UNIVERSITY OF BIRMINGHAM

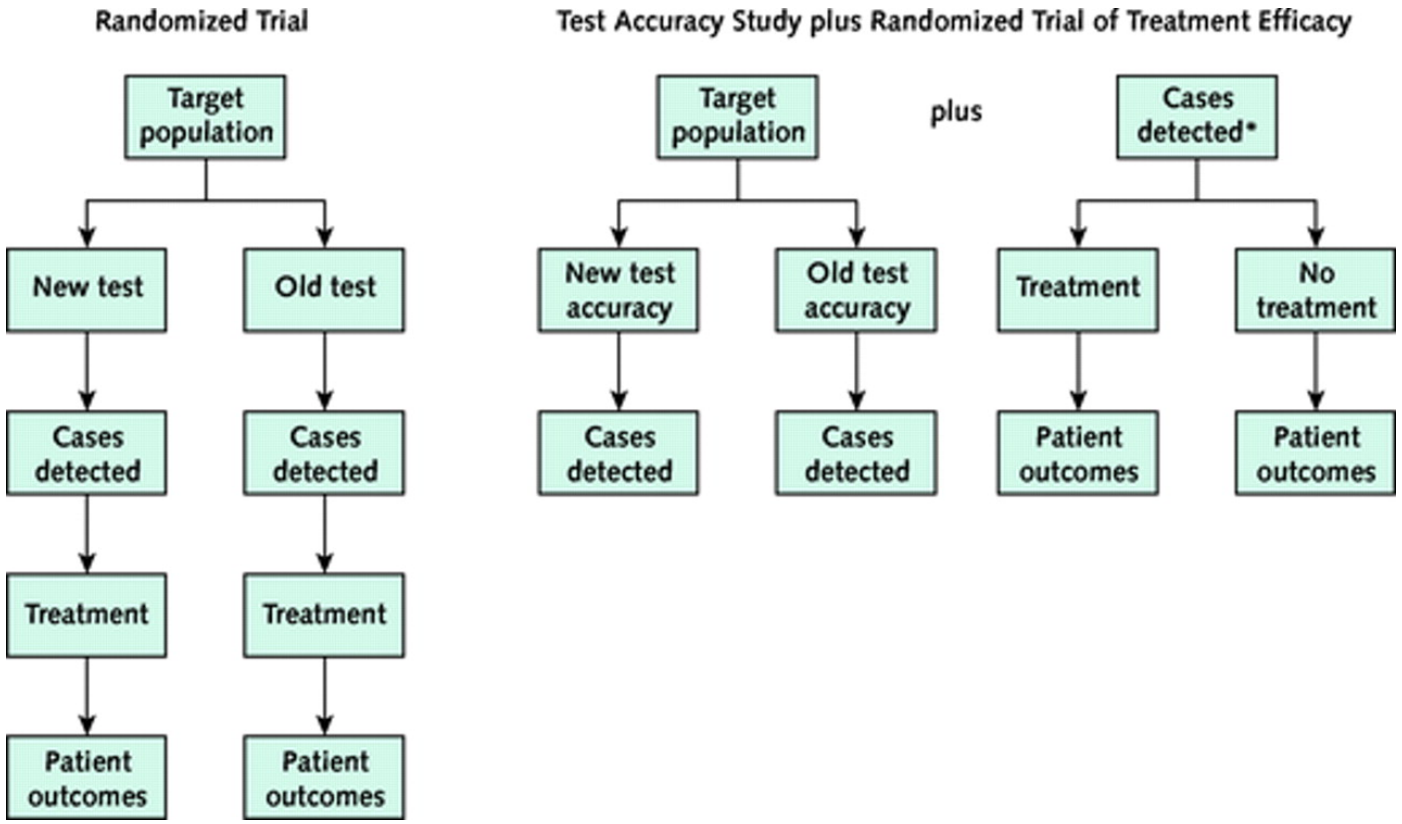
Models or end-to-end studies in monitoring studies? A case study

Deeks J, Sitch A, Dinnes J, Parkes J, Gregory W, Hewison J, Altman D
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

