AP&T Alimentary Pharmacology and Therapeutics

Nonalcoholic fatty liver disease (NAFLD) in the Veterans Administration population: development and validation of an algorithm for NAFLD using automated data

N. Husain*,†, P. Blais*, J. Kramer‡,§, M. Kowalkowski‡,§, P. Richardson‡,§, H. B. El-Serag*,†,‡ & F. Kanwal*,†,‡

*Department of Medicine, Baylor College of Medicine, Houston, TX, USA.
†Section of Gastroenterology and Hepatology, Baylor College of Medicine and Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA.
‡Center for Innovations in Quality, Effectiveness and Safety (IQuESt), Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA.
§Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX, USA.

Correspondence to:
Dr F. Kanwal, 2002 Holcombe Blvd. (152), Houston, TX 77030, USA.
E-mail: fasiha.kanwal@va.gov

Publication data
Submitted 9 April 2014
First decision 4 June 2014
Resubmitted 23 July 2014
Accepted 27 July 2014
EV Pub Online 26 August 2014

This article was accepted for publication after full peer-review.

SUMMARY

Background
In practice, nonalcoholic fatty liver disease (NAFLD) is diagnosed based on elevated liver enzymes and confirmatory liver biopsy or abdominal imaging. Neither method is feasible in identifying individuals with NAFLD in a large-scale healthcare system.

Aim
To develop and validate an algorithm to identify patients with NAFLD using automated data.

Methods
Using the Veterans Administration Corporate Data Warehouse, we identified patients who had persistent ALT elevation (≥2 values ≥40 IU/mL ≥6 months apart) and did not have evidence of hepatitis B, hepatitis C or excessive alcohol use. We conducted a structured chart review of 450 patients classified as NAFLD and 150 patients who were classified as non-NAFLD by the database algorithm, and subsequently refined the database algorithm.

Results
The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) for the initial database definition of NAFLD were 78.4% (95% CI: 70.0–86.8%), 74.5% (95% CI: 68.1–80.9%), 64.1% (95% CI: 56.4–71.7%) and 85.6% (95% CI: 79.4–91.8%), respectively. Reclassifying patients as having NAFLD if they had two elevated ALTs that were at least 6 months apart but within 2 years of each other, increased the specificity and PPV of the algorithm to 92.4% (95% CI: 88.8–96.0%) and 80.8% (95% CI: 72.5–89.0%), respectively. However, the sensitivity and NPV decreased to 55.0% (95% CI: 46.1–63.9%) and 78.0% (95% CI: 72.1–83.8%), respectively.

Conclusions
Predictive algorithms using automated data can be used to identify patients with NAFLD, determine prevalence of NAFLD at the system-wide level, and may help select a target population for future clinical studies in veterans with NAFLD.

Aliment Pharmacol Ther 2014; **40**: 949–954

doi:10.1111/apt.12923

BACKGROUND

Nonalcoholic fatty liver disease (NAFLD) encompasses a spectrum of disease ranging from simple hepatic steatosis to inflammatory steatohepatitis with fibrosis and ultimately cirrhosis and possibly hepatocellular carcinoma.[1] NAFLD typically occurs in patients with features of metabolic syndrome including visceral adiposity and insulin resistance.[2] Coincident with large secular increases in both obesity and diabetes, NAFLD is now the leading cause of chronic liver disease in the US with general population prevalence estimated at 30%. Indeed, recent data show that NAFLD may be the fastest growing cause of cirrhosis in the US.[3]

