

BMJ Open Metabolic dysfunction-associated steatotic liver disease and colorectal neoplasms risk: a global propensity score-matched retrospective cohort study

Mohammad Aldiabat ^{1,2}, Ali Osman,¹ Malek Ayoub,¹ Mahmoud Y Madi,³ Kamran Qureshi,³ Wing-Kin Syn^{3,4}

To cite: Aldiabat M, Osman A, Ayoub M, *et al*. Metabolic dysfunction-associated steatotic liver disease and colorectal neoplasms risk: a global propensity score-matched retrospective cohort study. *BMJ Open* 2025;**15**:e104934. doi:10.1136/bmjopen-2025-104934

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2025-104934>).

Received 11 May 2025
Accepted 27 October 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

¹Department of Medicine, Washington University in St Louis, St. Louis, Missouri, USA

²Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA

³Division of Gastroenterology & Hepatology, Saint Louis University, St. Louis, Missouri, USA

⁴Department of Physiology, Faculty of Medicine and Nursing, University of Basque Country UPV/EHU, Leioa, Vizcaya, Spain

Correspondence to

Mohammad Aldiabat;
Mohameddiabat88@gmail.com

ABSTRACT

Objectives To evaluate the association between metabolic dysfunction-associated steatotic liver disease (MASLD), metabolic-associated steatohepatitis (MASH), and the risk of colorectal cancer (CRC) and benign colorectal neoplasms (BCN), and to explore whether liver fibrosis/cirrhosis modifies these associations.

Design Retrospective cohort study with 1:1 propensity score matching.

Setting Global, multicentre real-world analysis using deidentified electronic health records from over 130 healthcare organisations in the TriNetX Global Collaborative Network.

Participants Hospitalised adults aged 45–75 years between October 2019 and October 2024. Patients with prior diagnoses of colorectal neoplasia or other chronic liver diseases were excluded. Final matched cohorts included 138 902 MASLD and non-MASLD patients, 3715 MASH and non-MASH patients, and 1312 MASH patients with and without fibrosis.

Primary and secondary outcome measures Primary outcomes: Incidence of CRC and BCN. Secondary outcome: Combined incidence of CRC and BCN. Outcomes were assessed with and without controlling for metabolic risk factors using Cox proportional hazards models.

Results MASLD was associated with increased risks of CRC (HR 2.71, 95% CI 2.29 to 3.20) and BCN (HR 2.50, 95% CI 2.38 to 2.63), both $p < 0.001$. MASH patients had a 5-fold higher risk of CRC (HR 5.03, 95% CI 1.43 to 17.72, $p = 0.005$) and nearly 2-fold risk of BCN (HR 1.91, 95% CI 1.38 to 2.67, $p < 0.001$). No significant differences in CRC or BCN risk were observed between MASH patients with versus without fibrosis/cirrhosis.

Conclusions MASLD and MASH are independent risk factors for CRC and BCN, irrespective of metabolic comorbidities. Fibrosis/cirrhosis does not significantly influence CRC risk. These findings support the need to revisit CRC screening guidelines for patients with MASLD/MASH. Further prospective studies are warranted to explore underlying mechanisms and evaluate preventative interventions.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Use of a large, diverse real-world dataset from over 130 healthcare organisations worldwide enhances generalisability and external validity.
- ⇒ Propensity score matching minimised confounding by balancing key demographic and metabolic variables across comparison groups.
- ⇒ Standardised ICD-10 codes were used for exposure and outcome definitions, ensuring reproducibility and consistency.
- ⇒ Retrospective design introduces potential for residual confounding and misclassification bias from reliance on coded diagnoses.
- ⇒ Lack of data on colonoscopy history, lifestyle factors and severity of liver or metabolic disease may limit adjustment for unmeasured confounders.

INTRODUCTION

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the second leading cause of cancer-related mortality globally, underscoring the substantial burden it imposes on healthcare systems worldwide.¹ Recent epidemiological data suggest that the global incidence of CRC is increasing, driven in part by sedentary lifestyles, ageing populations and escalating rates of metabolic disorders. Among the metabolic conditions is metabolic dysfunction-associated steatotic liver disease (MASLD), formerly known as non-alcoholic fatty liver disease (NAFLD).² Increasingly, investigators have noted that MASLD extends beyond merely a hepatic condition, potentially contributing to an elevated risk of extrahepatic malignancies, such as CRC.³

Multiple cross-sectional and cohort studies have repeatedly demonstrated a higher prevalence of colorectal adenomas and carcinoma in patients with MASLD compared

with individuals without underlying liver disease.^{4–6} A meta-analysis by Zeng *et al*⁷ synthesised 14 studies and reported that the risk of colorectal neoplasms increases with MASLD severity, particularly for left-sided lesions. Zhao *et al*⁸ conducted an updated cohort-only meta-analysis of nearly 10 million participants and found MASLD to be associated with higher risks of incident CRC and colorectal adenoma. Most recently, Celsa *et al*⁹ compared MASLD and Metabolic Dysfunction and Alcohol-Associated Steatotic Liver Disease across 24 cohort studies and demonstrated distinct risks for hepatic and extrahepatic cancers.

This association between MASLD and CRC stems from a shared pathophysiological mechanism involving insulin resistance, which promotes a pro-inflammatory state through adipocytokine dysregulation, hepatic lipid accumulation and oxidative stress.¹⁰ These processes can accelerate colorectal carcinogenesis by altering cell signalling pathways and fostering the development of adenomatous polyps.⁴ In addition, intestinal dysbiosis—a disturbance of gut microbiota that often accompanies obesity and metabolic syndrome—may further augment this risk by affecting epithelial integrity, inflammatory responses and local immune surveillance.¹⁰

Despite these insights, most available studies have been limited by region-specific populations, heterogeneous definitions of MASLD, MASH and fibrosis, and incomplete adjustment for metabolic comorbidities. Our global study addresses these gaps by (1) separately evaluating MASLD and MASH, (2) examining the influence of liver fibrosis/cirrhosis and (3) isolating the effect of these conditions on colorectal neoplasm risk after controlling for traditional metabolic risk factors using propensity score matching (PSM). This design allows for a more nuanced understanding of how metabolic liver disease contributes to CRC and benign colorectal neoplasm (BCN) risk, providing evidence to inform early detection and prevention strategies.