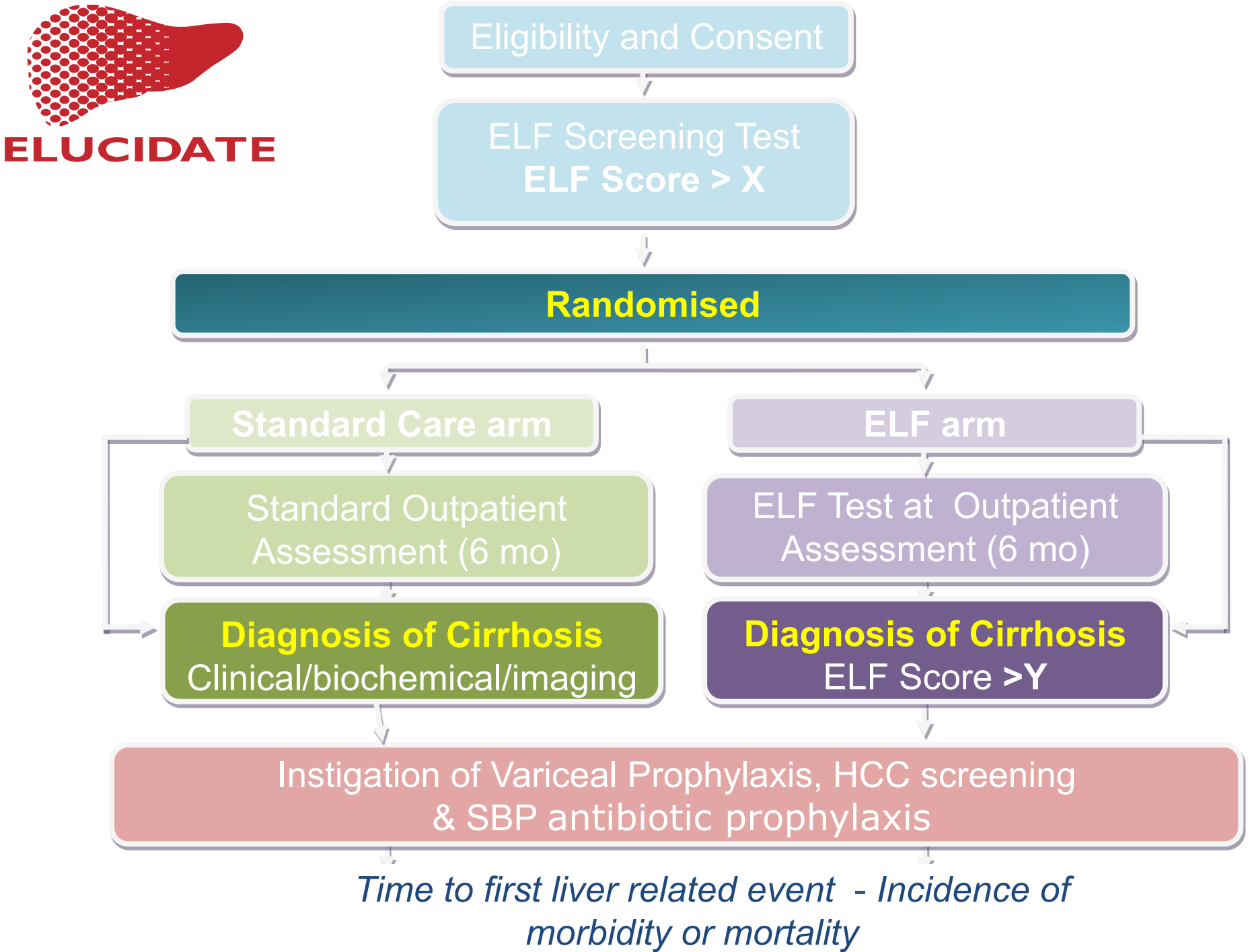


Lord, S. J. et. al. Ann Intern Med 2006;144:850-855

Annals of Internal Medicine

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?