Determining NAFLD prevalence requires a valid and reliable case definition. In clinical practice, NAFLD is diagnosed based on biochemical tests – such as elevated liver enzymes in the absence of other liver diseases (e.g. viral hepatitis, excessive alcohol use) – and increased echogenicity on abdominal imaging.[4] While liver biopsy is currently the gold standard to definitively diagnose NAFLD[2] and specifically nonalcoholic steatohepatitis (NASH), it has poor patient acceptance and potential for complications. Liver imaging (particularly ultrasound) may be feasible but is also inefficient and cost-prohibitive when applied to large numbers of patients. Neither liver biopsy nor imaging is a feasible option in large-scale representative determination of NAFLD prevalence because they need to be applied to several thousand individuals. Previous epidemiological studies have relied on algorithms based on clinical and laboratory parameters to define NAFLD; most relied on raised alanine aminotransferase (ALT) levels in the absence of high alcohol consumption, hepatitis B virus (HBV) or hepatitis C virus (HCV) infection. However, the accuracy of this algorithm in identifying individuals with NAFLD remains unclear.[5–7]

The Veterans Administration (VA) is the largest healthcare system in the US providing care to over 5 million Veterans nationwide. Given the higher prevalence of diabetes, dyslipidaemia and obesity in the VA population,[8, 9] it is likely that the burden of NAFLD is substantial. However, despite the increasing recognition of NAFLD as a major public health burden, there are no data regarding its prevalence and associated complications in the VA population.

In this study, we sought to adapt and validate an algorithm derived from VA automated administrative and clinical data to identify patients with NAFLD. Because our goal was to identify a population that had the highest likelihood of having NAFLD, *a priori*, we opted to select a NAFLD definition with the highest specificity and positive predictive value (PPV). This algorithm could be used to determine NAFLD prevalence in the national VA system and to identify a target population for treatment and secondary prevention including future clinical studies in Veterans with NAFLD.

METHODS

Data sources used to develop the algorithm

Data elements for developing the NAFLD algorithm were obtained from the VA Corporate Data Warehouse data available on the VA Informatics and Computing Infrastructure (VINCI) platform. CDW is a relational database that contains multiple data elements extracted from the VA electronic medical record such as out-patient and in-patient utilisation with associated diagnostic and procedure codes, laboratory results and pharmacy data.[10] CDW also houses data from Alcohol Use Disorders Identification Test (AUDIT-C), which is a brief alcohol screening instrument that reliably identifies patients who are hazardous drinkers or have active alcohol use disorders.[11, 12] Over 90% of VA ambulatory patients nationwide are screened for alcohol use disorders with AUDIT-C.[13]

NAFLD algorithm development

We used laboratory data from the CDW to identify patients with persistently elevated ALT levels, defined as two or more values ≥40 IU/mL at least 6 months apart during 2001–2011. We used the cut-off of 40 IU/mL as the upper range for ALT because the majority of clinical and epidemiological studies have used this value to define elevated ALT. We excluded patients with any positive tests for hepatitis B surface antigen (HbsAg), hepatitis C antibody (HCV Ab) or hepatitis C RNA (HCV RNA) from this cohort. We also excluded individuals with alcohol use disorder (ICD-9 codes: 291.x, 303.0x, 303.9x and 305.0x) or an AUDIT-C score >4 for males (>3 for females)[14] in the 12 months prior to or after the first elevated ALT.

We then randomly selected 150 patients classified as having NAFLD based on the criteria above who received their care at Michael E DeBakey VA Medical Center (MEDVAMC) in Houston, Texas. In this sample, we conducted an implicit qualitative review of patients' medical records in the Computerized Patient Record System (CPRS). CPRS is the electronic medical record in

the VA and contains a problem list, pharmacy data, orders, laboratory test results, progress notes, radiology results and pathology results. The purpose for this review was to gain insight into what additional information may be needed from the database to refine the algorithm and improve its PPV.

Algorithm refinement and validation

For algorithm refinement and validation, we randomly selected a new sample consisting of 600 patients who received their care at MEDVAMC. This included 450 patients who met the database-derived criteria for NAFLD definition above and 150 patients who did not meet these criteria and thus were classified as non-NAFLD based on the database. The purpose of this review was to verify NAFLD diagnosis and calculate the operating characteristics of the algorithm. A trained clinician, who was blinded to patients' NAFLD status, conducted an explicit structured chart review using the CPRS.

