Considerations for developing algorithms

• Moving between clinical definitions and the data available in databases and EMRs is NOT an exercise in mapping codes
• Defining elements include considerations of
  • Specificity of diagnostic terms
  • Overlap of diagnostic terms
  • Timing
  • Clinical plausibility
  • Differential diagnosis
• Iterative algorithm development
  • “Supervised learning”
  • Clinical expertise
  • Database expertise
• Achieves local validity
An algorithm for acute liver disease

1. Source population
2. A priori screen
3. Candidate cases
4. Sample 50
5. Profiles
6. Classify by machine rule
7. Classify by expert
8. Compare
9. Dissimilar
10. Update machine rule
11. Similar
12. Done
Sodium (NA)=140mmol/l
eGFR=89

NLP:
ALT--ALT(ELEVATED)
NAUSEA--NAUSEA()

01NOV2010  Dx:
ABDOMINAL PAIN RIGHT UPPER QUADRANT(78901)

02NOV2010  Lab:
Creatinine=1mg/dl
Bilirubin.total=.4mg/dl
Alanine aminotransferase (ALT)=300u/l
Aspartate aminotransferase (AST)=152u/l

03NOV2010  Dx:
NONSPEC ELEV LEVEL TRANSAMINASE/LDH(7904)
Lab:
O2 saturation.oximetry=95%

NLP:
ALT--ALT(ELEVATED)
ABDOMEN--PAIN(ABDOMENOCCASIONAL)
ABDOMEN--PAIN(ABDOMENSOME;OCCASIONAL;CRAMPY)
ABDOMEN--PAIN(ABDOMENWORSENING)

OBS:
BMI=21.5
DBP=76
HR=72
SBP=122
WT=64.9
Inclusion Criteria

*Inclusion 1.* Presence of an ICD-9 code in the record corresponding to “Acute and subacute necrosis of the liver,” “Other disorders of liver,” or “Jaundice, unspecified, not of newborn”;

*Inclusion 2.* AST or ALT values of at least three times the upper limit of normal (≥3xULN), provided that there had been a preceding value less than two times the upper limit of normal;

*Inclusion 3.* Either AST or ALT ≥2xULN and total bilirubin ≥3xULN or AST/ALT ≥3xULN and bilirubin ≥2xULN. The exclusion criteria come from both structured and NLP data.

Onset

The first date an inclusion criterion was met, or up to a week earlier if there were nonspecific hepatic ICD9 or NLP terms consistent with the onset of acute liver disease.
Expert-derived machine rule (3rd Iteration) 2

Exclusion criteria related to onset

Exclusion 1. Bile duct obstruction codes within seven days of onset.
Exclusion 2. Pancreatitis codes within 30 days of onset.
Exclusion 3. Viral hepatitis codes or NLP within 30 days of onset.
Exclusion 4. Metastatic cancer codes within 30 days of onset.
Exclusion 5. Chemotherapy codes within 30 days of onset.

Exclusion criteria related to underlying conditions

Exclusion 7. Cholangitis codes appearing at least two times in the record. These were considered to represent sclerosing cholangitis.

Exclusion 8. Persistent transaminase elevation. Multiple values of AST/ALT $\geq$3xULN, not contained within a 183-day window and with no AST/ALT values $<2xULN$ intervening. Does not exclude when there is an AST/ALT $\geq$6xULN.

Exclusion 9. Other chronic conditions. Codes for sarcoidosis or other chronic nonalcoholic liver disease, plus NLP terms for cirrhosis.
<table>
<thead>
<tr>
<th>Visit Date or Result Date</th>
<th>Diagnosis(Dx), Prescription(Rx), Lab, Procedure(PROC), Observations(OBS), Hospitalization(Hosp.), ER Visit, Dr. Notes(NLP)</th>
</tr>
</thead>
</table>
| 31 OCT 2010              | **Dx:**  
|                          | NAUSEA ALONE(78702)  
|                          | NONSPEC ELEV LEVEL TRANSMINASE/LDH(7904)  
|                          | OTH NONSPEC ABN SERUM ENZYM LEVLS(7905)  
|                          | **Lab:**  
|                          | Albumin=4.2g/dl  
|                          | Alkaline phosphatase (ALP) (ALKP)=172u/l  
|                          | Alkaline phosphatase (ALP) (ALKP)=186u/l  
|                          | Bilirubin.direct=.1mg/dl  
|                          | Blood urea nitrogen=16mg/dl  
|                          | CO2.total=28mmol/l  
|                          | Calcium.total=9.1mg/dl  
|                          | Chloride=105mmol/l  
|                          | Glucose.random=110mg/dl  
|                          | Hepatitis B surface antigen (HBsAg).qualitative=non-reactive  
|                          | Hepatitis C antibody (HCAb) (anti-HCV).qualitative=negative  
|                          | O2 saturation.oximetry=95%  
|                          | Potassium=3.1mmol/l  

Concordance between machine rule and clinical review (50 new candidates)

<table>
<thead>
<tr>
<th>Machine Rule</th>
<th>Clinical Review</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case</td>
<td>Non-Case</td>
</tr>
<tr>
<td>Case</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>Non-Case</td>
<td>36</td>
<td>1</td>
</tr>
</tbody>
</table>

Sensitivity = \( \frac{11}{12} = 91\% \)
Specificity = \( \frac{35}{36} = 97\% \)
Predictive Value when the rule is positive = \( \frac{11}{14} = 79\% \)
Notes for change after the 50-patient test assessment

Most discrepancies resulted in further proposed rule updates.

1. An NLP symptom flag should not set onset date if the symptom has been repeatedly present in the past.
2. Remove ICD9 573.3 (Hepatitis, unspecified) and 573.8 (Other specified disorders of liver) as qualifying codes for ALD.
3. Shorten requirement for persistent ALT>3xULN five months (≥152 days).
4. Change rule to capture jumps to >6xULN as acute transaminase elevation even with no preceding normal value.

This is a continuous process.
Considerations for developing algorithms

• Moving between clinical definitions and the data available in databases and EMRs is NOT an exercise in mapping codes
• Defining elements include considerations of
  • Specificity of diagnostic terms
  • Overlap of diagnostic terms
  • Timing
  • Clinical plausibility
  • Differential diagnosis
• Iterative algorithm development
  • “Supervised learning”
  • Clinical expertise
  • Database expertise
• Achieves local validity