28.857	$1.25 \cdot 10^{-180}$	[0.122, 0.140]
Weight	0.0062	0.000	29.376	$5.17 \cdot 10^{-187}$	[0.006, 0.007]
Liver PDFF	-0.0069	0.001	-12.339	$6.83 \cdot 10^{-35}$	[-0.008, -0.006]
Liver volume	0.5186	0.011	46.730	$< 5 \cdot 10^{-324}$	[0.497, 0.540]
FIB-4	0.0468	0.003	15.467	$9.33 \cdot 10^{-54}$	[0.041, 0.053]
T2D*	-0.0065	0.005	-1.427	0.154	[-0.015, 0.002]
				<b>n</b>	<b>R<sup>2</sup></b>
				39,880	0.375

The natural logarithm of spleen volume was used in the analysis.  $R^2$  = adjusted  $R^2$ . Beta = unstandardized coefficient. Abbreviations: NAFLD, non-alcoholic fatty liver disease; T2D, type 2-diabetes; CI, confidence interval; SE, standard error; PDFF, proton density fat fraction; FIB-4, fibrosis-4 score. Units: age, years; weight, kg; liver PDFF, %; liver volume, l. \* indicates dummy variable (0 = female, 1 = male; 0 = diabetes unlikely, 1 = diabetes possible, 2 = diabetes probable).

#### 4. Discussion

This large-scale imaging study showed the spleen volume to be independently associated with NAFLD, liver volume and liver fibrosis, in addition to a variation with respect to sex, age, and body size, according to multiple linear regression. Larger spleens were generally observed in men, relatively young subjects with large body size and in subjects with large livers, high liver fat content and high FIB-4.

The spleen volume increased already at low liver volumes and PDFF levels, indicating a relationship between spleen volume also in early NAFLD. Considering the association of spleen volume with sex, age, NAFLD and T2D in simple linear regression and analysis of moving average curves, the spleens of men were observed to be ~41% larger than those of women. In addition, the spleens of control subjects were ~18% and ~21% smaller than those of NAFLD and T2D subjects, respectively. However, differences varied slightly with age. The inverse relationship with age is in line with previous results from He et al. [21], showing an age-related decrease in spleen mass. The different measures of body size all showed weak to moderate positive correlation with spleen volume. These results contradict a previous publication by Mubbunu et al. [20], reporting no correlation between spleen weight with either subject weight or height. Caglar et al. [22], however, studied post-mortem spleens and observed BSA, BMI, body weight and body height to all be positively correlated with spleen volume and weight. The Pearson correlation coefficients reported by Caglar et al. [22] were  $r = 0.512$  (weight),  $r = 0.324$  (height),  $r = 0.416$  (BMI), and  $r = 0.519$  (BSA). These correlations were stronger than those obtained by simple linear regression in this study, potentially due to autopsies generating more accurate spleen measurements. However, the results from the two studies follow a similar relative pattern of height showing the weakest, and BSA, the strongest correlation with spleen volume.

Among the body size measurements, weight, BSA and TLT were obtained as the best explanatory variables (providing highest  $R^2$ ) in simple and multiple linear regression. Weight provided the single highest  $R^2$  in multiple linear regression (with sex and age as confounders), and therefore, was chosen to represent body size in subsequent regression with multiple variables related to NAFLD, T2D, liver fibrosis and basic characteristics (Table 3). In further association analyses of spleen volume with sex, age, NAFLD and T2D (Fig. 3), normalization of spleen volume by TLT reduced the sex-related difference whereas normalization by weight and TLT preserved the sex-related differences. In contrast, normalization of spleen volume by weight reduced the differences between controls and each disease group, with the three cohorts almost completely overlapping, whereas the differences largely remained for normalization by BSA and TLT. This indicates that TLT and weight correct for different aspects of the body size. Normalization by TLT appears to, at least partially, counterbalance the natural body size-related difference in spleen volume between men and women, possibly by targeting the typically larger muscle volume in men. Normalization by weight, however, includes adjustment for both lean and adipose tissue (similar to using TLT and TAT as complementary explanatory variables for body size) and might therefore compensate also for more pathologically related differences in spleen volume between controls and individuals with NAFLD and T2D. Normalization by BSA, however, does not change the inter-relationships between the sexes and cohorts as a function of age. Normalization of spleen volume by weight, BSA and TLT should therefore not be considered as interchangeable methods. However, a full investigation of the underlying aspects causing the differences in the normalized spleen volumes is beyond the scope of this study.

