RESEARCH ARTICLE

Phenotyping disease progression in Metabolic Dysfunction-Associated Steatohepatitis (MASH): Insights from medical claims data and clinical trial representation

Magda Berberova*, Petar Nikolov, Deepa Kumar, Liuqiang Lu

Berberova M, Nikolov P, Kumar D, et al. Phenotyping disease progression in Metabolic Dysfunction-Associated Steatohepatitis (MASH): Insights from medical claims data and clinical trial representation. J Hepato Gastroenterol. 2026;10(1):1-5.

ABSTRACT

Background and aims: This study evaluated the U.S. MASH population using medical records to characterize disease features, identify fast-progressing phenotypes, and compare them with late-phase clinical trial cohorts.

Methods: A retrospective cohort analysis of anonymized U.S. medical and prescription records (Oct 2018–Sep 2020) identified newly diagnosed MASH patients, defined by no prior MASH or cirrhosis records within two years. Progression was assessed by cirrhosis-related conditions recorded within three years post-diagnosis, classifying patients as fast progressors or others. Descriptive statistics, significance testing ($p<0.05$), and progression analyses were performed. Population characteristics were compared with late-phase clinical trial cohorts.

Results: Among 38,823 newly diagnosed MASH cases, 5,993 (15%) met the fast progressors definition. Over half were female, with significantly higher female representation and older age in fast progressors (61.3 ± 13.1 *vs.* 53.7 ± 14.2). Comorbidities were more frequent in this group: Type 2 diabetes (40% *vs.* 28%), dyslipidemia (40% *vs.* 37%), anemia (16% *vs.* 9%), coronary artery disease (12% *vs.* 7%), and heart failure (5% *vs.* 1%). Clinical trial cohorts showed similar gender and age profiles, higher metabolic comorbidities, and lower cardiovascular comorbidities. Obesity, age, and female gender were identified as independent risk factors for progression.

Conclusion: Approximately 15% of MASH patients exhibit rapid progression with distinct clinical features, aligning well with clinical trial representation and comparability.

Key Words: *MASH (Metabolic Dysfunction-Associated Steatohepatitis); Fast progressors; Disease progression; Cirrhosis risk; Comorbidities; Obesity; Gender differences; Age-related risk; Clinical trial comparability, Real-world evidence*

INTRODUCTION

The major liver-related clinically important sequelae of Metabolic Dysfunction-Associated Fatty Liver Disease (MAFLD) is liver inflammation and fibrosis progression in Metabolic Dysfunction Associated Steatohepatitis (MASH) to overt liver cirrhosis and it's complications [1].

Global epidemiology data show broad distribution of MAFLD in around 25% of global population, with the estimation of around 5 to 10% of the affected population to develop cirrhosis-related complications with risk of death or liver transplantation need [2]. While Cardiovascular Diseases (CVD) and non-hepatic malignancies are the most common causes of mortality in patients with MAFLD without advanced fibrosis, the rates of fibrosis progression and hepatic decompensation vary depending on baseline disease severity, genetic, individual environmental, and comorbid disease determinants. Fibrosis and the presence of steatohepatitis are the primary predictors of liver disease progression in MASH when compared to other fibrosis developing chronic liver diseases. Currently the MASH - driven progression to cirrhosis is considered less predictable, and the speed of progression and its risk factors are a subject of thorough investigation in the recent years to identify the fibrogenic and inflammatory drivers of disease progression [1,3]. The identification of clinical factors associated with MASH progression would support clinical practitioners for more targeted medical care for those MASH patients who are at higher or imminent risk of disease progression [4-7].

At the same time, both obesity, metabolic syndrome, MAFLD and MASH are currently part of intensive research for investigating deeper their pathophysiology, progression risk factors and potentially effective therapies and a lot of clinical research trials are conducted in the Metabolic syndrome liver-related diseases like MAFLD and MASH. In addition to the above, and as part of a global CRO team conducting a significant proportion of clinical research in the field of MAFLD and MASH we would consider valuable to compare the available data from general MASH population medical claims and late stages clinical research population of MASH and try to assess the extent of representation and comparability between the 2 [8,9].