METHODS

Study design and data source

This was a retrospective cohort study using the TriNetX Global Collaborative Network, a federated research platform designed for real-time analysis of deidentified patient data. The platform complies with HIPAA privacy standards and includes records from over 150 million individuals across 134 healthcare organisations (HCOs) worldwide. Participating HCOs span North America, Europe, Asia, the Middle East and other regions, providing a geographically diverse patient population. By integrating data from a diverse range of HCOs, TriNetX enables comprehensive real-world analyses that extend beyond the limitations of single-centre or region-specific studies. To ensure patient privacy, the platform obfuscates any patient counts ≤ 10 , grouping them into a single category. As the database exclusively contains deidentified information, the study did not require informed consent. All clinical data were

derived from electronic health records (EHRs) maintained by participating HCOs. This study adhered to the Strengthening the Reporting of Observational Studies in Epidemiology guidelines for observational research,¹¹ following established standards for methodological rigour and transparent reporting of findings.

Participants, eligibility criteria and exposure

The study population was identified from the TriNetX Global Collaborative Network by screening hospitalised patients aged 45–75 years over a 5-year period (31 October 2019–31 October 2024). Patients were stratified based on the presence or absence of diagnoses for MASLD (International Classification of Diseases, Tenth Revision (ICD-10): K76.0), MASH (ICD-10: K75.81) and liver fibrosis/cirrhosis (ICD-10: K74) using standardised ICD-10 codes. ICD-10 codes for CRC have demonstrated high validity in administrative datasets; for example, Hirano *et al* reported a positive predictive value of 86.3% (rising to 94.0% when using site-specific codes) for identifying CRC from hospital claims data.¹² The sample age range was chosen to align with the typical window for CRC screening, ensuring relevance to the study's outcomes. Three primary comparisons were made: (1) MASLD versus non-MASLD controls, (2) MASH versus non-MASH controls and (3) MASH with liver fibrosis/cirrhosis vs MASH without liver fibrosis/cirrhosis. Liver fibrosis/cirrhosis were identified using ICD-10 codes recorded in the EHR; histological or non-invasive imaging confirmation was not uniformly available across TriNetX and, therefore, was not used as part of the case definition. Exclusion criteria were applied to minimise confounding factors. Patients younger than 18 years were excluded, as were those with chronic liver diseases other than MASLD, MASH or liver fibrosis/cirrhosis. Excluded liver conditions included alcoholic and toxic liver diseases, hepatic failure, chronic hepatitis, autoimmune hepatitis and viral hepatitis, as well as other specific liver diseases (ICD-10: K70, K71, K72, K73, K75.0–K75.9, K76.1–K76.9, K77, B15–B19). Furthermore, individuals with any prior diagnosis of BCN (ICD-10: D12.2–D12.8) or malignant colorectal neoplasms (ICD-10: C18–C20) before the diagnosis of MASLD, MASH or liver fibrosis/cirrhosis were excluded. This ensured that differences in outcomes were not influenced by pre-existing colorectal pathology. The final study sample, selected after applying these criteria, is summarised in [figure 1](#), which outlines the flow of participant inclusion and exclusion.

Variables

Baseline demographic data included age, gender, race (categorised as white, black or African American, Asian, Native Hawaiian or Other Pacific Islander) and ethnicity (Hispanic or Latino, Not Hispanic or Latino). Clinical comorbidities were captured to account for metabolic and cardiovascular risk factors. These included hypertension (ICD-10: I10), diabetes mellitus (ICD-10: E08–E13), dyslipidaemia (ICD-10: E78), tobacco use (ICD-10: Z72.0), chronic ischaemic heart disease (ICD-10: I25), atrial

