Validation of using early modelling to predict the performance of a monitoring test – The use of the ELF biomarker in liver disease modelling and the ELUCIDATE trial

Jon Deeks, Alice Sitch, Jac Dinnes, Julie Parkes, Walter Gregory, Jenny Hewison, Doug Altman, William Rosenberg, Peter Selby
The project story ...

- “A multi-centre programme into the evaluation of biomarkers suitable for use in patients with kidney and liver diseases”

- NIHR funded Programme Grant with 3 workstreams:
  - Methodology workstream
  - Laboratory translational workstream
  - ELUCIDATE RCT workstream
The clinical story ...

**Fibrosis**
- Scar tissue in liver (various causes)
- But liver is still able to replace with new tissue

**Cirrhosis**
- Scaring damage becomes permanent
- Liver function affected, portal hypertension, etc

** Decompensated cirrhosis**
- Severe life threatening clinical disease - jaundice, ascites, variceal haemorrhage, or hepatic encephalopathy.

Identification at this point allows preventive interventions.
The ELF test

- 3 serum biomarkers shown to correlate to the level of liver fibrosis assessed by liver biopsy.
  - Hyaluronic acid (HA)
  - Procollagen III amino terminal peptide (PIIINP)
  - Tissue inhibitor of metalloproteinase 1 (TIMP-1)

Calculating the ELF Score
To calculate the ELF score for the ADVIA Centaur systems, first obtain results for the ADVIA Centaur HA, PIIINP, and TIMP-1 assays and then use the following equation/algorithm to calculate the ELF score:

**ADVIA Centaur XP/XPT:**
ELF score = 2.278 + 0.851 \ln(C_{HA}) + 0.751 \ln(C_{PIINP}) + 0.394 \ln(C_{TIMP1})

Interpretation of Results
Interpretation of the ELF score is as follows:

<table>
<thead>
<tr>
<th>Score Range</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 7.7</td>
<td>None to mild</td>
</tr>
<tr>
<td>≥ 7.7 to &lt; 9.8</td>
<td>Moderate</td>
</tr>
<tr>
<td>≥ 9.8</td>
<td>Severe</td>
</tr>
</tbody>
</table>
The ELUCIDATE trial

Patients randomised between monitoring as usual (TAU) with or without ELF tests every 6 months

N=878 patients
ELUCIDATE research questions

- Does monitoring with ELF permit earlier detection of liver cirrhosis in patients with Chronic Liver Disease to allow earlier interventions?
- result in patient benefit through improved survival and reduced liver-related morbidity and mortality?
- improve the cost-effectiveness of the management of end-stage liver disease?
Value of developing an early model

- Can we use a simulation model to
  - predict the likely event rates in the trial? (progression to cirrhosis)
  - evaluate the impact of choice of ELF threshold and frequency of monitoring?
  - predict the proportion of ELF+ve events who are likely to be false positives?
Modelling of monitoring strategies

Obtain estimates regarding disease progression or recurrence and test performance (measurement error, accuracy, variability)

Simulate patient cohort modelling disease progression and results of the monitoring test based on evidence

Evaluate the performance of alternative monitoring strategies (different thresholds, test frequencies, decision rules)

Identify optimum strategies for further evaluation
Requires data on the mean and variance of ELF values at each fibrosis stage

Obtained from cross-sectional cohort studies of patients undergoing liver biopsy with ELF measurements
Requires data on the rate and variability of progression of fibrosis over time

Obtained from an IPD review of disease progression rates

“Random slope random intercept model”
Variability in disease stage at the start

Requires data on the prevalence of fibrosis grade in those attending clinic

Obtained from cross-sectional study of liver biopsy at clinic
Incorporating measurement error

Requires data on the measurement error in ELF measurements

Difficult to obtain!

Lack of biological variability study data

Quality control values from company, longitudinal cohort data
Identifying disease end points - cirrhosis

Incorporating the error allows us to generate the ELF values that would be observed.