MATERIALS AND METHODS

We have searched the ICON's United States (US) medical and prescription longitudinal Health Insurance Portability and Accountability Act (HIPAA) compliant anonymized longitudinal database capturing healthcare claims for approximately 300M patients annually for medical claims features and diagnoses reported to try to define and assess the phenotype of progressing MASH in comparison to the general MASH population in US.

We have assessed all available newly diagnosed MASH patient data records from October 2018 to September 2020 to collect age, gender and observed conditions information, applying the case selection criterion of only new diagnoses confirmation by the absence of MASH, other liver diseases and cirrhosis-related conditions, defined as any hepatic event like ascites, oesophageal varices, portal hypertension bleed (varices or gastropathy) and hepatic encephalopathy claim records in the two years prior to the MASH diagnosis. We are putting the MASH initial diagnosis date of each case as an index date. We have then followed the eligible population of newly diagnosed MASH for 3 years after the index date in regards of development of cirrhosis-related condition records, defined as any of the ascites, oesophageal varices, portal hypertension bleed (varices or gastropathy) and hepatic encephalopathy event claim record to assess the progression pace and progressing population profile. We are basing the assessment on applying the categorization of fast progressors for the population with at least 1 cirrhosis-related conditions record in the 3-years follow up period

ICON plc, Leopardstown, Dublin 18, Ireland

Correspondence: *Magda Berberova, ICON plc, Leopardstown, Dublin 18, Ireland; E-mail: Magda.Berberova@iconplc.com*

Received: *18-Mar-2026, Manuscript No. PULHG-26-7734;* ***Editor assigned:*** *20-Mar-2026, PreQC No. PULHG-26-7734 (PQ);* ***Reviewed:*** *03-Apr-2026, QC No. PULHG-26-7734;* ***Revised:*** *17-Apr-2026, Manuscript No. PULHG-26-7734 (R);* ***Published:*** *24-Apr-2026, DOI: 10.37532/pulhg.26.10.1.1-5*

OPEN ACCESS

This open-access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC) (http://creativecommons.org/licenses/by-nc/4.0/), which permits reuse, distribution and reproduction of the article, provided that the original work is properly cited and the reuse is restricted to noncommercial purposes. For commercial reuse, contact reprints@pulsus.com

versus Rest of the population with no cirrhosis-related conditions recorded in that period (Figure 1).

For the MASH features comparison between the general MASH population and the MASH patients participating in late phase clinical research trials we have compared the data of the medical records described above to the already published data [10,11] from late stage phase III clinical research study in a close to the claims data collection time period of enrolment within the period March 2019 to July 2021 [11].

Statistical significance was determined by examining p-values less than 0.05. Descriptive statistics, significance tests (p-value<0.05) of findings specific parameters/morbidities and speed of progression, and correlation analyses are performed. ICON's US medical and prescription longitudinal anonymized claims data was leveraged for this analysis.

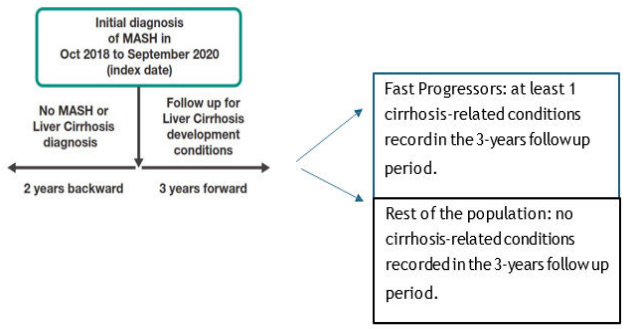

Figure 1) *Medical records assessment algorithm*

RESULTS

A total of 38,823 individual records with newly diagnosed MASH have been identified for the assessment. For the 3 years medical records follow up of this group we have been able to identify a subpopulation of 5993 cases (15%) with at least 1 cirrhosis-related conditions record after the index date, meeting the fast progression definition. The age and gender of fast progressors differ significantly from the rest of the population, cases identified as fast progressors were significantly older at index with the mean age of 61.3 ± 13.1 *vs.* 53.7 ± 14.2. and a higher proportion (58% *vs.* 53%)of patients among the fast progressors were female (Figure 2).