Our gold standard definition for verifying NAFLD derived from chart review was elevated liver enzymes (i.e. ALT >40 IU/mL for males and ALT >31 IU/mL for females[7, 8]) and no other documented liver diseases (HCV, HBV, alcoholic liver disease and other less common etiologies of chronic liver disease) in the presence of either: (i) a histological diagnosis of NAFLD, (ii) radiological NAFLD features such as fatty liver, increased or heterogeneous echogenicity, or (iii) patients not meeting criteria #1 or #2 but had features of metabolic syndrome defined as two or more of the following: BMI >25, hypertension defined by systolic blood pressure ≥130 mmHg or diastolic blood pressure ≥85 mmHg; high density lipoprotein (HDL) <40 g/dL; triglycerides >150 mg/dL or presence of diabetes (haemoglobin A1c >7.0). For each measurement, if there were multiple values we used the value that was closest to the date of the first elevated ALT. If patients were on medications for hypertension, hyperlipidaemia or diabetes, then we considered them as having met the respective criteria.

Statistical analyses

We present the calculations of the specificity, sensitivity, PPV and negative predictive value (NPV) of the various versions of the algorithm for identifying NAFLD correctly as defined in the chart reviews in the 600 patients in the final cohort. The specificity of the algorithm was calculated by dividing the number of cases that were negative for NAFLD by the algorithm and the chart review (true negatives) by the total number of records negative by *chart review*. The sensitivity of the algorithm was calculated by dividing the number of cases that were positive for NAFLD by the algorithm and confirmed on chart review (true positives) by the total number of records positive by *chart review*. The PPV of the algorithm was calculated by dividing the number of cases for which NAFLD was confirmed on chart review (true positives) by the total number of records positive for NAFLD by the *algorithm* from the database. The NPV was calculated by dividing the number of cases without NAFLD based on the algorithm and confirmed negative by chart review (true negatives) by the total number of records negative for NAFLD by the *algorithm*. The specificity, sensitivity, PPV and NPV were also adjusted for oversampling of the NAFLD group from the database by applying inverse probability sampling weights.

Sensitivity analysis

AUDIT-C screening was not used consistently before 2004; therefore a few patients with alcohol use may have been misclassified as having NAFLD particularly if they only used the VA prior to 2004. Therefore, in a sensitivity analysis, we only included patients who had any documented elevation in ALT tests after 1 January 2004 (after AUDIT-C was widely implemented in the VA as an alcohol use disorder screening tool).

RESULTS

Demographics of validation sample

The mean age of our sample of 600 patients was 53.5 years (standard deviation, s.d. 14.3), almost all were male, 59.0% were White, 24.5% were African Americans, 6.7% were Hispanics and 9.8% belonged to other racial groups. Table 1 details the key demographical and clinical characteristics of our study sample. Patients who met the database-derived criteria for NAFLD were older, more likely to be white, obese (defined on the basis of body mass index ≥30 kg/m^2), and had a higher prevalence of diabetes and hypercholesterolaemia than patients who did not meet NAFLD criteria.

Performance of original NAFLD algorithm

Of the 450 patients selected by the original NAFLD algorithm (which included patients with two elevated ALT values >6 months apart, with no positive serologic testing for HBV and HCV, and no positive AUDIT-C scores within 1 year of the elevated ALT), 358 patients had evidence of NAFLD based on the chart review. Of the 150 patients selected without NAFLD by the original algorithm, 122 did not have NAFLD in the chart.

Table 1 | Demographical and clinical characteristics of 600 patients included in the validation sample

Characteristics (%)	Met NAFLD definition (n = 450)	Did not meet NAFLD definition (n = 150)
Demographics		
Age, median (IQR)	55.6 (46.4–61.7)	54.2 (42.8–59.8)
Race		
White	61.8	48.8
African Americans	22.4	32.0
Hispanics	7.2	6.6
Other race	8.6	12.6
Clinical characteristics (%)		
Diabetes	54.7	32.0
Obesity (BMI >30 kg/m^2)	54.1	36.8
Hypercholesterolaemia	34.8	20.0
Hypertension	10.4	11.5

We defined diabetes, hypercholesterolaemia and hypertension based on presence of ICD-9 codes in the administrative data. We used the AHRQ Clinical Classification System to classify all patients ICD9 codes into the relevant diagnosis.