In general, the spleen volume was observed to increase with increasing liver PDFF, on average by 36 ml (~18% for men and ~26% for women) between 5% and 20% in liver PDFF. This positive relationship remained also after spleen volume normalization by TLT and BSA but not after normalization by weight. This is likely related to large subjects (with large spleens) tending to have higher liver PDFF,



but that this positive relationship effectively is cancelled out by normalization by weight as including both an adipose (i.e., more pathologically related) and lean tissue component. BSA does seemingly not correct for both adipose and lean tissue, to the same extent as weight, and TLT only corrects for lean tissue per definition. Absolute and normalized spleen volume were observed to be positively associated with liver volume. With an increase in liver volume from 1 to 2 l, the absolute spleen volume increased by ~17% in both men and women. These overall findings, largely supported by the moving average curves, support the theory of a link between spleen size and early stages of NAFLD, although the increase in spleen volume was subtle for relatively large increases in liver PDFF and volume. This indicates that portal hypertension, which is one of the main known mechanisms behind splenomegaly in CLD, increases even in early, pre-cirrhotic stages of NAFLD. This is in line with recent literature, where Nababan et al. [41] reviewed studies suggesting portal hypertension to potentially develop already in the early course of NAFLD, before the onset of significant fibrosis and cirrhosis. The discussion was based on observations of mildly elevated HVPG (5–9 mmHg) in a fraction of non-cirrhotic NAFLD patients [42–44] and reports of no significant fibrosis in 12% of examined NAFLD patients with signs of portal hypertension (e.g., splenomegaly) [45]. In these cases, portal hypertension was instead positively associated or even suggested to be driven by hepatic steatosis. A reason for the potential early onset of portal hypertension in NAFLD could be the pattern of fibrosis being different from that in other forms of CLD (e.g., viral hepatitis) [45]. In one of the studies, the fibrosis stage was observed to be positively associated with both the presence and number of findings of portal hypertension in NAFLD patients, which supports the assumption of their interrelationship [45].

The isolated conditions of NAFLD and T2D resulted in similar spleen enlargement, and their coexistence, to further enlargement. This suggests that NAFLD and T2D are associated with independent and similar effects on spleen size, compared to controls, which combines during their coexistence. Splenomegaly has previously been observed in Caucasians with T2D, without specific information provided on the subsets with/without NAFLD [46]. In the present study, however, T2D was not observed as an independent explanatory variable in the linear regression including multiple variables related to NAFLD, T2D and liver fibrosis. This might be due to other explanatory variables, related to NAFLD and T2D, being more important for the prediction.

The linear regression model with multiple variables related to NAFLD, T2D, liver fibrosis and basic characteristics (see Table 3) showed statistically significant correlations between the spleen volume and all explanatory variables except for T2D. This suggests that liver PDFF, liver volume, FIB-4, sex, age, and weight, but not T2D, are independent predictors for spleen volume in the UKBB cohort. However, the correlation was moderate ( $R^2 = 0.375$ ). This is probably related to both limitations of the model (i.e., that variables not considered in this study have a relatively large influence on spleen volume) and variations in data collection, timing of data collections, reporting, processing and analysis of data in the multi-scale and multi-centre UKBB project (see limitations of the study below). Attempts to improve the predictive model for the spleen volume was, however, beyond the scope of this study and an avenue for future research aiming to elucidate spleen physiology and its link to pathologies. In addition, liver PDFF showed a negative correlation in this multiple linear regression, contradicting the positive correlation obtained in previous analyses. This deviating result was possibly caused by certain covariation with other explanatory variables, whereby the negative relationship between spleen volume and liver PDFF should be interpreted with caution.