The red diamonds indicate the point at which an individual reaches cirrhosis.
Implementing a monitoring strategy

Different monitoring rules can be applied to the cohort to evaluate impact of monitoring

Rule applied
- 6 monthly observations
- Threshold of 10
Implementing a monitoring strategy

Observation is curtailed at 1st test positive

At event (red diamond)

Up to three months before event true positive (circle)

Never has event, or event over three months later - false positive (cross)
Alternative monitoring strategies

- Using different thresholds
- Using different monitoring frequency
- Sophisticated thresholds
  - absolute or relative increases
  - rates of change
- Retesting positives
Statistics used to evaluate performance

- Sensitivity (proportion with cirrhosis detected at a time point)
- Mean diagnostic delay (evaluated across timepoints)
- Positive predictive value (false positives)
- Mean number of tests per person
Refining the model in real time

- Data from the ongoing trial allows refinement of key parameters in the model
  - Screening and randomisation ELF values provided a better estimate of measurement error
  - We could also have used baseline ELF values from the trial
Validation - comparing trial and model

Trial data

Simulation data

Observed ELF
# Model validation – ELF trajectories

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Trial</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF at 1(^{st}) time point - mean (sd)</td>
<td>9.57 (1.21)</td>
<td>9.71 (1.15)</td>
</tr>
<tr>
<td>ELF between individual SD</td>
<td>0.93</td>
<td>0.76</td>
</tr>
<tr>
<td>ELF within individual SD</td>
<td>0.53</td>
<td>0.51</td>
</tr>
<tr>
<td><strong>In those with &gt;1 measurement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increase in ELF per year</td>
<td>0.31</td>
<td>0.24</td>
</tr>
<tr>
<td>Mean ELF at baseline</td>
<td>8.73</td>
<td>8.84</td>
</tr>
</tbody>
</table>
Model validation – trial outcomes

- Comparison of the cirrhosis rates
  - In ELF arm made on the basis of ELF > 9.5.
  - In TAU arm made on the basis of symptomatic presentation and other biomarkers
- Leads to endoscopy and intervention

<table>
<thead>
<tr>
<th>Outcome</th>
<th>RCT ELF</th>
<th>RCT TAU</th>
<th>Model ELF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of cirrhosis during trial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis of cirrhosis after first measurement</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Model validation – trial outcomes

- Comparison of the cirrhosis rates
  - In ELF arm made on the basis of ELF > 9.5.
  - In TAU arm made on the basis of symptomatic presentation and other biomarkers
  - Leads to endoscopy and intervention

<table>
<thead>
<tr>
<th>Outcome</th>
<th>RCT ELF</th>
<th>RCT TAU</th>
<th>Model ELF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of cirrhosis during trial</td>
<td></td>
<td></td>
<td>13586/20000 67.9%</td>
</tr>
<tr>
<td>Diagnosis of cirrhosis after first measurement</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Model validation – trial outcomes

- Comparison of the cirrhosis rates
  - In ELF arm made on the basis of ELF>9.5.
  - In TAU arm made on the basis of symptomatic presentation and other biomarkers
  - Leads to endoscopy and intervention

<table>
<thead>
<tr>
<th>Outcome</th>
<th>RCT ELF</th>
<th>RCT TAU</th>
<th>Model ELF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of cirrhosis during trial</td>
<td></td>
<td></td>
<td>13586/20000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>67.9%</td>
</tr>
<tr>
<td>Diagnosis of cirrhosis after first measurement</td>
<td></td>
<td></td>
<td>2365/20000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.8%</td>
</tr>
</tbody>
</table>
Model validation – trial outcomes

- Comparison of the cirrhosis rates
  - In ELF arm made on the basis of ELF>9.5.
  - In TAU arm made on the basis of symptomatic presentation and other biomarkers
  - Leads to endoscopy and intervention