Our original algorithm yielded a weighted specificity, sensitivity, PPV and NPV of 74.5%, 78.4%, 64.1% and 85.6%, respectively (definition #1 in Table 2).

NAFLD algorithm refinement

During our initial implicit qualitative chart review, we observed several reasons for false positive results per the database-derived NAFLD algorithm. Specifically, patients were inappropriately captured by the algorithm if they had malignancy with metastases to the liver or if they had elevated liver enzymes during hospital admissions (for reasons including choledocholithiasis, cholecystitis, congestive heart failure exacerbation, etc.) that returned to normal after discharge. In addition, some patients met the database algorithm but had only a few instances of elevated ALT that were several years apart (e.g. two elevated ALT more than 3–5 years apart) with majority of ALTs being in the normal range.

Final algorithm refinement and testing

We made step-wise modifications to the algorithm based on our observations from the implicit review above.

We first excluded patients with malignancy with metastases. This resulted in specificity, sensitivity, PPV and NPV of 79.4%, 71.2%, 66.7% and 82.6%, respectively (definition #2 in Table 2).

Next, we excluded ALT values that were measured within 3 months before or after an inpatient hospitalisation. This resulted in specificity, sensitivity, PPV and NPV of 83.1%, 67.8%, 70.0% and 81.7%, respectively (definition #3 in Table 2).

Last, we re-classified patients as NAFLD in the database if they had two elevated ALTs in the ambulatory setting, greater than 6 months apart but within 2 years each other. Other criteria were same as above and included patients with no positive serologic testing for HBV and HCV, and no positive AUDIT-C scores within 1 year of the elevated ALT. This analysis yielded specificity, sensitivity, PPV and NPV of 92.4%, 55.0%, 80.8% and 78.0%, respectively (definition #4 in Table 2).

Limiting to only those patients who had any documented elevation in ALT tests after 1 January 2004 did not change the specificity, sensitivity, PPV, or NPV of the algorithm appreciably [90.8% (95% CI: 86.5–95.2%); 60.8% (95% CI: 51.4–70.2%); 80.8% (95% CI: 72.5–89.0%) and 78.5% (95% CI: 72.0–85.0%) respectively].

Table 2 | Specificity, sensitivity, positive predictive value (PPV) and negative predictive value (NPV) with corresponding 95% confidence intervals (CIs) of different algorithms for defining NAFLD based on automated data and compared to a gold standard based on chart reviews

Definitions	Specificity (95% CI)	Sensitivity (95% CI)	PPV (95% CI)	NPV (95% CI)
#1: Two elevated ALT ≥ 6 months apart, no positive tests for HBV or HCV, no alcohol use within 1 year of first elevated ALT	74.5 (68.1–80.9)	78.4 (70.0–86.8)	64.1 (56.4–71.7)	85.6 (79.4–91.8)
#2: #1 plus patients with metastatic cancer removed	79.4 (73.6–85.1)	71.2 (62.2–80.3)	66.7 (58.9–74.4)	82.6 (76.3–89.0)
#3: #2 plus 2 elevated ALT ≥6 months apart in ambulatory setting	83.1 (77.6–88.6)	67.8 (58.7–76.9)	70.0 (61.7–78.3)	81.6 (75.4–87.8)
#4: #3 *but* second elevated ALT within 2 years of first elevated ALT	92.4 (88.8–96.0)	55.0 (46.1–63.9)	80.8 (72.5–89.0)	78.0 (72.1–83.8)

Thus, our final algorithm classified patients as having NAFLD if they had at had ≥2 elevated ALT values performed in an ambulatory setting, greater than 6 months apart but within 2 years of each other, with no evidence of metastatic cancer, positive serologic testing for HBV and HCV, and no positive AUDIT-C scores within 1 year of the elevated ALT (definition #4 in Table 2).