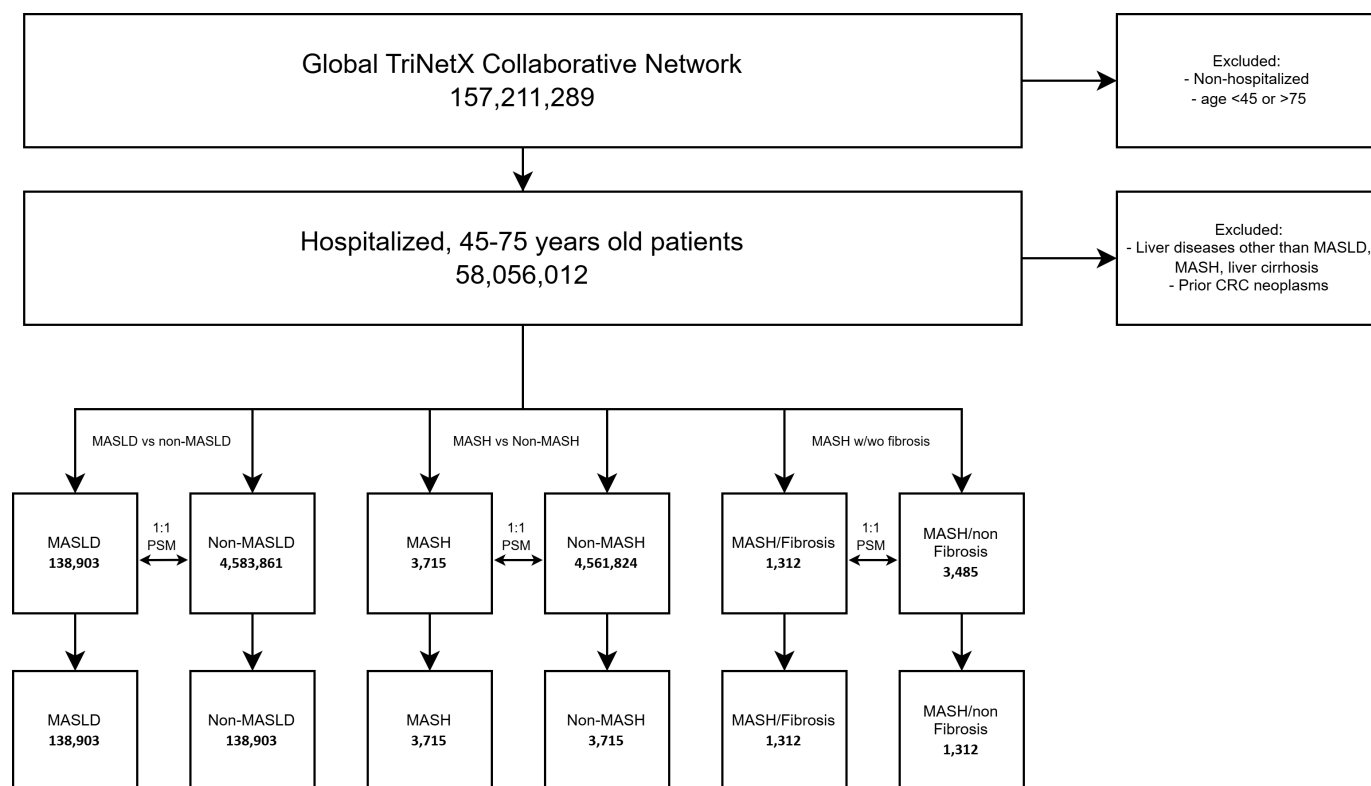


Figure 1 Flow chart of patient selection and propensity score matching. Selection of hospitalised adults aged 45–75 years from the TriNetX Global Collaborative Network (31 October 2019–31 October 2024). Exclusions included non-hospitalised patients, age <45 or >75 years, other chronic liver diseases and prior colorectal neoplasms. Three comparisons were analysed: MASLD versus non-MASLD, MASH versus non-MASH and MASH with versus without fibrosis/cirrhosis. Each exposure group was 1:1 propensity score-matched to its control group to balance demographics and metabolic comorbidities. Final matched sample sizes are shown. CRC, colorectal cancer; MASH, metabolic-associated steatohepatitis; MASLD, metabolic dysfunction-associated steatotic liver disease; PSM, propensity score matching.

fibrillation and flutter (ICD-10: I48), chronic lower respiratory diseases (ICD-10: J40–J4A), and alcohol-related disorders (ICD-10: F10). Medication usage was assessed for Glucagon-like peptide-1 (GLP-1) analogues (ATC code: A10BJ) and aspirin (RxNorm: 1191). Additionally, laboratory measures were extracted to evaluate metabolic and hepatic function, including body mass index (BMI) (TNX Curated: 9083), aspartate aminotransferase (AST) (TNX Curated: 9047), alanine aminotransferase (ALT) (TNX Curated: 9044), low-density lipoprotein (LDL) cholesterol (TNX Curated: 9002), triglycerides (TNX Curated: 9004) and high-density lipoprotein (HDL) cholesterol (TNX Curated: 9001). Comorbid conditions and laboratory values were captured using the most recent codes and results documented at or before the index MASLD/MASH diagnosis date.

Outcomes

The primary outcomes were the risk of CRC and BCN, including adenomas and hyperplastic polyps, in the subgroups of patients with MASLD, MASH and MASH with liver fibrosis/cirrhosis compared with disease-free controls. Additionally, a composite outcome combining CRC and BCN was analysed to provide a broader assessment of colorectal pathology risk. Outcomes were identified using the standardised ICD-10 codes previously

outlined. Patients were followed from the date of the index event, defined as the diagnosis of MASLD, MASH or MASH with liver fibrosis/cirrhosis, until the earliest of the following: the first recorded outcome (CRC or BCN), loss to follow-up or the conclusion of the study period on 31 October 2024.

Statistical analysis

Statistical analyses were conducted using the real-time analytics tool within the TriNetX Live platform. Baseline cohort characteristics were summarised as means and SD for continuous variables, while categorical variables were presented as counts and percentages. Continuous variables are presented as mean±SD as provided by the TriNetX analytic platform; because patient-level raw data are not accessible, formal testing for normality could not be performed. A one-to-one PSM was performed using a nearest-neighbour greedy matching algorithm with a calliper width of 0.25 times the SD to ensure close matching between groups. Variables used in the initial matching included age, sex, race and ethnicity. A second, separate PSM model was conducted for each subset of patients to additionally incorporate metabolic risk factors such as hypertension, diabetes mellitus, dyslipidaemia and BMI. This additional matching was designed to isolate the effect of MASLD, MASH and MASH with liver fibrosis/

Table 1 Baseline characteristics of MASLD, MASH, MASH with fibrosis/cirrhosis cohorts compared with controls

