A comparison of Natural Language Processing, ICD10 codes, and manual chart abstraction for identifying patients with NASH in an EHR system.

Pankil Shah, Alex Bokov, Dimpy Shah, …

We present the comparative performance analysis of the NLP based CliniThink classifier in identifying patients with NASH (Non-Alcoholic SteatoHepatitis) as compared to manual chart abstraction and ICD10 based query. These methods represent three different approaches to patient cohort identification from electronic medical records. We identified the study population from a convenience sample of bariatric clinic patients. Study protocol and methods can be found here. In identifying NASH cases, the CliniThink classifier performed comparably to the manual chart review. CliniThink had similar sensitivity levels and marginally lower specificity. Yoden’s Index (expressed here on a 0 to 100 scale) was also similar; however, chart review had overall higher accuracy. Upon inspection of divergent cases, we observed that CliniThink NLP excelled at phrases matching, but fell short because its query did not account for corroborative clinical criteria (e.g. fib4 score, MetSyn, elevated LFTs). The CliniThink query took only a few minutes during the demo and had substatial time saving benefits over chart review. Both CliniThink and manual chart review performed better than ICD10 based query. This finding could be attributed to a lack of specific coding for NASH in ICD10 and inadequate coding practice as NASH was not always the principal diagnosis. None of the differences were statistically significant. The analysis was expanded to include expanded diagnostic criteria for NASH (clinical NASH) as well as probable classification. We observed a similar trend in the results.

# Introduction

Nonalcoholic fatty liver disease (NAFLD) is a condition in which excess fat is stored in the liver but not caused by heavy alcohol use. About 20% of NAFLD cases progress to nonalcoholic steatohepatitis (NASH) characterized by inflammation of the liver and liver cell damage, in addition to fat deposits in the liver. This in turn can cause fibrosis, or scarring, of the liver. NASH is definitively diagnosed by biopsy or fibroscan. In the absence of such results NASH is suspected if the clinical diagnosis of NAFLD is accompanied by elevated liver enzymes, CT/USG imaging reports mentioning fibrosis, and cluster of related conditions such as diabetes (DM), hypertension, MetSyn.

* NASH is a good case study
  + No ICD9
  + Biopsy confirmatory – cost, peads
  + Treatment
  + Need for Clinical Trials
* Search Strategy: for search in case of say DM vs NASH
* Some background on each of the 4 approaches
  + Time
  + Effort
  + Funding / pilot
* Discussion:
  + Address Bias

# Materials and Methods

## Patient Cohort

A sample of 286 patients seen at the MARC Bariatric Surgery Clinic from January 1 2019 through [???] was randomly selected.

## Chart Review

We used a combination of a resident, medical student/s and graduate student/s to conduct chart review of sequential cases at the Bariatric Surgery Clinic to identify NASH / hepatic fibrosis cases. For each patient, each of the reviewers read the [X, Y, and Z] panels in the Epic Hyperspace EHR system. IRR was required to be greater than 0.95 to maintain consistent standard of evaluation. Reviewers that did not meet this threshold were trained or replaced.

### Expert Chart Review (Gold Standard)

Any patient who was screened as a case of NASH / hepatic fibrosis by at least one of the screening methods was included in the expert review. Additionally, a random sequential sample of 10% of all cases were included in the expert review to address verification bias. The diagnosis status of the patients not identified as cases by any of the screening methods and also not included in the 10% random sequential expert review were imputed based on lab results, billing and treatment variables for FibroTest, hepatitis B/C, alcoholic liver disease, cirrhosis, etc. using random forest or other statistical methods.

### 

### Clinithink NLP

[describe Clinithink method]

### ICD10 Codes

Patients with any ICD10 code in K75.81, XX, or YY were judged positive for NASH or hepatic fibrosis.

# Results

For each of the three screening methods only the definitive verdicts are shown. For the ICD10 screen there were no probable verdicts.

