Machine Learning Tool Can Predict Risk of Fatty Liver Disease

The tool relies on 14 clinical variables, such as liver enzymes, body mass index, triglycerides, height and sex.

November 8, 2019 By Benjamin Ryan

Researchers developed a machine-learning tool that can predict the risk of non-alcoholic steatohepatitis (NASH) among people with other health conditions.

Jörn Schattenberg, MD, of the department of medicine at the University Medical Center in Mainz, Germany, presented findings from the study at the Annual Meeting of the American Association for the Study of Liver Diseases in Boston (The Liver Meeting). The research team also included investigators from Novartis Pharma AG in Basel, Switzerland, and ZS Associates in New Jersey.

NASH and its less severe form, non-alcoholic liver disease (NAFLD), occur when fat accumulates in the liver. Over time, this can lead to fibrosis (scarring), cirrhosis and liver cancer.

Because NASH is underdiagnosed, new methods for identifying the liver disease are greatly needed.

The researchers developed machine-learning algorithms that could predict the risk of having NASH based on noninvasive, routinely collected clinical parameters. They tested their algorithms on the NAFLD Adult Database from the National Institute of Diabetes, Digestive and Kidney Diseases (NIDDK), which includes data on about 450 people confirmed to have NASH and about 250 people who have NAFLD.

Next, the study authors took the algorithm that performed the best and tested it on the Optum Humedica electronic medical records database, which included 3 million people about whom there were sufficient data to validate the model. This cohort includes 23,000 people with NASH diagnosed over a 10-year period. The 1,016 members of this group who had NASH confirmed via a liver biopsy were used to evaluate the performance of the model.

As a performance measure, the researchers looked at the area under the curve (AUC), which is a measure of how well a diagnostic tool can identify the existence of a disease, with 1.0 being a perfect score. They also looked at how accurate the models were at correctly identifying the presence of NASH, the absence of NASH and the overall accuracy of the tools.

The best-performing model had an AUC of 0.82 with the NIDDK data set. This model drew upon 14 variables, which, in order of most to least importance, included hemoglobin A1c (HbA1c, a cumulative measure of blood sugar), AST liver enzymes, ALT liver enzymes, total protein, AST-to-ALT liver enzyme ratio, body mass index, triglycerides, height, platelet count, white blood cell count, hematocrit (the proportion of red blood cells), albumin, high blood pressure and sex.

The model had an AUC of 0.76 when tested on the Optum records.

A simplified model used just five variables, including HbA1c, AST, ALT, triglycerides and total protein. It had a slightly lower level of accuracy, with an AUC of 0.80 for the NIDDK data and 0.74 for the Optum data.

The researchers estimated that by using the model on patients in the Optum database, they could predict up to 29,000 additional previously unidentified people with NASH per 100,000 people with the disease.

The study authors concluded that the model may be used with existing electronic health records data as an effective and scalable means of prescreening for NASH and referring those at risk of the condition to specialists.

More research is planned to validate the model in actual clinical practices to determine its value in such a setting and as a tool to help recruit participants for clinical trials.

To read a press release about the study, click here.