The image segmentation deep neural network achieved a Dice score slightly lower than the measured intra-operator variability (0.928 vs 0.956). Cross-validation results (Supplementary Table 1) showed that inclusion of problematic cases in the training dataset slightly improved the ability of the network to segment problematic cases (from Dice score 0.840 to 0.850), meanwhile causing a small reduction in Dice-score for normal cases (0.928 vs 0.932). As the benefit of improved segmentation performance for problematic cases was considered to outweigh the reduced segmentation performance for normal cases, the final spleen segmentation was based on training data from both normal and problematic cases. This was also in line with a previously published kidney segmentation approach [25]. The kidney segmentation method reached a Dice score of 0.956 and an intra-operator variability of 0.962, despite a lower number of training data compared to this study ( $n = 64$  vs  $n = 115$ ). The lower performance of the spleen segmentation method may be due to several factors. The consistency of the manual spleen segmentations, by the novice operator, is a possible factor. Morphological differences between the two organs, where the kidney shape, size and location might vary less compared to the spleen, is another factor. A previous publication by Kart et al., on deep learning-based spleen segmentation of the UKBB cohort, presented a superior Dice score to that obtained in the present study (Dice = 0.946) [27]. This higher Dice score was likely influenced by the large number of manual segmentations performed by radiologists ( $n = 200$  datasets), forming the training data. A similar publication on the UKBB cohort, by Liu et al. [26], achieved a validation Dice score of 0.910 and presented spleen volumes of (mean  $\pm$  standard deviation)  $0.20 \pm 0.078$  l for males and  $0.14 \pm 0.054$  l for females. These measurements are very similar to the  $0.19 \pm 0.086$  l for males and  $0.14 \pm 0.058$  l for females obtained in this study, which supports adequate segmentation performance and spleen volume estimates in this study. The same research group also recently published findings of increased spleen iron content with age, based on deep learning segmentations in the UKBB cohort, with an even more similar Dice score (0.922) to that of the present study [47].

An obvious strength of the UKBB is the large number of subjects available for inclusion, which provides high statistical power and possibility to include multiple variables in the analyses. The scale of this study would not be practically feasible, however, without the use of automatic assessment of for example volumes and liver PDFF from the MR images. Manual spleen segmentation in a single subject using SmartPaint required an average duration of 40 min, serving as a benchmark for segmentation efficiency. This implies that manual segmentation of 40,000 subjects would require roughly 26,700 h. In contrast, the network completed its inference run on all MRI examinations in approximately 40 h, without the need for human intervention.

Another strength of the study was the 3D volume measurements of the spleen size in-vivo, possible from MRI. The traditional approach to assess spleen size in studies, however, is by 2D measurements from ultrasound in-vivo [19] or measurements from autopsy.

The largest limitation of the study was the definition of its cohorts, especially the “healthy” control cohort. It is not fully known to

which extent these subjects were healthy, other than demonstrating low FIB-4 scores, low liver PDFF levels and were unlikely of having T2D. They could potentially have had other diseases affecting their spleen volume (e.g., haematological diseases and steatosis of other origin). Such diseases could possibly have influenced all cohorts to a certain extent. In addition, the NAFLD and T2D cohorts lacked a gold standard measurement for verification of diagnosis. By using additional data, such as alcohol habits, an attempt to robust exclusions was carried out.

At first, ICD-codes were used for creating cohorts. However, this approach was abandoned due to the relatively small number of subjects with registered liver-related disease as compared to the expected prevalence, which indicated hidden statistics. Based on this consideration, the ICD-codes were merely used as exclusion criteria. The new nomenclature and redefinition of the subclasses of steatotic liver disease include MASLD, which has a largely overlapping aetiology with NAFLD [4]. The new subclass would likely have improved the definition of the cohorts in the present study but was introduced after study completion and therefore not applied.

Another possible limitation was the algorithmic quality control employed to remove potential failure spleen segmentations. However, these potential failures were not evaluated by human visual inspection.

Some of the clinical data, such as those from blood samples, were collected at a different, earlier visit than the imaging visit. This mismatch in timing between data acquisitions was suboptimal for interpretation and linking between data as the health of the subjects could change over time. This problem was somewhat mitigated by CLD being a chronic disease, resulting in for example similar high FIB-4 scores over a couple of years. However, a recent study showed that approximately 20% of patients with NAFLD, who demonstrated initial low risk of advanced fibrosis ( $FIB-4 < 1.45$ ) at time of diagnosis, progressed to indeterminate or high risk ( $FIB-4 \geq 1.45$ ) after three years [48]. The uncertainty and timing of the T2D diagnosis were related limitations of relevance in the present study.

Lastly, this study was limited by the relatively old age of the participants, spanning from 44 to 80 years. As the results showed a decrease in spleen volume with age, it would have been interesting to investigate the corresponding relationship also in younger adults and whether it was affected by for example menopause in women.