DISCUSSION

This is the first study to systematically examine the validity of a predictive algorithm for identifying patients with NAFLD using automated data. We found that the algorithm that relies on two or more ambulatory ALT values that were elevated for at least 6 months (but within 2 years of each other) in the absence of evidence for chronic HCV, HBV or alcohol use disorders is associated with a specificity of 92.4% (95% CI: 88.8–96.0%) and a PPV of 80.8% (95% CI: 72.5–89.0%) for the presence of NAFLD. This means that 92.4% of the patients without NAFLD in the charts did not have NAFLD based on the algorithm and 80.8% of the patients who were classified as NAFLD by the algorithm truly had NAFLD as determined by our chart review.

We believe that the results of this study are useful in several ways. Our data can guide researchers, clinicians and policy makers interested in using pre-existing automated databases to examine questions regarding patients with NAFLD. We recognise that the NAFLD algorithm, like other tests, has different operating characteristics (i.e. PPV and NPV, sensitivity and specificity) under different specifications. Therefore, we calculated and presented these data for different NAFLD definitions (Table 1). We believe that the exact approach used to identify patients with NAFLD will be guided by the underlying goals of the programme and available resources. Some may opt to focus on the group most likely to have NAFLD (as we did), whereas others may prefer to cast a broader net. In the former scenario, our selected algorithm may identify patients who can serve as a high yield target for the studies, such as those evaluating effectiveness of existing as well as new treatment options for NAFLD. In contrast, using a less restrictive definition (such as definition #1 in Table 2) with high sensitivity and NPV may be well suited for epidemiological studies geared towards understanding the burden of disease as well as for projects focusing on preventive care services for those at risk for NAFLD.

In addition to being the first study examining the validity of a NAFLD algorithm, we believe that our structured approach relying on both implicit and explicit chart reviews is an important strength. For example, we noted that patients with malignancy with metastasis to the liver often had mild elevations in their ALT values. In addition, we found that ALT and AST were often elevated immediately before, during and after hospital admissions but that these usually normalised after discharge. Not accounting for these instances would have resulted in misclassification of patients as NAFLD. Indeed, the specificity increased from 74.0% to 92.4% as we sequentially refined our algorithm based on lessons learnt as part of the built-in implicit review.

This study has several limitations. First, although, we used the best available objective data in medical records combined with clinical expertise (in cases of ambiguity) to determine NAFLD status in the charts (gold standard), some misclassification may have occurred. Second, NAFLD can exist in the absence of elevated ALT.[15] Although the clinical importance of this presentation is unclear, it likely represents the benign end of NAFLD spectrum. Our approach captured clinically important and apparent NAFLD. We used a cut-off of 40 IU/mL to define elevated ALT because it has been commonly used by previous clinical and epidemiological studies. However, the absolute cut-off to define high ALT in obese individuals and those with diabetes remains unclear. Furthermore, the absolute value of ALT may vary based on the platform used for analysis. However, the alternative approach of using individual laboratories' upper limit of normal (ULN) to define elevated ALT would have introduced a greater error given the wide variability in the ULN reported by clinical laboratories.[16] Lastly, our findings may have limited generalisability as they are derived from a single VA center. Conducting a wider review with patients from different geographical locations would be the ideal way to determine the algorithms' validity; however, this was not practical due to time and feasibility constraints.

In conclusion, predictive algorithms using automated data can be used to identify patients with NAFLD. These algorithms may help determine the prevalence of NAFLD in large datasets and to define a target population for prevention and treatment including future clinical studies in Veterans with NAFLD. With the wide implementation of electronic medical records in a variety of clinical settings, this algorithm may also serve as a potential screener to identify patients with NAFLD in healthcare systems outside the VA.

AUTHORSHIP

Guarantor of the article: Fasiha Kanwal, MD, MSHS.