ELUCIDATE
Eligibility and Consent
ELF Screening Test
ELF Score > X
Randomised
Standard Care arm
ELF arm
Standard Outpatient Assessment (6 mo)
ELF Test at Outpatient Assessment (6 mo)
Diagnosis of Cirrhosis
Clinical/biochemical/imaging
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

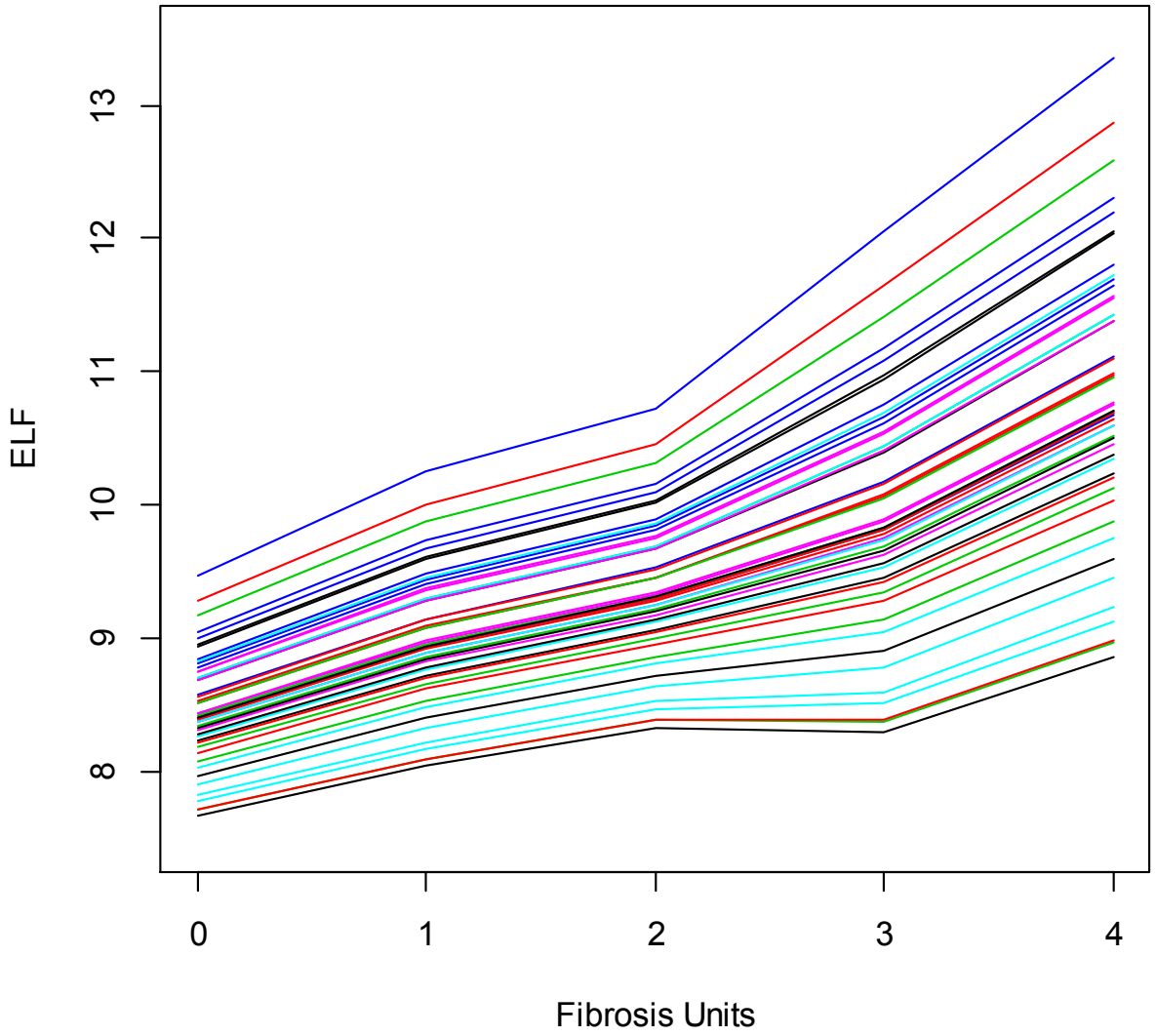
Obtain data and elicit expert opinion regarding disease progression or recurrence and test performance (measurement error, accuracy, variability)