## 5. Conclusion

In the age span ~50–80 years, the spleen volume decreased with age, correlated positively with body size and was smaller for females compared to males. Larger spleens were observed in subjects with NAFLD and T2D compared to controls. Spleen volume was also positively and independently associated with liver PDFF, liver volume and FIB-4, with notable increases already at low liver PDFF levels and liver volumes. These results suggest a link between spleen volume and NAFLD already at an early stage of the disease, potentially due to initial rise in portal vein pressure.

## Ethics approval and consent to participate

This study was reviewed and approved by The Swedish Review Authority, with the approval number: Dnr 2019–03073. All participants/patients provided informed consent to participate in the study.

## Data availability statement

The generated spleen volume data will be shared via the UKBB data return system. Other datasets analysed during the current study are available in the UKBB repository, <https://www.ukbiobank.ac.uk/>. The trained segmentation model will be shared upon reasonable request directed to E. Lundström by email ([elin.lundstrom@uu.se](mailto:elin.lundstrom@uu.se)).

## Funding

This research was funded by the Swedish Heart Lung Foundation (Hjärtlungfonden, 20220129) and the Swedish Research Council (Vetenskapsrådet, 2016–01040, 2019–04756). This research has been conducted using the UKBB Resource under Application Number 14237 and the Swedish National Infrastructure for Computing (SNIC, sens2019016). The computation and data handling were enabled by resources provided by the SNIC at Uppsala University, partially funded by the Swedish Research Council through grant agreement 2018-05973.

## CRedit authorship contribution statement

**S. Helgesson:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **S. Tarai:** Writing – review & editing, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **T. Langner:** Writing – review & editing, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **H. Ahlström:** Writing – review & editing, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **L. Johansson:** Writing – review & editing, Methodology, Investigation. **J. Kullberg:** Writing – review & editing, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **E. Lundström:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Håkan Ahlström reports financial support was provided by the Swedish Heart Lung Foundation. Håkan Ahlström reports financial support was provided by the Swedish Research Council. Joel Kullberg reports financial support was provided by the Swedish Research Council. Sambit Tarai, Taro Langner, Elin Lundström, Joel Kullberg, Håkan Ahlström and Lars Johansson report a relationship with Antaros Medical AB that includes: consulting or advisory, employment, and equity or stocks. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e28123>.