Author contributions: Nisreen Husain, Peter Blais and Fasiha Kanwal collected data; Nisreen Husain and Fasiha Kanwal prepared the manuscript; Marc Kowalkowski and Peter Richardson analysed the data, Jennifer Kramer and Fasiha Kanwal designed the study; Jennifer Kramer and Hashem B El-Serag interpreted the data and performed critical review of the manuscript. All authors approved the final version of the manuscript.

ACKNOWLEDGEMENTS

Declaration of personal interests: None.

Declaration of funding interests: This study is funded in part by the Center for Innovations in Quality, Effectiveness and Safety (#CIN 13-413). Dr El-Serag is also supported by National Institute of Diabetes and Digestive and Kidney Diseases K24-04-107.

DISCLAIMER

The opinions and assertions contained herein are the sole views of the authors and are not to be construed as official or as reflecting the views of the Department of Veterans Affairs.

REFERENCES

1. White DL, Kanwal F, El-Serag HB. Association between nonalcoholic fatty liver disease and risk for hepatocellular cancer, based on systematic review. *Clin Gastroenterol Hepatol* 2012; **10**: 1242–59.
2. Chalasani N, Younossi Z, Lavine JE, *et al.* The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Gastroenterological Association, American Association for the Study of Liver Diseases, and American College of Gastroenterology. *Gastroenterology* 2012; **142**: 1592–609.
3. Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. *Aliment Pharmacol Ther* 2011; **34**: 274–85.
4. Jepsen P, Gronbaek H. Prognosis and staging of non-alcoholic fatty liver disease. *BMJ* 2011; **343**: d7302.
5. Ruhl CE, Everhart JE. Determinants of the association of overweight with elevated serum alanine aminotransferase activity in the United States. *Gastroenterology* 2003; **124**: 71–9.
6. Ioannou GN, Boyko EJ, Lee SP. The prevalence and predictors of elevated serum aminotransferase activity in the United States in 1999–2002. *Am J Gastroenterol* 2006; **101**: 76–82.
7. Patt CH, Yoo HY, Dibadj K, Flynn J, Thuluvath PJ. Prevalence of transaminase abnormalities in asymptomatic, healthy subjects participating in an executive health-screening program. *Dig Dis Sci* 2003; **48**: 797–801.
8. Das SR, Kinsinger LS, Yancy WS Jr, *et al.* Obesity prevalence among veterans at Veterans Affairs medical facilities. *Am J Prev Med* 2005; **28**: 291–4.
9. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data *Diabetes Care* 2004; **27**(Suppl. 2): B10–21.
10. Health Services Research and Development. VA Informatics and Computing Infrastructure. Available at: http://www.hsrd.research.va.gov/for_researchers/vinci/(accessed April 26, 2012).
11. Bush K, Kivlahan DR, McDonell MB, *et al.* The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Arch Intern Med* 1998; **158**: 1789–95.
12. Bradley KA, Williams EC, Achtmeyer CE, *et al.* Implementation of evidence-based alcohol screening in the Veterans Health Administration. *Am J Manag Care* 2006; **12**: 597–606.
13. Lapham GT, Achtmeyer CE, Williams EC, Hawkins EJ, Kivlahan DR, Bradley KA. Increased documented brief alcohol interventions with a performance measure and electronic decision support. *Med Care* 2012; **50**: 179–87.
14. Using AUDIT-C Alcohol Screening Data in VA Research: Interpretation, Strengths, Limitations, & Sources. Available at: http://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/vci-031511.pdf (accessed April 27, 2012).
15. Soderberg C, Stal P, Askling J, *et al.* Decreased survival of subjects with elevated liver function tests during a 28-year follow-up. *Hepatology* 2010; **51**: 595–602.
16. Neuschwander-Tetri BA, Unalp A, Creer MH; Nonalcoholic Steatohepatitis Clinical Research Network. Influence of local reference populations on upper limits of normal for serum alanine aminotransferase levels. *Arch Intern Med* 2008; **168**: 663–6.