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

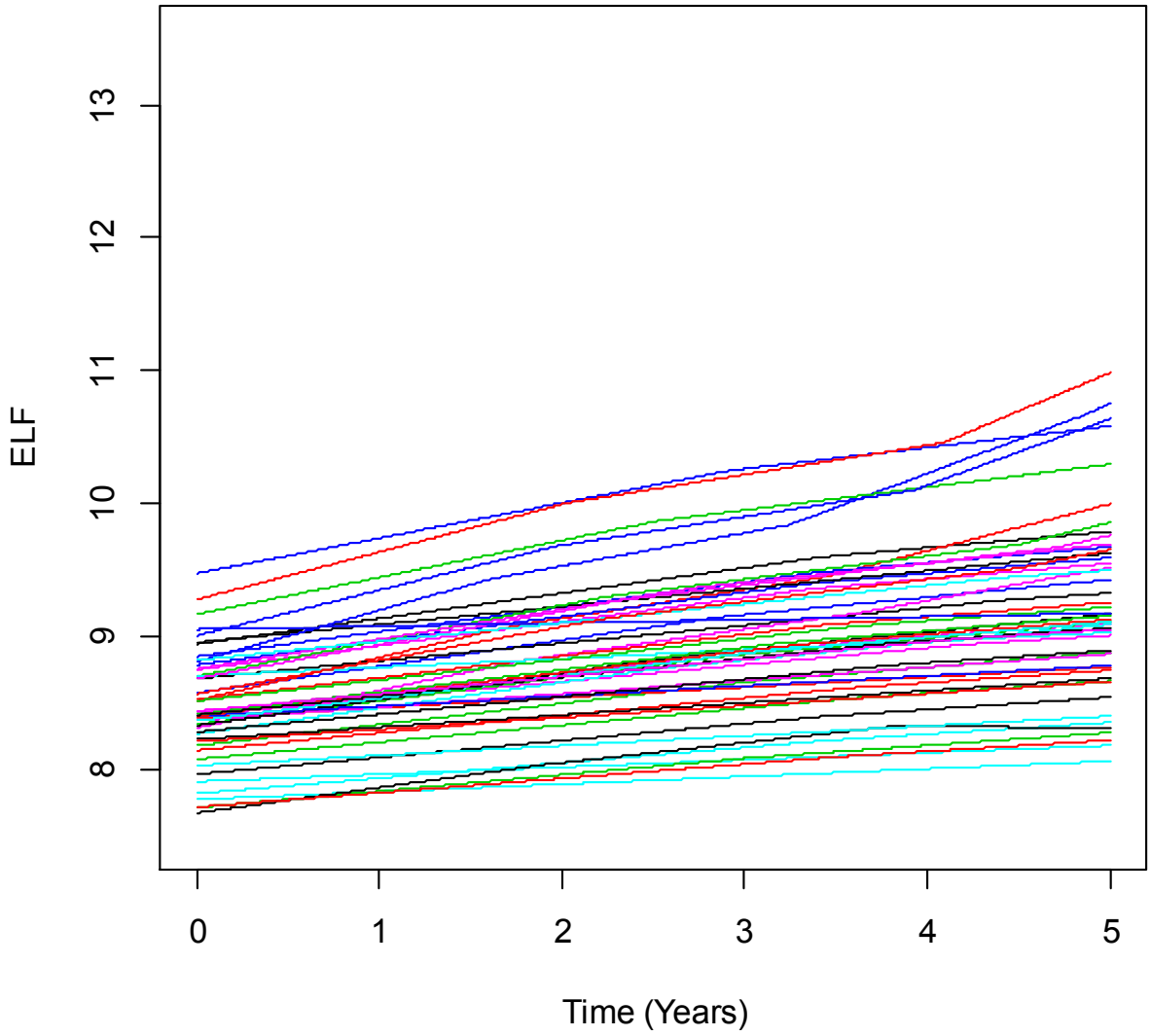
Model relationship of biomarker and disease: example trajectories for 50 participants