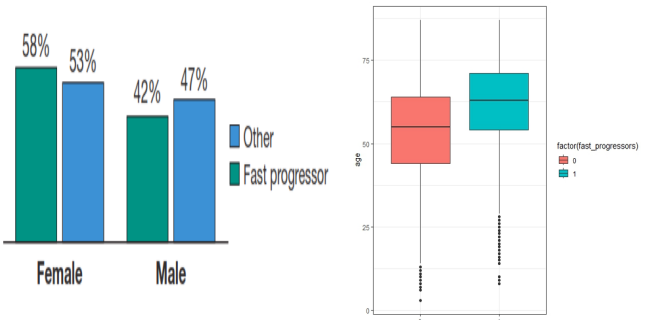

Figure 2) *Gender and age distribution between the fast progressors and rest of the population*

We have assessed the character and frequency of co-morbidities at index date and if any differences in the morbidity profiles between the groups. Among the most frequent observed comorbidities are the Type 2 Diabetes Mellitus (T2DM, 40% *vs.* 28%) and dyslipidemia (40% *vs.* 38%), followed by anemia (16% *vs.* 9%), coronary artery disease (12% *vs.* 7%) and heart failure (5% *vs.* 1%) presented significantly more in the Fast Progressing group (Figure 3).

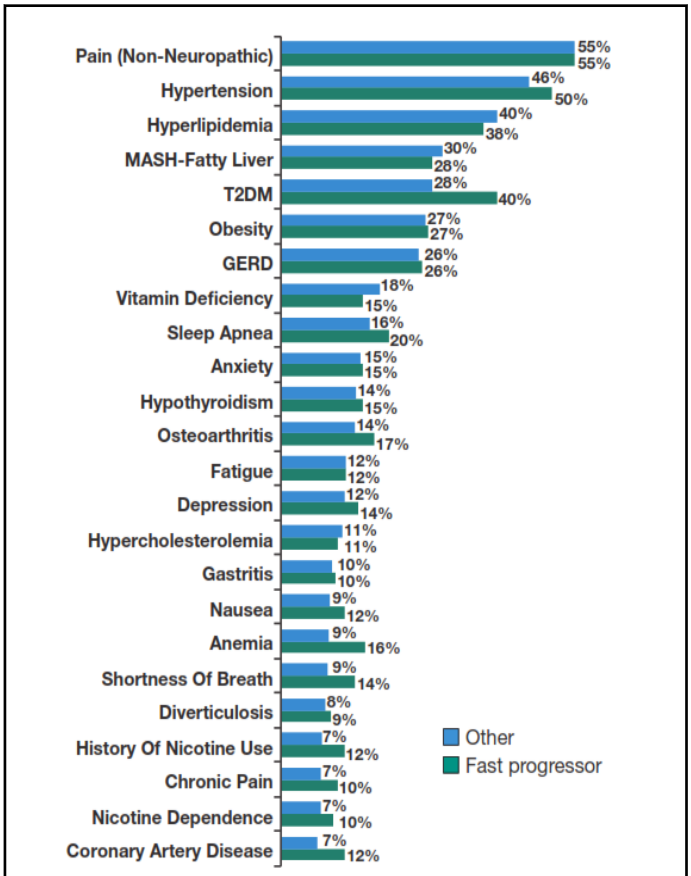

Figure 3) *Top-down list of observed co-morbidities and difference between the fast progressors and rest of the population*