Characteristic	MASLD (N=1 38 903)	Non-MASLD (N=1 38 903)	MASH (N=3715)	Non-MASH (N=3715)	MASH with fibrosis/ cirrhosis (N=1312)	MASH without fibrosis/ cirrhosis (N=1312)
Demographics						
Age at Index (years, mean±SD)	58.0±8.6	58.0±8.6	58.9±8.6	58.9±8.6	62.6±7.7	62.5±7.7
Male (n, %)	64 561 (46.5%)	64 561 (46.5%)	1626 (43.8%)	1626 (43.8%)	531 (40.5%)	535 (40.8%)
White (n, %)	91 658 (66.0%)	91 658 (66.0%)	2576 (69.3%)	2576 (69.3%)	1037 (79.0%)	1043 (79.5%)
Black or African American (n, %)	12 936 (9.3%)	12 936 (9.3%)	276 (7.4%)	276 (7.4%)	72 (5.5%)	71 (5.4%)
Asian (n, %)	5120 (3.7%)	5120 (3.7%)	107 (2.9%)	107 (2.9%)	20 (1.5%)	13 (1.0%)
Native Hawaiian or Other Pacific Islander (n, %)	604 (0.4%)	604 (0.4%)	18 (0.5%)	18 (0.5%)	10 (0.8%)	10 (0.8%)
American Indian or Alaska Native (n, %)	407 (0.3%)	407 (0.3%)	15 (0.4%)	15 (0.4%)	10 (0.8%)	10 (0.8%)
Unknown race (n, %)	22 038 (15.9%)	25 736 (18.5%)	151 (4.1%)	60 (1.6%)	54 (4.1%)	31 (2.4%)
Other race (n, %)	6140 (4.4%)	2442 (1.8%)	572 (15.4%)	663 (17.8%)	124 (9.5%)	151 (11.5%)
Hispanic or Latino (n, %)	14 246 (10.3%)	6307 (4.5%)	331 (8.9%)	158 (4.3%)	129 (9.8%)	92 (7.0%)
Not Hispanic or Latino (n, %)	84 523 (60.9%)	70 302 (50.6%)	2492 (67.1%)	1877 (50.5%)	889 (67.8%)	934 (71.2%)
Unknown Ethnicity (n, %)	40 134 (28.9%)	62 294 (44.8%)	892 (24.0%)	1680 (45.2%)	294 (22.4%)	286 (21.8%)
Essential (primary) hypertension (n, %)	76 983 (55.4%)	34 873 (25.1%)	2101 (56.6%)	972 (26.2%)	728 (55.5%)	785 (59.8%)
Diabetes mellitus (n, %)	45 419 (32.7%)	16 978 (12.2%)	1494 (40.2%)	449 (12.1%)	660 (50.3%)	556 (42.4%)
Disorders of lipoprotein metabolism (n, %)	62 733 (45.2%)	26 151 (18.8%)	1802 (48.5%)	703 (18.9%)	585 (44.6%)	704 (53.7%)
Tobacco use (n, %)	10 327 (7.4%)	4379 (3.2%)	174 (4.7%)	115 (3.1%)	59 (4.5%)	59 (4.5%)
Chronic ischaemic heart disease (n, %)	24 877 (17.9%)	10 917 (7.9%)	764 (20.6%)	301 (8.1%)	343 (26.1%)	300 (22.9%)
Atrial fibrillation and flutter (n, %)	10 363 (7.5%)	5132 (3.7%)	326 (8.8%)	163 (4.4%)	194 (14.8%)	134 (10.2%)
Chronic lower respiratory diseases (n, %)	37 085 (26.7%)	15 894 (11.4%)	858 (23.1%)	458 (12.3%)	295 (22.5%)	334 (25.5%)
Alcohol related disorders (n, %)	11 359 (8.2%)	3942 (2.8%)	153 (4.1%)	100 (2.7%)	36 (2.7%)	48 (3.7%)
Glucagon-like peptide-1 analogues (n, %)	10 250 (7.4%)	2702 (1.9%)	353 (9.5%)	66 (1.8%)	139 (10.6%)	132 (10.1%)
Aspirin (n, %)	43 592 (31.4%)	19 774 (14.2%)	1192 (32.1%)	557 (15.0%)	429 (32.7%)	455 (34.7%)
BMI (kg/m ² , mean±SD)	33.8±8.4	30.4±7.8	34.8±8.9	30.4±7.9	35.2±8.4	34.2±8.5
Aspartate aminotransferase (U/L, mean±SD)	36.8±69.2	26.9±47.3	38.1±48.2	26.6±27.4	39.0±38.7	38.7±48.7
Alanine aminotransferase (U/L, mean±SD)	40.2±68.1	27.1±56.8	43.5±56.9	27.4±43.5	34.4±38.7	41.0±47.0

Continued

Table 1 Continued

Characteristic	MASLD (N=1 38 903)	Non-MASLD (N=1 38 903)	MASH (N=3715)	Non-MASH (N=3715)	MASH with fibrosis/ cirrhosis (N=1312)	MASH without fibrosis/ cirrhosis (N=1312)
Low-density lipoprotein (LDL) Cholesterol (mg/dL, mean±SD)	99.8±40.4	104.1±39.5	96.1±39.8	102.0±38.6	84.1±61.0	94.3±40.3
Triglycerides (mg/dL, mean±SD)	178.4±192.8	144.3±136.2	174.2±155.2	135.8±90.8	154.6±158.5	163.9±126.0
High-density lipoprotein (HDL) cholesterol (mg/dL, mean±SD)	46.3±18.1	48.2±21.0	44.7±19.0	48.1±21.5	43.1±16.2	47.6±18.3

BMI, body mass index; MASH, metabolic-associated steatohepatitis; MASLD, metabolic dysfunction-associated steatotic liver disease.

cirrhosis on the risks of CRC and BCN by accounting for the influence of these metabolic conditions. Following matching, Cox proportional hazards models were used to estimate HRs with 95% CIs for the risks of CRC, BCN and the composite outcome of colorectal neoplasms. A threshold of $p<0.05$ was used to determine statistical significance in all analyses.