## Chart Review

|  |  |  |  |
| --- | --- | --- | --- |
|  | NASH + | NASH - | Total |
| **Chart Review +** | 8 | 11 | 19 |
| **Chart Review -** | 3 | 264 | 267 |
| **Total** | 11 | 275 | 286 |

|  |  |  |
| --- | --- | --- |
|  | Estimate | CI |
| Apparent prevalence | 0.07 | (0.04, 0.10) |
| True prevalence | 0.04 | (0.02, 0.07) |
| Sensitivity | 0.73 | (0.39, 0.94) |
| Specificity | 0.96 | (0.93, 0.98) |
| Positive predictive value | 0.42 | (0.20, 0.67) |
| Negative predictive value | 0.99 | (0.97, 1.00) |
| Positive likelihood ratio | 18.18 | (9.19, 35.99) |
| Negative likelihood ratio | 0.28 | (0.11, 0.75) |

### Inter-rater reliability

### Personnel-hours

## Diagnosis Codes

|  |  |  |  |
| --- | --- | --- | --- |
|  | NASH + | NASH - | Total |
| **Diagnosis Codes +** | 6 | 19 | 25 |
| **Diagnosis Codes -** | 5 | 256 | 261 |
| **Total** | 11 | 275 | 286 |

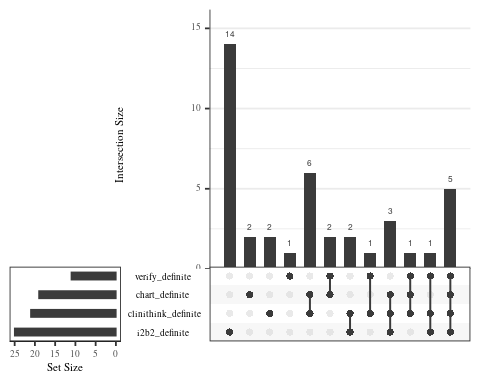
|  |  |  |
| --- | --- | --- |
|  | Estimate | CI |
| Apparent prevalence | 0.09 | (0.06, 0.13) |
| True prevalence | 0.04 | (0.02, 0.07) |
| Sensitivity | 0.55 | (0.23, 0.83) |
| Specificity | 0.93 | (0.89, 0.96) |
| Positive predictive value | 0.24 | (0.09, 0.45) |
| Negative predictive value | 0.98 | (0.96, 0.99) |
| Positive likelihood ratio | 7.89 | (3.95, 15.78) |
| Negative likelihood ratio | 0.49 | (0.26, 0.93) |

## Clinithink NLP

|  |  |  |  |
| --- | --- | --- | --- |
|  | NASH + | NASH - | Total |
| **Diagnosis Codes +** | 8 | 13 | 21 |
| **Diagnosis Codes -** | 3 | 262 | 265 |
| **Total** | 11 | 275 | 286 |