Leading body systems affected in the fast progressors with significant magnitude of difference to the rest of MASH population are assessed as well. We've observed statistically significant differences between fast progressors and non-fast progressors for the frequency of following individual and body-system comorbidities tested:

- Anemia 16% *vs.* 9%
- Coronary artery disease 12% *vs.* 7%
- Heart failure 5% *vs.* 1%
- T2DM 40% *vs.* 28%
- Dyslipidemia 40% *vs.* 37%
- CV disease (general) 30% *vs.* 20%
- Hemostatic disorders 10% *vs.* 2%
- Nervous system disorders 34% *vs.* 29%
- Renal system disorders 18% *vs.* 12%
- Respiratory disorders 36% *vs.* 29%

The statistically different data between the groups show that the comorbidities and progressing status of the MASH are related.

In addition, we have looked for some correlations between frequent co-morbidity parameters within the groups and assessed if there are any correlations difference on comparison of fast progressors versus rest of the MASH population.

While the correlations found were weak within a group for the parameters of anemia-dyslipidemia, anemia-T2DM and hemostatic-respiratory disorders, we have found a statistically significant difference of those correlations when compared between the Fast-progressing MASH and Rest of the MASH population (Table 1).

Table 1: Correlation of anemia to dyslipidemia, anemia to T2DM and hemostatic to respiratory disorders within and between the groups

Parameter	Correlation within the group	Correlation difference between the groups (Fisher's z-test)	Difference magnitude between the groups (Zou's confidence intervals)
Anemia and dyslipidemia	Weak	z=3.55, p=0.0004	CI: 0.0219 to 0.0753
Anemia and T2DM	Weak	z=2.2273, p=0.0259	CI: 0.0037 to 0.0574
Hemostatic and respiratory disorders	Weak	z=5.26, p < 0.05	CI: 0.0461 to 0.1006

We have specifically investigated cases with medically diagnosed obesity and searched if any clinically relevant and statistically valid differences. Table 2 presents the number of diagnosed obesity cases and the mean and median age differences if diagnosed with obesity or not within the 2 groups. When comparing the mean of factors, the test result indicates there are significant difference in age between factors (p<.05), however the size of the effect is small (ges=0.02)<0.06.

Table 2: Presence of medically diagnosed obesity within the groups assessed

	Mean age (std)	Median age (years)	Count of cases
Non-fast progressor without obesity	54.2 (14.2)	56	24,048
Non-fast progressor with obesity	52.4 (14.3)	54	8,782
Fast progressor without obesity	62.4 (12.9)	64	4,380
Fast progressor with obesity	58.5 (13.2)	60	1,613

Age is significantly lower in concomitant Obesity cases versus no obesity diagnosed for both groups of fast progressors and rest of the population, ((FP w/ *vs.* w/o obesity mean age 58.5 *vs.* 62.4 years) and (Rest of the population w/ *vs.* w/o obesity mean age 52.4 *vs.* 54.2 years)), factoring obesity as a speeding MASH disease parameter (Figure 4).