## References

- [1] I.S. Idilman, O. Keskin, A. Celik, B. Savas, A.H. Elhan, R. Idilman, M. Karcaaltincaba, A comparison of liver fat content as determined by magnetic resonance imaging-proton density fat fraction and MRS versus liver histology in non-alcoholic fatty liver disease, *Acta Radiol.* 57 (2016) 271–278.
- [2] E.R. Heba, A. Desai, K.A. Zand, et al., Accuracy and the effect of possible subject-based confounders of magnitude-based MRI for estimating hepatic proton density fat fraction in adults, using MR spectroscopy as reference, *J. Magn. Reson. Imag.* 43 (2016) 398–406.
- [3] M. Noureddin, J. Lam, M.R. Peterson, et al., Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials, *Hepatology* 58 (2013) 1930–1940.
- [4] M.E. Rinella, J.V. Lazarus, V. Ratziu, et al., A multisociety Delphi consensus statement on new fatty liver disease nomenclature, *J. Hepatol.* 79 (2023) 1542–1556.
- [5] L.S. Szczepaniak, P. Nurenberg, D. Leonard, et al., Magnetic resonance spectroscopy to measure hepatic triglyceride content: prevalence of hepatic steatosis in the general population, *Am. J. Physiol. Endocrinol. Metab.* 288 (2005) E462–E468.
- [6] P. Angulo, J.M. Hui, G. Marchesini, et al., The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD, *Hepatology* 45 (2007) 846–854.
- [7] J. Parkes, P. Roderick, S. Harris, et al., Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease, *Gut* 59 (2010) 1245–1251.
- [8] Y. Vali, J. Lee, J. Boursier, et al., Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: a systematic review and meta-analysis, *J. Hepatol.* 73 (2020) 252–262.
- [9] N. Peleg, A. Issachar, O. Sneh-Arbib, A. Shlomai, AST to Platelet Ratio Index and fibrosis 4 calculator scores for non-invasive assessment of hepatic fibrosis in patients with non-alcoholic fatty liver disease, *Dig. Liver Dis.* 49 (2017) 1133–1138.
- [10] Y. Sumida, M. Yoneda, K. Tokushige, et al., FIB-4 first in the diagnostic algorithm of metabolic-dysfunction-associated fatty liver disease in the era of the global metabodemic, *Life* 11 (2021).
- [11] P.R. Gibson, R.N. Gibson, M.R. Ditchfield, J.D. Donlan, Splenomegaly—an insensitive sign of portal hypertension, *Aust. N. Z. J. Med.* 20 (1990) 771–774.
- [12] M. Bolognesi, C. Merkel, D. Sacerdoti, V. Nava, A. Gatta, Role of spleen enlargement in cirrhosis with portal hypertension, *Dig. Liver Dis.* 34 (2002) 144–150.
- [13] S. Voutilainen, R. Kivisaari, J. Lohi, H. Jalanko, M.P. Pakarinen, A prospective comparison of noninvasive methods in the assessment of liver fibrosis and esophageal varices in pediatric chronic liver diseases, *J. Clin. Gastroenterol.* 50 (2016) 658–663.
- [14] J.S. Bae, D.H. Lee, J. Yoo, et al., Association between spleen volume and the post-hepatectomy liver failure and overall survival of patients with hepatocellular carcinoma after resection, *Eur. Radiol.* 31 (2021) 2461–2471.
- [15] Y. Huang, B. Huang, T. Kan, B. Yang, M. Yuan, J. Wang, Liver-to-spleen ratio as an index of chronic liver diseases and safety of hepatectomy: a pilot study, *World J. Surg.* 38 (2014) 3186–3192.
- [16] O. Etzion, V. Takyar, V. Novack, et al., Spleen and liver volumetrics as surrogate markers of hepatic venous pressure gradient in patients with noncirrhotic portal hypertension, *Hepatol Commun* 2 (2018) 919–928.
- [17] W.K. Loftus, L.T. Chow, C. Metreweli, Sonographic measurement of splenic length: correlation with measurement at autopsy, *J. Clin. Ultrasound* 27 (1999) 71–74.
- [18] M. Sienz, A. Ignee, C.F. Dietrich, [Reference values in abdominal ultrasound - biliopancreatic system and spleen], *Z. Gastroenterol.* 49 (2011) 845–870.
- [19] K.U. Chow, B. Luxembourg, E. Seifried, H. Bonig, Spleen size is significantly influenced by body height and sex: establishment of normal values for spleen size at US with a cohort of 1200 healthy individuals, *Radiology* 279 (2016) 306–313.
- [20] L. Mubbunu, K. Bowa, V. Petrenko, M. Silitongo, Correlation of internal organ weights with body weight and body height in normal adult Zambians: a case study of ndola teaching hospital, *Anat Res Int* 2018 (2018) 4687538.
- [21] Q. He, S. Heshka, J. Albu, L. Boxt, N. Krasnow, M. Elia, D. Gallagher, Smaller organ mass with greater age, except for heart, *J. Appl. Physiol.* 106 (2009) 1780–1784, 1985.
- [22] V. Caglar, b. Kumral, R. Uygur, O.A. Alkoç, O. Ozen, H. Demirel, Study of volume, weight and size of normal pancreas, spleen and kidney in adults autopsies, *Forensic Med. Anat. Res.* 2 (2014) 63–69.
- [23] C. Sudlow, J. Gallacher, N. Allen, et al., UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age, *PLoS Med.* 12 (2015) e1001779.
- [24] J. West, O. Dahlqvist Leinhard, T. Romu, et al., Feasibility of MR-based body composition analysis in large scale population studies, *PLoS One* 11 (2016) e0163332.
- [25] T. Langner, A. Ostling, L. Maldonis, et al., Kidney segmentation in neck-to-knee body MRI of 40,000 UK Biobank participants, *Sci. Rep.* 10 (2020) 20963.
- [26] Y. Liu, N. Bastý, B. Whittecher, et al., Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning, *Elife* 10 (2021).
- [27] T. Kart, M. Fischer, T. Kustner, et al., Deep learning-based automated abdominal organ segmentation in the UK biobank and German national cohort magnetic resonance imaging studies, *Invest. Radiol.* 56 (2021) 401–408.
- [28] F. Malmberg, R. Nordenskjöld, R. Strand, J. Kullberg, SmartPaint — a tool for interactive segmentation of medical volume images, *Comput. Methods Biomech. Biomed. Eng.: Imaging & Visualization* (2014) 36–44.
- [29] A. Carass, S. Roy, A. Gherman, et al., Evaluating white matter lesion segmentations with refined sorensen-dice analysis, *Sci. Rep.* 10 (2020) 8242.
- [30] UK Biobank online showcase, Available from: <https://biobank.ndph.ox.ac.uk/showcase/search.cgi>.
- [31] J. Linge, M. Borga, J. West, et al., Body composition profiling in the UK biobank imaging study, *Obesity* 26 (2018) 1785–1795.
- [32] T. Langner, R. Strand, H. Ahlström, J. Kullberg, Large-scale inference of liver fat with neural networks on UK Biobank body MRI, in: 23rd International Conference of Medical Image Computing and Computer Assisted Intervention – MICCAI, Springer, 2020.