<table>
<thead>
<tr>
<th>Outcome</th>
<th>RCT ELF</th>
<th>RCT TAU</th>
<th>Model ELF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of cirrhosis during trial</td>
<td>281/438</td>
<td></td>
<td>13586/20000</td>
</tr>
<tr>
<td></td>
<td>64.2%</td>
<td></td>
<td>67.9%</td>
</tr>
<tr>
<td>Diagnosis of cirrhosis after first measurement</td>
<td></td>
<td></td>
<td>2365/20000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.8%</td>
</tr>
</tbody>
</table>
Model validation – trial outcomes

- Comparison of the cirrhosis rates
  - In ELF arm made on the basis of ELF>9.5.
  - In TAU arm made on the basis of symptomatic presentation and other biomarkers
  - Leads to endoscopy and intervention

<table>
<thead>
<tr>
<th>Outcome</th>
<th>RCT ELF</th>
<th>RCT TAU</th>
<th>Model ELF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of cirrhosis during trial</td>
<td>281/438</td>
<td></td>
<td>13586/20000</td>
</tr>
<tr>
<td></td>
<td>64.2%</td>
<td></td>
<td>67.9%</td>
</tr>
<tr>
<td>Diagnosis of cirrhosis after first measurement</td>
<td>84/438</td>
<td></td>
<td>2365/20000</td>
</tr>
<tr>
<td></td>
<td>29.9%</td>
<td></td>
<td>11.8%</td>
</tr>
</tbody>
</table>
Model validation – trial outcomes

- Comparison of the cirrhosis rates
  - In ELF arm made on the basis of ELF > 9.5.
  - In TAU arm made on the basis of symptomatic presentation and other biomarkers
  - Leads to endoscopy and intervention

<table>
<thead>
<tr>
<th>Outcome</th>
<th>RCT ELF</th>
<th>RCT TAU</th>
<th>Model ELF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of cirrhosis during trial</td>
<td>281/438</td>
<td>20/440</td>
<td>13586/20000</td>
</tr>
<tr>
<td></td>
<td>64.2%</td>
<td>4.5%</td>
<td>67.9%</td>
</tr>
<tr>
<td>Diagnosis of cirrhosis after first measurement</td>
<td>84/438</td>
<td>20/440</td>
<td>2365/20000</td>
</tr>
<tr>
<td></td>
<td>29.9%</td>
<td>4.5%</td>
<td>11.8%</td>
</tr>
</tbody>
</table>
Some indications that many ELF+ves are likely to be false positives

<table>
<thead>
<tr>
<th>Process / Diagnosis / Treatment</th>
<th>ELF</th>
<th>TAU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endoscopies</td>
<td>153/438 (34.9%)</td>
<td>12/440 (18.%)</td>
</tr>
<tr>
<td>Diagnosis of varices</td>
<td>9/438 (2.1%)</td>
<td>7/440 (1.6%)</td>
</tr>
<tr>
<td>Started beta-blockers or had band ligation</td>
<td>5/438 (1.1%)</td>
<td>3/440 (0.7%)</td>
</tr>
<tr>
<td>Treatment to normalise LFTs</td>
<td>57/438 (13.0%)</td>
<td>48/440 (12.0%)</td>
</tr>
</tbody>
</table>

Long term follow-up for clinical events will provide further evidence
Observations

- Comparability of the findings from the simulation and trial provide a degree of validation of the model.
- Faced challenges in obtaining necessary data.
- Model improved when using data obtained from baseline assessments in trial.
- Model findings were too late to influence trial design but help with interpretation of findings.
- High measurement error and slow progression lead to lots of false positive monitoring tests.
Conclusions

- Early modelling of monitoring strategies has an important role in
  - helping identify optimal strategies
  - make the case for or against a trial
- Developing prior to and maintaining a model during a trial seems a good idea
- Emphasises the importance of studies of biological variability and natural history