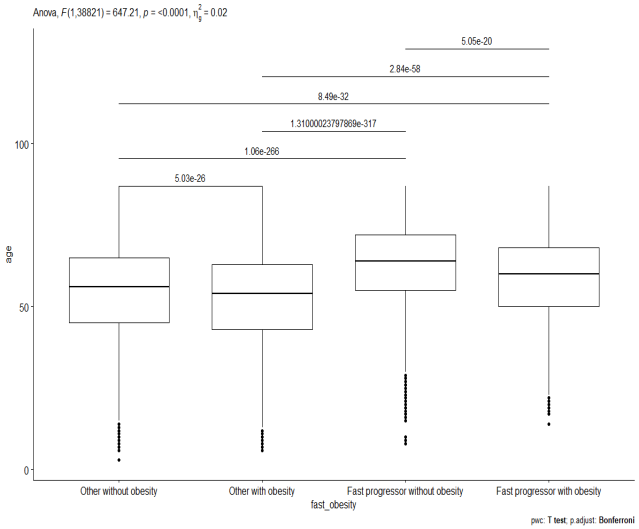

Figure 4) *ANOVA plot of diagnosed obesity within the groups assessed*

Our analyses show that when holding age unchanged, and sex status remains constant, diagnosed obesity increases the odds of MASH progressing fast (versus non-fast progression) by a factor of 1.1. When comorbidities status remains unchanged, and sex factor remains constant, one year of aging also increases the odds of MASH fast progression compared to non-fast progression by a factor of 1.04. Additionally, when co-morbidities status and age remain unchanged, the factor of being female increases the odds of MASH progressing fast (versus non-fast progression) by a factor of 1.08 (fit model predictions close to 70% correct (AUC=0.66), Table 3. Factoring each of the 3 parameters independently to the other 2 is showing that each of the parameter's obesity, female gender and age is independently associated with MASH disease progression.

Table 3: Factoring, odds ratios and AUC of age, gender and obesity analysis. Area under the Curve (AUC) 0.6558

Parameter	P value	Odds ratio	2.5%	97.5%
Age	0.04163191	1.04251067	1.04018569	1.0448567
Female	0.07353916	1.07631069	1.01651421	1.1397293
Obesity	0.09611455	1.10088517	1.03284404	1.1729721

When comparing the data presented above to the MASH phase III clinical study population demographic and clinical characteristics (11, Table 1) the mean age of the clinical study trial participants of 56.6 (± 10.9) years is close to the mean age of the non-progressing population we found (53.7 ± 14.2 years) and lower than the mean age of the fast progressors (61.3 ± 13.1 years), while the gender representation with clinical trial females gender in 56% is close to the female representation in the medical claims – based

database (58% females in the fast progressors group and 53% in the rest of the population).

There are some differences in the incidence of metabolic risk factors and co-morbidities, especially for T2DM, hypertension and dyslipidemia, represented in 67%, 78.1% and 71.3% of the clinical trial population and in only 40%, 50% and 40% of the fast progressors and 28%, 46% and 38%of the rest of the population, respectively (Table 4).

Table 4: Most frequent metabolic co-morbidities distribution between medical claims- based medical records and late phase clinical research MASH population (adapted from 11)

Co-morbidity/Percentage	MASH Clinical research population	Medical claims-database fast progressors	Medical claims-database rest of the population
T2DM	67%	40%	28%
Hypertension	78.10%	50%	46%
Dyslipidemia	71.30%	40%	38%

While we found medical records of obesity in only 27% of the MASH medical claims database, the mean BMI of the phase III MASH research population reported is 35.7 ± 6.8, potentially showing the researched population as significantly more obese than the general population of diagnosed MASH. Together with the potential under-reporting of obesity as medical claim/diagnosis as potential explanation, our analyses above show that both obesity, age and female gender, all well represented in the late phase clinical research study, are an independent risk factors for the MASH disease progression.

Regarding the other metabolically related co-morbidity of cardiovascular diseases concomitant Atherosclerotic Cardiovascular Disease (ASCVD) is reported in 3% of the MASH clinical trial population, while the medical claims-based database reports total cardiovascular diseases frequency of 25%, and CAD specifically in 12% of the fast progressors and in 7% of the rest of the MASH population. This discrepancy might potentially be explained by a kind of protective pre-selection bias for clinical studies participation of relatively more "healthy" MASH patients in regards of less advanced co-morbidities within the everyday clinical practice.

While we found similar frequency of reported hypothyroidism in the medical claims database (14-15%) and in the clinical trial population (13.4%), we were not able to compare the representation of anemia, as not reported in the published clinical study data, but with significant difference of representation in the fast progressors versus rest of the MASH medical claims-based population, even with relatively low frequency, when compared to the metabolically related co-morbidities (16% in the fats progressors group versus 9% in the rest of the MASH population).