RESULTS

Participant characteristics

From the Global TriNetX Collaborative Network of over 157 million patients, the study identified 58 056 012 hospitalised individuals aged 45–75 years. After 1:1 PSM, cohorts included 138 903 MASLD patients and controls, 3715 MASH patients and controls, and 1312 MASH patients with and without fibrosis/cirrhosis. [Table 1](#) summarises the baseline demographic and clinical characteristics of patients across the MASLD, MASH and MASH with or without fibrosis/cirrhosis cohorts and their respective matched controls.

The mean age at index was similar across cohorts, with MASLD and non-MASLD groups having a mean age of 58.0 ± 8.6 years, MASH and non-MASH groups averaging 58.9 ± 8.6 years and the MASH with and without fibrosis/cirrhosis subgroups showing slightly older means of 62.6 ± 7.7 and 62.5 ± 7.7 years, respectively. Male representation was consistent across comparisons, with proportions ranging from 40.5% to 46.5%. The majority of patients in all groups identified as white, with the highest percentage seen in the MASH with fibrosis/cirrhosis subgroup (79.0%).

Key metabolic risk factors were significantly more prevalent among MASLD and MASH patients compared with their respective controls. For instance, the prevalence of hypertension in MASLD was 55.4% vs 25.1% in non-MASLD patients, and in MASH, it was 56.6% compared with 26.2% in non-MASH. Similarly, diabetes mellitus was observed in 32.7% of MASLD patients and 40.2% of MASH patients compared with 12.2% and 12.1% in their non-MASLD and non-MASH counterparts, respectively.

Patients with MASH with fibrosis/cirrhosis showed a slightly higher prevalence of hypertension (55.5% vs 59.8%) and diabetes mellitus (50.3% vs 42.4%) compared with those without fibrosis/cirrhosis. Dyslipidaemia was also more common in MASLD (45.2% vs 18.8%), MASH (48.5% vs 18.9%) and MASH with fibrosis/cirrhosis (44.6% vs 53.7%) groups than their respective controls.

Metabolic dysfunction was further reflected in laboratory profiles, with MASLD and MASH groups showing elevated BMI ($33.8\pm8.4\text{ kg/m}^2$ and $34.8\pm8.9\text{ kg/m}^2$, respectively) compared with controls ($30.4\pm7.8\text{ kg/m}^2$ and $30.4\pm7.9\text{ kg/m}^2$). Liver enzymes, including AST and ALT, were higher in MASLD and MASH groups, with the highest values observed in MASH patients (AST: $38.1\pm48.2\text{ U/L}$, ALT: $43.5\pm56.9\text{ U/L}$). Lipid profiles showed higher triglycerides and lower HDL cholesterol in MASLD and MASH groups compared with controls, reflecting underlying metabolic disturbances. Interestingly, LDL cholesterol was lower in the MASH with fibrosis/cirrhosis subgroup ($84.1\pm61.0\text{ mg/dL}$) compared with the non-fibrosis/cirrhosis MASH group ($94.3\pm40.3\text{ mg/dL}$).

Outcomes

In the matched cohort of 138 902 MASLD and non-MASLD patients, without controlling for metabolic risk factors, MASLD was associated with significantly higher risks for CRC and BCN. The HR for CRC was 2.488 (95% CI 2.098 to 2.951, $p<0.001$), while for BCN, it was 2.954 (95% CI 2.790 to 3.128, $p<0.001$). For the composite outcome of CRC and BCN, the HR was 2.931 (95% CI 2.768 to 3.104, $p<0.001$). After controlling for metabolic risk factors, MASLD patients maintained elevated risks, with an HR of 2.708 (95% CI 2.291 to 3.201, $p<0.001$) for CRC, 2.501 (95% CI 2.378 to 2.631, $p<0.001$) for BCN and 2.505 (95% CI 2.380 to 2.636, $p<0.001$) for the composite outcome ([table 2](#), [figure 2](#)).

Among the MASH subgroup, the risks for CRC and BCN were substantially higher compared with non-MASH controls. Without controlling for metabolic factors, the HR for CRC was 6.288 (95% CI 1.418 to 27.882, $p=0.006$)

Table 2 Neoplasm outcomes in MASLD with and without metabolic risk factor control

Outcome	Without metabolic risk factor control			With metabolic risk factor control		
	MASLD (N=138903)	Non-MASLD (N=138903)		MASLD (N=138902)	Non-MASLD (N=138902)	
Benign neoplasms	4636 (3.3%)	1593 (1.1%)	2.954 (2.790 to 3.128)	4636 (3.3%)	2255 (1.6%)	2.501 (2.378 to 2.631)
CRC	474 (0.3%)	184 (0.1%)	2.488 (2.098 to 2.951)	474 (0.3%)	195 (0.1%)	2.708 (2.291 to 3.201)
Combined neoplasms	4582 (3.3%)	1581 (1.1%)	2.931 (2.768 to 3.104)	4582 (3.3%)	2211 (1.6%)	2.505 (2.380 to 2.636)
CRC, colorectal cancer; MASLD, metabolic dysfunction-associated steatotic liver disease.						

and for BCN, it was 2.484 (95% CI 1.694 to 3.644, $p<0.001$). The composite outcome showed an HR of 2.884 (95% CI 1.961 to 4.242, $p<0.001$). After adjusting for metabolic risk factors, the HR for CRC decreased to 5.032 (95% CI 1.429 to 17.718, $p=0.005$), while for BCN, it was 1.914 (95% CI 1.375 to 2.666, $p<0.001$), and for the composite outcome, it was 2.254 (95% CI 1.616 to 3.144, $p<0.001$) (table 3, figure 2).