- [33] A. Martínez Mora, T. Langner, A. Korenyushkin, D. Olmo, A. Karlsson, H. Ahlström, J. Kullberg, Automated liver volume measurements in more than 35,000 subjects from neck-to-knee MRI in UK-biobank, in: AASLD Liver Meeting 2020, 2020.
- [34] S.V. Eastwood, R. Mathur, M. Atkinson, et al., Algorithms for the capture and adjudication of prevalent and incident diabetes in UK biobank, *PLoS One* 11 (2016) e0162388.
- [35] A.J. Sanyal, E.M. Brunt, D.E. Kleiner, et al., Endpoints and clinical trial design for nonalcoholic steatohepatitis, *Hepatology* 54 (2011) 344–353.
- [36] E. Curovic Rotbain, D. Lund Hansen, O. Schaffalitzky de Muckadell, F. Wibrand, A. Meldgaard Lund, H. Frederiksen, Splenomegaly - diagnostic validity, work-up, and underlying causes, *PLoS One* 12 (2017) e0186674.
- [37] E. Valletta, M. Fornaro, M. Cipolli, S. Conte, F. Bissolo, C. Danchielli, Celiac disease and obesity: need for nutritional follow-up after diagnosis, *Eur. J. Clin. Nutr.* 64 (2010) 1371–1372.
- [38] D.S. Allende, D.E. Kleiner, Fatty liver disease that is neither metabolic nor alcoholic, *Hum. Pathol.* 141 (2023) 212–221.
- [39] Introduction — statsmodels, Available from: <https://www.statsmodels.org/stable/index.html>.
- [40] SciPy. Available from: <https://scipy.org/>.
- [41] S.H.H. Nababan, C.R.A. Lesmana, Portal hypertension in nonalcoholic fatty liver disease: from pathogenesis to clinical practice, *J Clin Transl Hepatol* 10 (2022) 979–985.
- [42] S. Francque, A. Verrijken, I. Mertens, et al., Noncirrhotic human nonalcoholic fatty liver disease induces portal hypertension in relation to the histological degree of steatosis, *Eur. J. Gastroenterol. Hepatol.* 22 (2010) 1449–1457.
- [43] S. Francque, A. Verrijken, I. Mertens, et al., Visceral adiposity and insulin resistance are independent predictors of the presence of non-cirrhotic NAFLD-related portal hypertension, *Int. J. Obes.* 35 (2011) 270–278.
- [44] L. Vonghia, T. Magrone, A. Verrijken, P. Michielsen, L. Van Gaal, E. Jirillo, S. Francque, Peripheral and hepatic vein cytokine levels in correlation with non-alcoholic fatty liver disease (NAFLD)-Related metabolic, histological, and haemodynamic features, *PLoS One* 10 (2015) e0143380.
- [45] F.D. Mendes, A. Suzuki, S.O. Sanderson, K.D. Lindor, P. Angulo, Prevalence and indicators of portal hypertension in patients with nonalcoholic fatty liver disease, *Clin. Gastroenterol. Hepatol.* 10 (2012) 1028–1033 e1022.
- [46] L.E. Davidson, D.E. Kelley, S. Heshka, et al., Skeletal muscle and organ masses differ in overweight adults with type 2 diabetes, *J. Appl. Physiol.* 117 (2014) 377–382, 1985.
- [47] E.P. Sorokin, N. Bastý, B. Whitcher, et al., Analysis of MRI-derived spleen iron in the UK Biobank identifies genetic variation linked to iron homeostasis and hemolysis, *Am. J. Hum. Genet.* 109 (2022) 1092–1104.
- [48] G. Cholankeril, J.R. Kramer, J. Chu, et al., Longitudinal changes in fibrosis markers are associated with risk of cirrhosis and hepatocellular carcinoma in non-alcoholic fatty liver disease, *J. Hepatol.* 78 (2023) 493–500.