DISCUSSION

The data analyzed are based on a significant number of MASH diagnoses, 38,823 individual medical records, between October 2018 to September 2020. We have found 15% of this population to show at least one liver cirrhosis – related condition diagnosed in a 3-years follow up after the initial MASH diagnosis, defining that subpopulation as fast progressors. The proportion of medical records on disease progression within three years follow up period is within the ranges of the general epidemiological data published on disease progression and the estimation of 5% to 10% of the general MAFLD population to develop liver cirrhosis – related complications and liver-related death or transplantation [2].

As our analysis is based on a clinical features of cirrhosis development disease progression recorded by healthcare practitioners and considering the stage of liver fibrosis histological assessment as the widely accepted surrogate endpoint for liver disease progression in MASH, the 15% liver cirrhosis related medical features recorded in 3 years follow up period is smaller than the fibrosis progression of 34.56% found on paired liver histology assessment data meta-analysis of 7 eligible for the assessment studies in 116 patients within an at least 1-year apart histological follow up [6]. We are acknowledging the potential confounding of the medical records-based data on the initially recorded diagnosis of MASH, put as an index date in our analysis, to content contemporary and delayed formal MASH diagnosis recording of the individual cases and thus to affect the pace and progression assessments. We consider the exclusion of cases with any MASH, other liver disease or cirrhosis - related medical records in the previous 2 years period before the index date and the extensive number of cases assessed as an attempt to minimize the late diagnosis potential confounding in our analysis, being aware that the progression versus late diagnosis is a commonly found weakness of MASH disease analyses including on big real world longitudinal data base cases explored by a machine learning technologies [5].

Another limitation of our data is the potential heterogeneity of the medical records database, as it includes all newly MASH diagnosed cases in the defined 2 years-time period who meet the eligibility criteria for the analysis. Retrospective cohorts analyses based on medical diagnosis records and ICD codes is a standard approach for large scale epidemiology reviews in a specific outcomes research for identification of the potential high risk factors, including in regards of the co-morbidities profile associated with those outcomes [12,13]. Considering those are real world medical records data the incremental heterogeneity has some potential advantages of assessing the typical clinical features, like age, gender and co-morbidities, that for such a large group of data might provide some epidemiological directions of the clinical profile of fast progressing MASH.

When comparing the comparability and how representable a late phase clinical trial in MASH population is compared to the medical claims-based data of the general MASH diagnosed population in US, we found good representation parameters of mean age with a slight deviation towards younger lower progression risk age, and for gender distribution, knowing the typical challenges of females enrolment in clinical studies. Not surprising, the typical metabolic syndrome-related MASH co-morbidities, like T2DM, hypertension and dyslipidemia are much more represented in the clinical study population, this might be explained by both patient's sourcing that comes from the co-morbidities medical care and referring for their liver disease management, as well as, by the clinical management and prioritization of patients showing the broader clinical features of metabolic syndrome and at increased risk of diseases speed up and progression. As our analysis conclusions confirm, metabolic co-morbidities are related to the faster progression of the liver face of the metabolic syndrome MASH. Our comparative analysis shows some discrepancies between the late phase clinical trial in MASH population features and the medical claims-based data of the general MASH diagnosed population in regards of both obesity and cardiovascular diseases. Considering obesity, the medical records database analyzed does not include data about BMI itself but only provide if a medical diagnosis of obesity as a disease is recorded. That database specificity could explain the relatively lower percentage of diagnosed obesity in the general MASH population compared to the mean BMI data in the late phase MASH clinical study. On contrast, the general MASH diagnosed population shows significantly higher frequency of both cardiovascular diseases bucket and CAD itself, when compared to the clinical trial population, that we could consider to be driven by the safety focused preventive pre-selection by physicians of MASH patients for a clinical trial, tending to a relatively good general medical status. That typical observation in clinical research is usually coped by the post-marketing full spectrum clinical practice patients follow up and surveillance. Importantly, the MASH features comparison performed showed that the 3 independent risk factors for the MASH disease progression of obesity, age and female gender that we have identified in our analyses are relatively well represented in the

late phase clinical research study, indirectly supporting the relevance of the high medical need MASH population selected for the investigations.

