Findings From a Non-Alcoholic Steatohepatitis (Nash) Cohort Developed via Artificial Intelligence in a Large Representative Population in the U.S.

Bandaria J, Boussios C, Donadio G, Li F, Mortimer K, Gliklich R, Starzyk K | OM1, Inc, Boston, MA, USA

Background

- Although tens of millions of people are estimated to have nonalcoholic steatohepatitis (NASH) in the US, a small fraction of those are readily identifiable in existing health-related data, which hampers epidemiologic studies and clinical development[1]
- Liver biopsy remains the gold standard for diagnosing NASH but carries risk, is contraindicated for some patients and is not feasible for longitudinal disease assessment. Even in the presence of biopsy results, there remain inconsistencies regarding how histologic parameters are documented, interpreted and translated to patient management [2]
- Despite recent progress in the development of non-invasive diagnostic and staging scores, biomarker panels and radiological approaches to identify and monitor patients, there remains a substantial proportion of patients that are either not diagnosed as having NASH or not documented as such
- Identification of both probable and confirmed NASH is limited by the frequent lack of symptoms, under-diagnosis, absence of targeted treatments and poor billing code specificity (a specific ICD code was not available until late 2016)
- Better characterization of patients likely to have transitioned to NASH is needed to improve patient identification and management early in the disease process and elucidate the natural history of the disease overall and amongst key subphenotypes

Methods

- The OM1 Data Cloud (OM1, Boston, MA) collects, links and leverages structured and unstructured data from electronic medical records, medical and pharmacy claims and other sources in an ongoing and continuously updating manner. Patient level data are deterministically linked to provide the most comprehensive longitudinal patient journeys for algorithm development
- The data include extensive clinical and claims data on patients seen in a variety of provider practice types across the US, which is important in understanding the multi-systemic burden of the disease and diagnostic pathway
- As previously reported, we used a combination of sophisticated artificial intelligence (AI) algorithms to characterize the likelihood that a patient who is not an ICD10-identified NASH patient is in fact a NASH patient
- We initially applied this algorithm to a sample of over 44 million obese and/or diabetic patients [receiver operating characteristic (ROC) area in out-of-sample 0.86]. The algorithm was then applied to a larger unselected population of patients in the U.S. of approximately 240 million.
- Methods were validated by systematic comparisons to clinically documented NASH patients and alignment of clinical characteristics and comorbidities
- Demographic, geographic and other descriptive statistics were generated using SQL, with categorical data expressed as n(%) and continuous results expressed as mean, standard deviation, median, interquartile range as appropriate. As an exploratory characterization of the cohort, no statistical testing was planned or implemented.
- The OM1 NASH Cohort is continuously updated with clinical and administrative claims data. The most updated data available (from January 2013 through April 2019, updated since abstract submission) were included in these descriptive analyses in order to maximize the sample size and reflect up-to-date patient distributions

Results

- A NASH cohort was identified applying a high confidence score threshold. A total of 971,903 patients were identified, including over 141,424 identifiable via diagnosis codes
- Geographic distribution of prevalent NASH patients identified by the algorithm and/or diagnosis codes is shown in **Figure 1**
- Mean age of identified patients was 56 years [standard deviation (SD) 14.4] (for categorical distribution, see **Figure 2**), with 11,584 patients under the age of 18 identified
- In the full cohort, the distribution of males and females was 51% and 49%, respectively (**Figure 3**), which was more evenly distributed compared to patients identified by diagnosis codes and/or documented biopsy alone
- The proportion of patients with documented with type 2 diabetes (T2DM) ranged from 31% to 53%, based on case identification method
- Overall, 2% of patients in the overall NASH cohort had a history of liver transplant
- Of the over 237,000 calculated FIB-4 scores in 77,862 patients, 24% of the results were indicative of advanced fibrosis, the majority of which were in patients not identified by relevant diagnosis codes
- The median FIB-4 score (interquartile range) was 1.40 (0.87, 2.60)

Conclusions

- Identifying probable cases of NASH, and understanding their early clinical characteristics, has important implications for clinical detection and optimizing patient management, as well as facilitating therapeutic clinical development
- This AI-based approach is a robust and broadly applicable means of identifying probable NASH patients that employs data collected during routine clinical care by the wide variety of healthcare providers managing these patients
- Understanding the pathway to diagnosis, which may be different based on socio-demographics, prevalent comorbidities and provider awareness will be key to improving disease detection in earlier, more manageable stages of disease.
- With identification of a larger potential NASH population, further research can target subphenotypes of NASH, such as “lean” non-alcoholic fatty liver disease, patients most likely to progress, pediatric and adolescent NASH, and those at higher risk of non-hepatic outcomes (e.g., cardiovascular sequelae, cancer)

Figure 1. Geographic distribution (heat map)

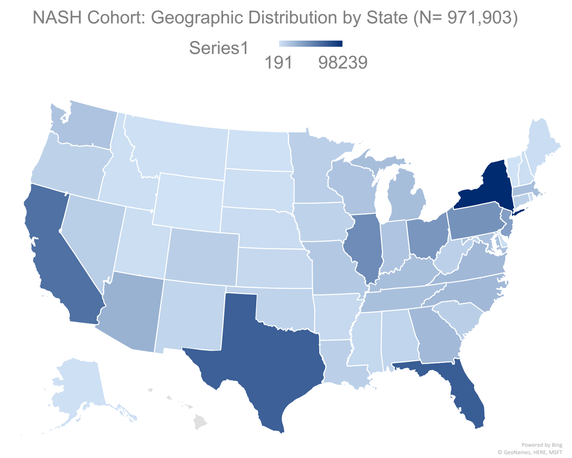


Figure 2. Distribution by age

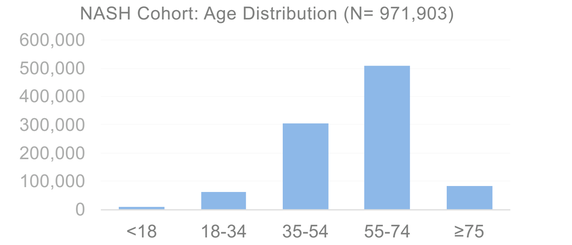
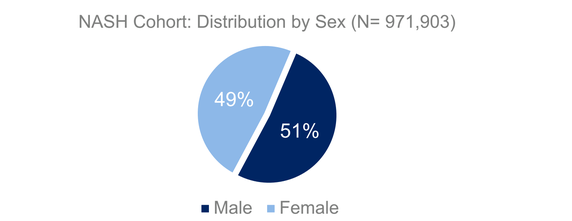


Figure 3. Distribution by sex



Conference

Presented at 2019 DDW Annual Conference. May 18-21, 2019. San Diego, CA, USA

Disclosure

2019 DDW disclosure: The authors have no conflicts to disclose.

References

1. Farrell GC and Larter CZ. Nonalcoholic fatty liver disease: from steatosis to cirrhosis. Hepatology. 2006 Feb;43(2 Suppl 1):S99-S112.
2. Rinella ME, Loomba R, Caldwell SH, et al. Controversies in the Diagnosis and Management of NAFLD and NASH. Gastroenterol Hepatol (N Y). 2014;10(4):219–227.