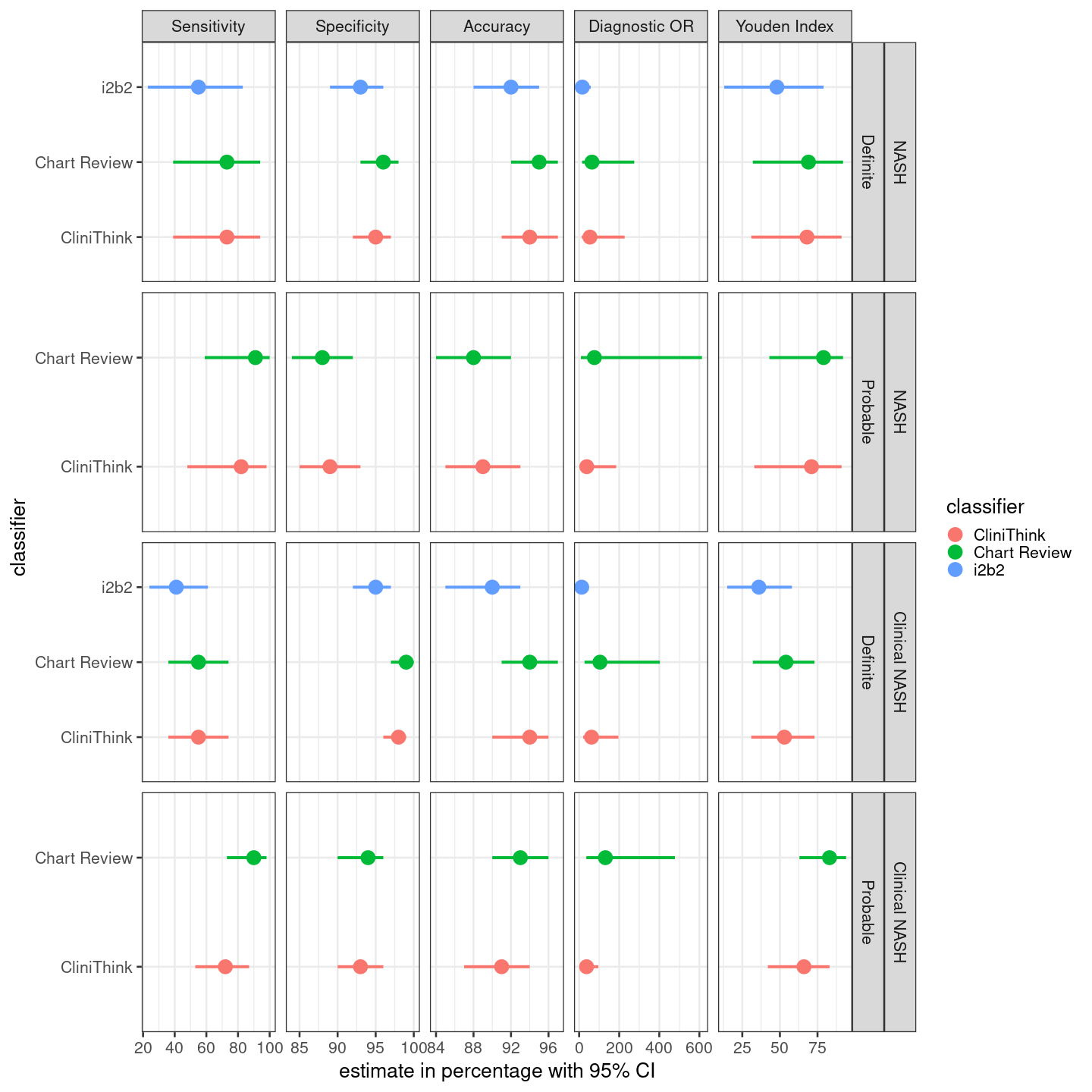
|  |  |  |
| --- | --- | --- |
|  | Estimate | CI |
| Apparent prevalence | 0.07 | (0.05, 0.11) |
| True prevalence | 0.04 | (0.02, 0.07) |
| Sensitivity | 0.73 | (0.39, 0.94) |
| Specificity | 0.95 | (0.92, 0.97) |
| Positive predictive value | 0.38 | (0.18, 0.62) |
| Negative predictive value | 0.99 | (0.97, 1.00) |
| Positive likelihood ratio | 15.38 | (8.09, 29.24) |
| Negative likelihood ratio | 0.29 | (0.11, 0.75) |

## Agreement and disagreement between screening methods



An UpSet diagram showing which patients were identified by which combination of screening methods. [Note that the largest single group currently are 14 patients identified by ICD10 codes only. Note also that there are 5 patients identified by all three screening methods and verified. There are also 3 patients identified by all three methods but not verified.]

## Overall performance



# Discussion

Recently, Van Vleck *et al.* (2019) published results from using this algorithm to identify NALD leaving open the question of […] which we address by […] .

Our results are broadly in agreement with theirs but in addition […]

We need to find a way to distinguish ourselves from Van Vleck et al. 2019, because if this is all we report it’s basically a clone of their paper but with a smaller sample size. Possible points to develop further:

1. Staff effort
2. Inter-rater reliability
3. What makes it easy for abstractors to overlook patients that are specifically coded for NASH (or, why are patients coded for NASH erroneously)
4. Technically, Van Velck et al. focus on NALD and we focus on NASH, so this could be framed as the next logical step building on their work instead of them simply scooping us

# Conclusions

# References

Van Vleck, T. T. *et al.* (2019) ‘Augmented intelligence with natural language processing applied to electronic health records for identifying patients with non-alcoholic fatty liver disease at risk for disease progression’, *International Journal of Medical Informatics*, 129, pp. 334–341. doi: [10.1016/j.ijmedinf.2019.06.028](https://doi.org/10.1016/j.ijmedinf.2019.06.028).