CONCLUSION

Given the more holistic approach we have been focused in regards of the interconnections between multiple parameters observed in our analysis and the different clinical features interrelations, as those between age, gender and obesity, and MASH comorbidities as not isolated variables but part of a dynamic system where each element influences and amplifies the others. The progression of MASH is not driven by a single factor but emerges from a network of interrelated conditions, echoing the philosophical insight that everything is connected and leads to one another. Understanding these relationships is essential for identifying high-risk patients and tailoring interventions that reflect the complexity of real-world clinical profiles.

We hope our holistic approach and the relations we have provided to be useful for clinical practice and clinical research in MASH, and, as introspection reveals inner resilience, be part of the ever-living paradigm, phrased in the classical Marcus Aurelius Meditations (Book VII, Section 59), "Look well into thyself; there is a source of strength which will always spring up if thou wilt always look."

CONFLICTS OF INTEREST

No conflicts of interest of the authors.

REFERENCES

1. Rinella M, Neuschwander-Tetri B, Shadab Siddiqui M, et al. AASLD practice guidance on the clinical assessment and management of nonalcoholic fatty liver disease. Hepatology. 2023;77:1797–835.
2. Kechagias S, Nasr P, Blomdahl J, et al. Established and emerging factors affecting the progression of nonalcoholic fatty liver disease. Metabolism. 2020;111:154183.
3. Schuppan D, Surabattula R, Wang X. Determinants of fibrosis progression and regression in NASH. J Hepatol. 2018;68(2):238–50.
4. Brookhart MA. Estimating causal effects of treatments in real-world data. 7th edition. Paris NASH Meeting. 2021.
5. Schattenberg JM, Balp MM, Reinhart B, et al. NASHmap: Clinical utility of a machine learning model to identify patients at risk of NASH in real-world settings. Sci Rep. 2023;13(1):5573.
6. Singh S, Murad MH, Allen AM, et al. Fibrosis progression in nonalcoholic fatty liver versus nonalcoholic steatohepatitis: A systematic review and meta-analysis of paired-biopsy studies. Clin Gastroenterol Hepatol. 2015;13(4):643–54.
7. Parameswaran M, Hasan HA, Sadeque J, et al. Factors that predict the progression of nonalcoholic fatty liver disease (NAFLD). Cureus. 2021;13(12):e20776.
8. Shaikh A, Pedra G, Ruiz-Casas L, et al. Risk factors for fibrosis progression in non-alcoholic steatohepatitis: Analysis of the European real-world cohort in the GAIN study. Gastroenterol Hepatol. 2024;47:463–72.
9. Chen VL, Morgan TR, Rotman Y, et al. Resmetirom therapy for metabolic dysfunction-associated steatotic liver disease: October 2024 updates to AASLD practice guidance. Hepatology. 2025;81(1):312-20.
10. Harrison SA, Taub R, Neff GW, et al. Resmetirom for non-alcoholic fatty liver disease: A randomized, double-blind, placebo-controlled phase 3 trial. Nat Med. 2023;29:2919–28.
11. Harrison SA, Bedossa P, Guy CD, et al. A phase 3, randomized, controlled trial of resmetirom in NASH with liver fibrosis. N Engl J Med. 2024;390:497-509.
12. Ochoa-Allemant P, Hubbard RA, Kaplan DE, et al. Adverse liver outcomes, cardiovascular events, and mortality in steatotic liver disease. JAMA Intern Med. 2025;185(8):986-95.
13. Issa G, Shang Y, Strandberg R, et al. Cause-specific mortality in 13,099 patients with metabolic dysfunction-associated steatotic liver disease in Sweden. J Hepatol. 2025;83:643–51.