For MASH patients with fibrosis/cirrhosis compared with those without, no significant differences were observed. Without controlling for metabolic risk factors, the HR for CRC was 1.062 (95% CI 0.230 to 4.910, $p=0.940$), while for BCN, it was 0.877 (95% CI 0.496 to 1.551, $p=0.650$), and for the composite outcome, it was 0.920 (95% CI 0.532 to 1.590, $p=0.766$). After adjusting for metabolic risk factors, the HRs remained non-significant, with CRC at 0.808 (95% CI 0.135 to 4.852, $p=0.815$), BCN at 0.899 (95% CI 0.503 to 1.607, $p=0.720$), and the composite outcome at 0.849 (95% CI 0.484 to 1.487, $p=0.566$) (table 4, figure 2).

DISCUSSION

This study highlights critical findings regarding the association between metabolic-associated fatty liver disease (MASLD), metabolic-associated steatohepatitis (MASH) and CRC as well as benign neoplasms. Patients with MASLD demonstrated a 2.7-fold higher risk of CRC and a 2.5-fold higher risk of benign neoplasms even after adjusting for metabolic risk factors such as hypertension, diabetes and obesity, indicating that MASLD is an independent risk factor for these outcomes. For patients with MASH, a more severe form of fatty liver disease, the risk of CRC was even more pronounced, with a fivefold increase compared with non-MASH controls, and nearly a twofold increased risk of benign neoplasms. Importantly, these elevated risks persisted after adjustment for metabolic factors, underscoring the need to distinguish MASLD from MASH in clinical practice, as disease severity appears to modulate CRC risk. MASLD and MASH are closely linked to obesity, insulin resistance and other metabolic risk factors, which are themselves established contributors to CRC risk.¹³ In our propensity-score-matched analysis adjusting for these factors, MASLD and MASH remained significantly associated with CRC, supporting an independent effect. This finding aligns with proposed mechanisms, such as hepatic and systemic inflammation, metabolic dysregulation and alterations in gut microbiota, that may create a pro-tumourigenic environment in MASLD and MASH.^{14–16} While these pathways are likely to influence both benign and malignant colorectal neoplasms, their intensity and persistence may play a greater role in driving progression from benign to malignant lesions, which could partly explain differences in observed risks.

Previous studies have suggested that liver fibrosis may be linked to increased risk of colorectal neoplasia. For example, Haimi *et al*¹⁷ reported that high FIB-4 scores were

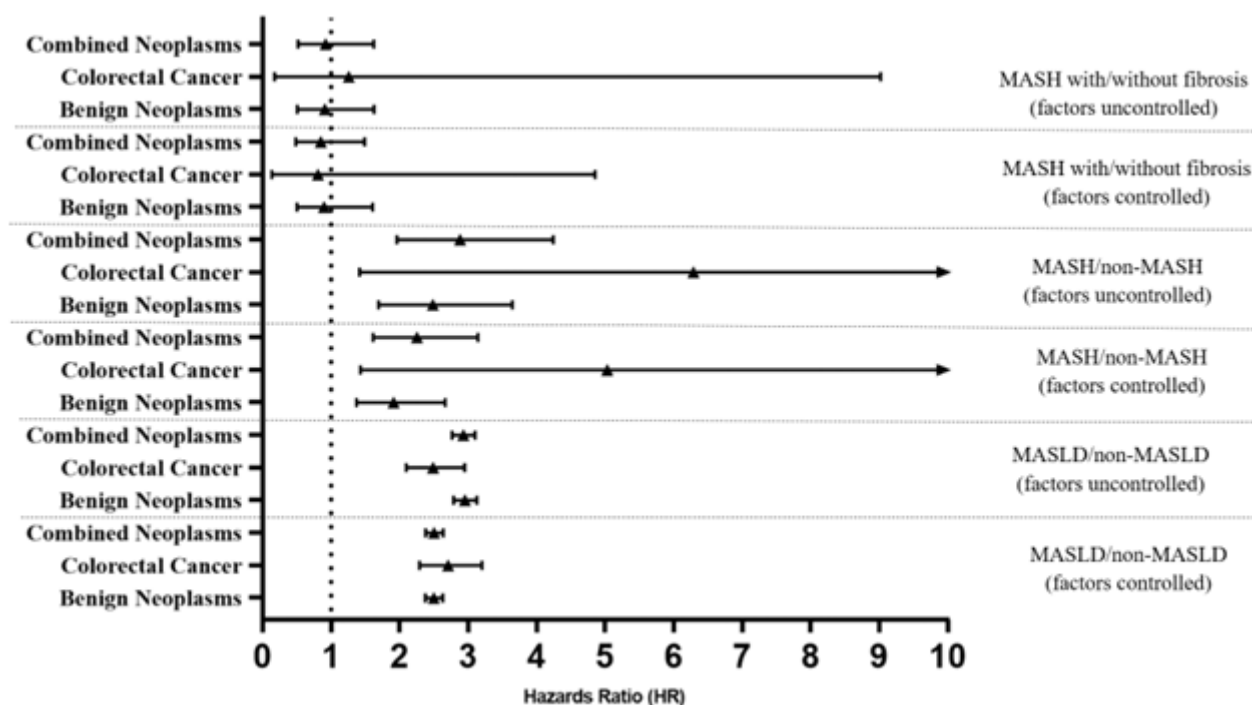


Figure 2 HRs for colorectal and benign colorectal neoplasms (BCN) across study groups. Forest plot showing HRs with 95% CIs for combined colorectal neoplasms, colorectal cancer (CRC) and BCN in MASLD versus non-MASLD, MASH versus non-MASH and MASH with versus without fibrosis/cirrhosis. Results are displayed for propensity score-matched analyses. HR >1 indicates higher risk in the exposure group; notably, MASLD and MASH remained significantly associated with CRC and BCN after matching and controlling for metabolic risk factors, whereas no significant difference was observed between MASH patients with and without fibrosis. MASH, metabolic-associated steatohepatitis; MASLD, metabolic dysfunction-associated steatotic liver disease.

