Models or end-to-end studies in monitoring studies? A case study

Universities of Birmingham, Southampton, Leeds and Oxford

This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (RP-PG-0707-10101: Evaluating the benefits for patients and the NHS of new and existing biological fluid biomarkers in liver and renal disease).

The views and opinions expressed by the authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR or the Department of Health.
Trial evidence versus linked evidence of test accuracy and treatment efficacy

ELF and Liver Fibrosis/Cirrhosis

- The ELF (Enhanced Liver Fibrosis) biomarker detects progression of fibrosis to cirrhosis
- Evidence exists of its accuracy (compared to liver biopsy)
- What is its value as a monitoring test in patients with Severe Liver Disease (SLD)?
- Can it better guide management to reduce the complications of cirrhosis that routine care?
Randomised

Eligibility and Consent

ELF Screening Test ELF Score > X

Randomised

Standard Care arm

Standard Outpatient Assessment (6 mo)

Diagnosis of Cirrhosis Clinical/biochemical/imaging

ELF arm

ELF Test at Outpatient Assessment (6 mo)

Diagnosis of Cirrhosis ELF Score > Y

Instigation of Variceal Prophylaxis, HCC screening & SBP antibiotic prophylaxis

Time to first liver related event - Incidence of morbidity or mortality
Evidence linkage methods

Monitoring strategies are rarely evidence-based:
- testing frequency based on routine care
- test thresholds chosen *ad hoc*

Can modelling

1. Help identify optimal strategies to be evaluated?
2. Predict the impact on outcomes?
Method for selection of monitoring strategies

1. Obtain data and elicit expert opinion regarding disease progression or recurrence and test performance (measurement error, accuracy, variability)
2. Simulate patient cohort modelling disease progression and results of the monitoring test based on evidence
3. Evaluate the performance of alternative monitoring strategies (different thresholds, test frequencies, decision rules)
4. Identify optimum strategies for further evaluation
Model relationship of biomarker and disease: example trajectories for 50 participants

1) ELF value at each fibrosis stage
Model relationship of disease and time: example trajectories for 50 participants

1) ELF value at each fibrosis stage

2) speed of disease progression
Modelling variability in disease stage at the start of trial

1) ELF value at each fibrosis stage
2) Speed of disease progression
3) Fibrosis stage at entry
Modelling measurement error

- Incorporating the error allows us to generate the ELF values that would be observed.
Identifying disease end points - compensated 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 compensated cirrhosis.
Implementing a monitoring strategy

- 6 monthly observations
- Threshold of 9.5
Implementing a monitoring strategy

- 6 monthly observations
- Threshold of 9.5
- $o = TP, x = FP$
Evaluating monitoring strategies

• Monitoring strategy
  – Test every 6 months
  – Duration 5 years
  – Threshold for positive result of 9.5

• Results
  – 53,336 tests for 20,000 patients
  – 2.54% of participants have delay to diagnosis over 12 months
  – Positive predictive value is 14.4%
Evaluating monitoring strategies

• Observation frequency
• Different thresholds:
  • Simple
  • Re-test (if a test value is within a specified range)
  • Absolute and relative changes from start and last value
  • Rate of change (regression)
• Outcomes to be assessed
  • number of tests
  • positive predictive value
  • delay to diagnosis.
• For a fixed PPV (25%), can compare number of tests and delay
Basic threshold strategy

% patients with delay in diagnosis of 12 months or more

Number of tests for 20,000 patients over 5 years

Basic threshold strategy
Threshold value 10.47
Observations every 6 months
No retesting
PPV 25%
Basic threshold strategy

% patients with delay in diagnosis of 12 months or more

Number of tests for 20,000 patients over 5 years

Basic threshold strategy
Threshold value 10.47
Observations every 6 months
No retesting
PPV 25%

Number of tests: 117,060
Addition of retesting

Targeted retest-at any point patients with an ELF value within 1 of the threshold have a retest. The mean of the two values is then used.

% patients with delay in diagnosis of 12 months or more

Number of tests for 20,000 patients over 5 years
% patients with delay in diagnosis of 12 months or more

Addition of retesting

Worse strategy
Increase in number of tests and in patients with delay

Basic threshold strategy
Threshold value 10.355
Observations every 6 months
Retesting
PPV 25%

Number of tests for 20,000 patients over 5 years
Decreasing observation frequency

% patients with delay in diagnosis of 12 months or more

Number of tests for 20,000 patients over 5 years

Number of tests

Delay (%)

% patients with delay in diagnosis of 12 months or more

0 50000 100000 150000 200000 250000

0 10 20 30 40

Basic threshold strategy
Threshold value 10.31
Observations every 12 months
No retesting
PPV 25%

UNIVERSITY OF BIRMINGHAM
% patients with delay in diagnosis of 12 months or more

Decrease in number of tests and slight increase in patients with delay

Number of tests for 20,000 patients over 5 years

Basic threshold strategy
Threshold value 10.31
Observations every 12 months
No retesting
PPV 25%
% patients with delay in diagnosis of 12 months or more

Number of tests for 20,000 patients over 5 years
Absolute increase from start value

% patients with delay in diagnosis of 12 months or more

Slight decrease in number of tests and slight increase in patients with delay

Number of tests for 20,000 patients over 5 years

Absolute increase from start value strategy
Trigger value 1.065
Observations every 6 months
No retesting
PPV 25%

117060
115365
14.46
+2.16
-1695

16.62
Linear regression

% patients with delay in diagnosis of 12 months or more

Number of tests for 20,000 patients over 5 years

Linear regression strategy
Trigger value 10.26
Observations every 6 months
No retesting
PPV 25%
% patients with delay in diagnosis of 12 months or more

Slightly better strategy
Slight decrease in number of tests and slight decrease in patients with delay

Linear regression strategy
Trigger value 10.26
Observations every 6 months
No retesting
PPV 25%

UNIVERSITY OF BIRMINGHAM
All strategies

% patients with delay in diagnosis of 12 months or more

Number of tests for 20,000 patients over 5 years
Data required

<table>
<thead>
<tr>
<th>Step</th>
<th>Data required</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural history</td>
<td>Progression of fibrosis</td>
<td>Questionable</td>
</tr>
<tr>
<td>Natural history</td>
<td>Variability at presentation</td>
<td>Questionable</td>
</tr>
<tr>
<td>Test performance</td>
<td>Measurement error in ELF</td>
<td>Contradictory</td>
</tr>
<tr>
<td>Test performance</td>
<td>Accuracy of ELF</td>
<td>Fair</td>
</tr>
</tbody>
</table>

Planned validation by comparison of the predicted ELF values with those observed in the ELF arm of ELUCIDATE
## Predicting impact on patient outcomes

<table>
<thead>
<tr>
<th>Step</th>
<th>Data required</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural history</td>
<td>Progression of fibrosis</td>
<td>Questionable</td>
</tr>
<tr>
<td>Natural history</td>
<td>Variability at presentation</td>
<td>Questionable</td>
</tr>
<tr>
<td>Test performance</td>
<td>Measurement error in ELF</td>
<td>Contradictory</td>
</tr>
<tr>
<td>Test performance</td>
<td>Accuracy of ELF</td>
<td>Fair</td>
</tr>
<tr>
<td>Natural history</td>
<td>Occurrence of liver related events</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>Effectiveness of interventions</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

• Many input variables are required for linked evidence of monitoring tests
• Hindered by poor data on disease progression or recurrence, and test performance
• Modelling allows candidate strategies for to be identified and compared, and assessment of the feasibility of an RCT
• The impact on patient outcomes may stretch the data too far