associated with both prevalent and incident malignancies, including colon cancer (OR 2.75 over 10 years), and Kim *et al.*¹⁸ found that NAFLD patients with advanced fibrosis had significantly higher risks of colorectal adenoma compared with those without fibrosis. In contrast, in our large, globally representative cohort, we did not observe a significant difference in CRC or BCN risk between MASH patients with and without fibrosis/cirrhosis, even after controlling for metabolic risk factors. Several factors may explain this discrepancy: (1) our study included only hospitalised adults aged 45–75 years, aligning with CRC screening age and reducing detection bias; (2) we used 1:1 PSM to balance demographic and metabolic comorbidities, minimising residual confounding present in prior studies and (3) fibrosis/cirrhosis was defined by coded diagnoses rather than non-invasive scores such as FIB-4, which may capture a different patient subset. Taken together, these differences suggest that while fibrosis/cirrhosis is a key predictor of liver-related outcomes, its independent effect on colorectal neoplasia risk may be attenuated once metabolic risk factors are carefully balanced. Our findings support the view that the metabolic component of MASLD/MASH, rather than fibrosis itself, exerts a stronger influence on CRC risk.

Given that CRC is the second most common cancer worldwide and a leading cause of cancer-related mortality,¹⁹ these findings carry significant clinical implications. Our results suggest that patients with MASLD

or MASH may benefit from enhanced CRC screening strategies. Revisiting current clinical guidelines to incorporate these populations into high-risk groups for CRC screening could potentially reduce CRC-related mortality. Furthermore, periodic liver imaging in patients with risk factors such as hypertension, diabetes, hyperlipidaemia and obesity may aid in the early detection of MASLD/MASH, enabling timely interventions to mitigate downstream risks, including CRC.

Our study has several notable strengths. The use of the TriNetX Global Collaborative Network, encompassing over 150 million patients from 134 HCOs worldwide, ensures a large and diverse sample size, enhancing the generalisability of the findings. The study's focus on hospitalised patients aged 45–75 years aligns with CRC screening guidelines, ensuring clinical relevance. By leveraging real-world EHR data, the study provides insights applicable to routine clinical practice. However, several limitations must be acknowledged. First, the study does not account for the severity or duration of metabolic risk factors. For instance, uncontrolled diabetes, severe dyslipidaemia and morbid obesity may confer higher CRC risk than milder forms of these conditions. Second, as a retrospective cohort study, causal relationships cannot be definitively established. Despite the use of PSM, unmeasured confounders, such as dietary habits, specific alcohol intake, physical activity and family history of CRC may influence our results. The reliance on clinical

Table 3 Neoplasm outcomes in MASH cohorts with and without metabolic risk factor control

Outcome	Without metabolic risk factor control		With metabolic risk factor control		
	MASH (N=3715)	Non-MASH (N=3715)	HR (95% CI)	P value	MASH (N=3715) No MASH (N=3715) HR (95% CI)
Benign neoplasms	91 (2.4%)	37 (1.0%)	2.484 (1.694 to 3.644)	<0.001	1.914 (1.375 to 2.666)
CRC	13 (0.4%)	10 (0.3%)	6.288 (1.418 to 27.882)	0.006	5.032 (1.429 to 17.718)
Combined neoplasms	100 (2.7%)	35 (0.9%)	2.884 (1.961 to 4.242)	<0.001	2.254 (1.616 to 3.144)
CRC, colorectal cancer; MASH, metabolic-associated steatohepatitis.					

Table 4 Neoplasm outcomes in MASH with fibrosis/cirrhosis cohorts with and without metabolic risk factor control

Outcome	Without metabolic risk factor control		With metabolic risk factor control		
	MASH with fibrosis/cirrhosis (N=1312)	MASH without fibrosis/cirrhosis (N=1312)	HR (95% CI)	P value	MASH with fibrosis/cirrhosis (N=1312) MASH without fibrosis/cirrhosis (N=1312) HR (95% CI)
Benign neoplasms	19 (1.4%)	29 (2.2%)	0.908 (0.508 to 1.625)	0.746	0.899 (0.503 to 1.607)
CRC	10 (0.7%)	10 (0.7%)	1.259 (0.176 to 9.022)	0.819	0.808 (0.135 to 4.852)
Combined neoplasms	20 (1.5%)	30 (2.3%)	0.919 (0.520 to 1.623)	0.771	0.849 (0.484 to 1.487)
CRC, colorectal cancer; MASH, metabolic-associated steatohepatitis.					

records introduces potential biases such as selection bias and underreporting of conditions, potentially affecting the robustness of the findings. Additionally, reliance on ICD-10 codes introduces the potential for diagnostic misclassification. Because colonoscopy data were not available in TriNetX, we could not directly adjust for CRC screening practices. Variation in screening uptake, diagnostic intensity and local guidelines across participating HCOs could influence detection rates and potentially bias our observed associations, possibly leading to underestimation or overestimation of risk differences between groups. The cohort consisted primarily of hospitalised patients, which may not represent the general population, especially individuals with milder forms of MASLD or MASH. Inclusion of only hospitalised patients may reduce the applicability of findings to outpatient populations. This limits the generalisability of the results. Lastly, the 5-year follow-up period may be insufficient to capture the full spectrum of cancer progression or other long-term outcomes related to MASLD and MASH.

In conclusion, our findings emphasise the importance of MASLD and MASH as independent risk factors for CRC and benign neoplasms. This study highlights an opportunity to optimise CRC screening guidelines and preventive strategies in populations with metabolic liver diseases, ultimately improving gastrointestinal and overall health outcomes. Prospective, longitudinal studies are needed to confirm the causal relationship between MASLD/MASH and CRC or benign neoplasms. Such studies should account for the severity and duration of metabolic risk factors, as well as the degree of liver involvement, to provide a more nuanced understanding of these associations. Moreover, future research should explore how early interventions targeting MASLD/MASH might influence CRC risk and outcomes and incorporate information on potentially chemopreventive medications such as statins or metformin, which may influence cancer risk and further refine effect estimates.

Contributors Using the following codes: A=study design, B=data collection, C=statistical analysis, D=data interpretation, E=manuscript preparation, F=literature search, G=manuscript review. MAI: A, B, C, D, E, F, G. AO: D, E, F, G. MAY: D, E, F, G. MYM: A, D, E, G. KQ: D, E, F, G. W-KS: A, D, E, F, G. All authors reviewed and approved the final version of the manuscript. Guarantor: MAI accepts full responsibility for the integrity of the work as a whole, from inception to published article, and is the guarantor of the study.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Institutional Review Board (IRB) approval or exemption was not required for this study, as the TriNetX platform exclusively provides access to fully deidentified patient data in accordance with the HIPAA Privacy Rule, consistent with prior observational studies using the TriNetX platform. This ensures that no information capable of identifying individual patients is included, maintaining compliance with ethical standards for research involving secondary data analysis. For any questions about research integrity or ethical oversight at our institution, please contact the Human Research Protection Office/Institutional Review Board,

Washington University in St. Louis at hrpo@wustl.edu, or the TriNetX Network team via <https://trinetx.com/contact/>.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. This study used data from the TriNetX Global Collaborative Network, a federated research platform that provides access to deidentified patient information from participating healthcare organisations. Access to TriNetX is subscription-based and provided under institutional licensing agreements. The authors analysed aggregated data within the platform under their institutional access.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <https://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Mohammad Aldiabat <https://orcid.org/0000-0002-4536-2799>

REFERENCES

- Bray F, Laversanne M, Sung H, *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024;74:229–63.
- Rinella ME, Sookoian S. From NAFLD to MASLD: updated naming and diagnosis criteria for fatty liver disease. *J Lipid Res* 2024;65:100485.
- Mantovani A, Dauriz M, Byrne CD, *et al.* Association between nonalcoholic fatty liver disease and colorectal tumours in asymptomatic adults undergoing screening colonoscopy: a systematic review and meta-analysis. *Metab Clin Exp* 2018;87:1–12.
- Wong VW-S, Wong GL-H, Tsang SW-C, *et al.* High prevalence of colorectal neoplasm in patients with non-alcoholic steatohepatitis. *Gut* 2011;60:829–36.
- Shen H, Lipka S, Kumar A, *et al.* Association between nonalcoholic fatty liver disease and colorectal adenoma: a systemic review and meta-analysis. *J Gastrointest Oncol* 2014;5:440–6.
- Pan S, Hong W, Wu W, *et al.* The relationship of nonalcoholic fatty liver disease and metabolic syndrome for colonoscopy colorectal neoplasm. *Medicine (Baltimore)* 2017;96:e5809.
- Zeng Y, Cao R, Tao Z, *et al.* Association between the severity of metabolic dysfunction-associated fatty liver disease and the risk of colorectal neoplasm: a systematic review and meta-analysis. *Lipids Health Dis* 2022;21:52.
- Zhao J-F, Zhou B-G, Lv Y, *et al.* Association between metabolic dysfunction-associated steatotic liver disease and risk of colorectal cancer or colorectal adenoma: an updated meta-analysis of cohort studies. *Front Oncol* 2024;14:1368965.
- Celsa C, Pennisi G, Tulone A, *et al.* Risk of hepatic and extrahepatic outcomes associated with metabolic dysfunction-associated steatotic liver disease and metabolic dysfunction and alcohol-associated steatotic liver disease: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* 2025;10:998–1012.
- Loomba R, Sanyal AJ. The global NAFLD epidemic. *Nat Rev Gastroenterol Hepatol* 2013;10:686–90.
- von Elm E, Altman DG, Egger M, *et al.* The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344–9.
- Hirano T, Negishi M, Kuwatsuru Y, *et al.* Validation of algorithms to identify colorectal cancer patients from administrative claims data of a Japanese hospital. *BMC Health Serv Res* 2023;23:274.
- Schoen RE, Tangen CM, Kuller LH, *et al.* Increased blood glucose and insulin, body size, and incident colorectal cancer. *J Natl Cancer Inst* 1999;91:1147–54.
- VanSaun MN, Lee IK, Washington MK, *et al.* High fat diet induced hepatic steatosis establishes a permissive microenvironment for colorectal metastases and promotes primary dysplasia in a murine model. *Am J Pathol* 2009;175:355–64.
- Mantovani A, Lonardo A, Stefan N, *et al.* Metabolic dysfunction-associated steatotic liver disease and extrahepatic gastrointestinal cancers. *Metab Clin Exp* 2024;160:156014.
- Kalligeros M, Henry L, Younossi ZM. Metabolic dysfunction-associated steatotic liver disease and its link to cancer. *Metab Clin Exp* 2024;160:156004.

- 17 Haimi M, Baker FA, Vinker S, *et al.* Association Between High FIB-4 Score and the Risk of Malignancy Development and Mortality: A Retrospective Longitudinal Case-Control Study. *Dig Dis Sci* 2025;70:3525–37.
- 18 Kim MC, Park JG, Jang BI, *et al.* Liver fibrosis is associated with risk for colorectal adenoma in patients with nonalcoholic fatty liver disease. *Medicine (Baltimore)* 2019;98:e14139.
- 19 Parkin DM, Bray F, Ferlay J, *et al.* Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74–108.