1) ELF value at each fibrosis stage

UNIVERSITY OF BIRMINGHAM

Model relationship of disease and time: example trajectories for 50 participants

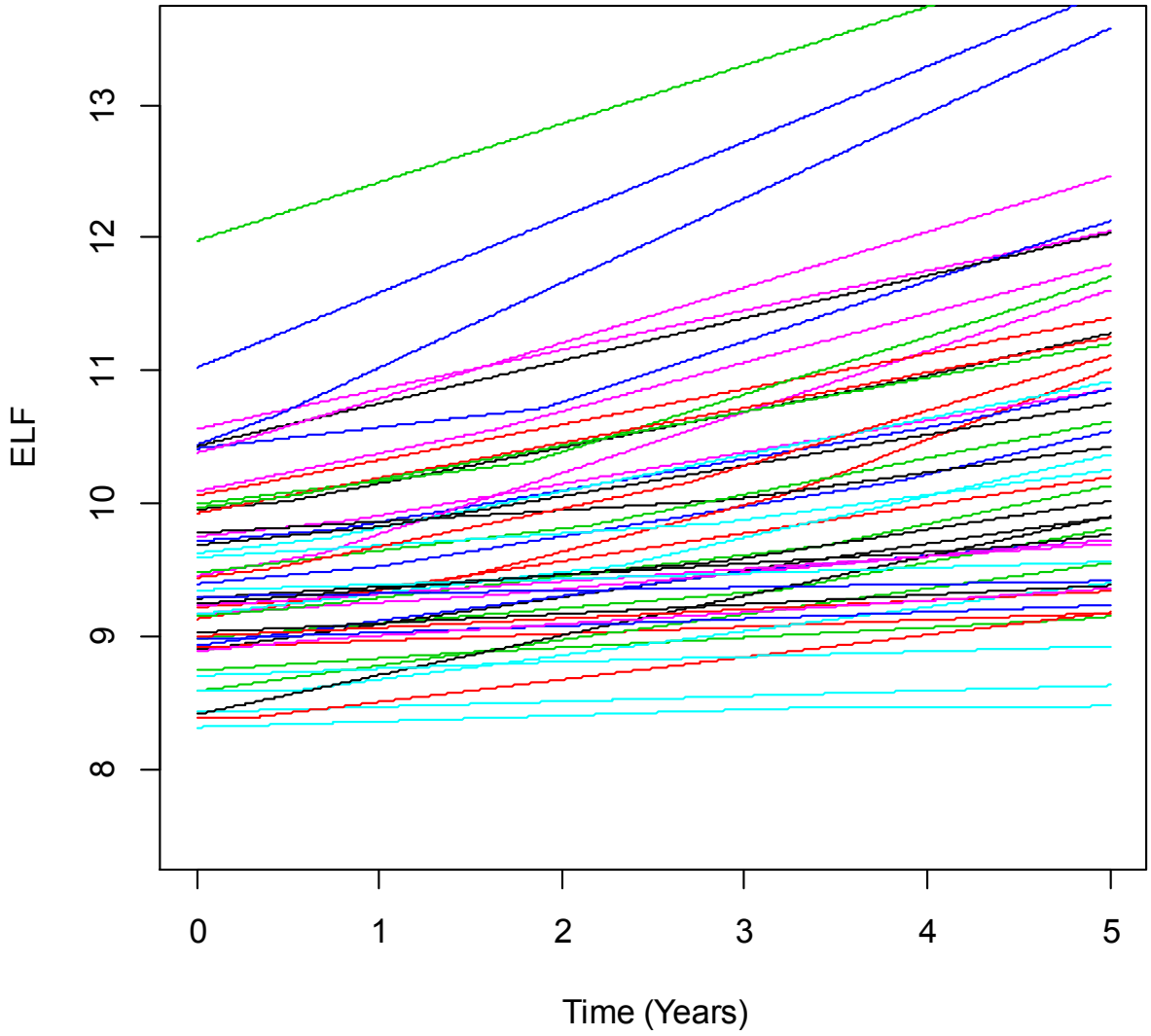


1) ELF value at each fibrosis stage

2) speed of disease progression

UNIVERSITY OF BIRMINGHAM

Modelling variability in disease stage at the start of trial



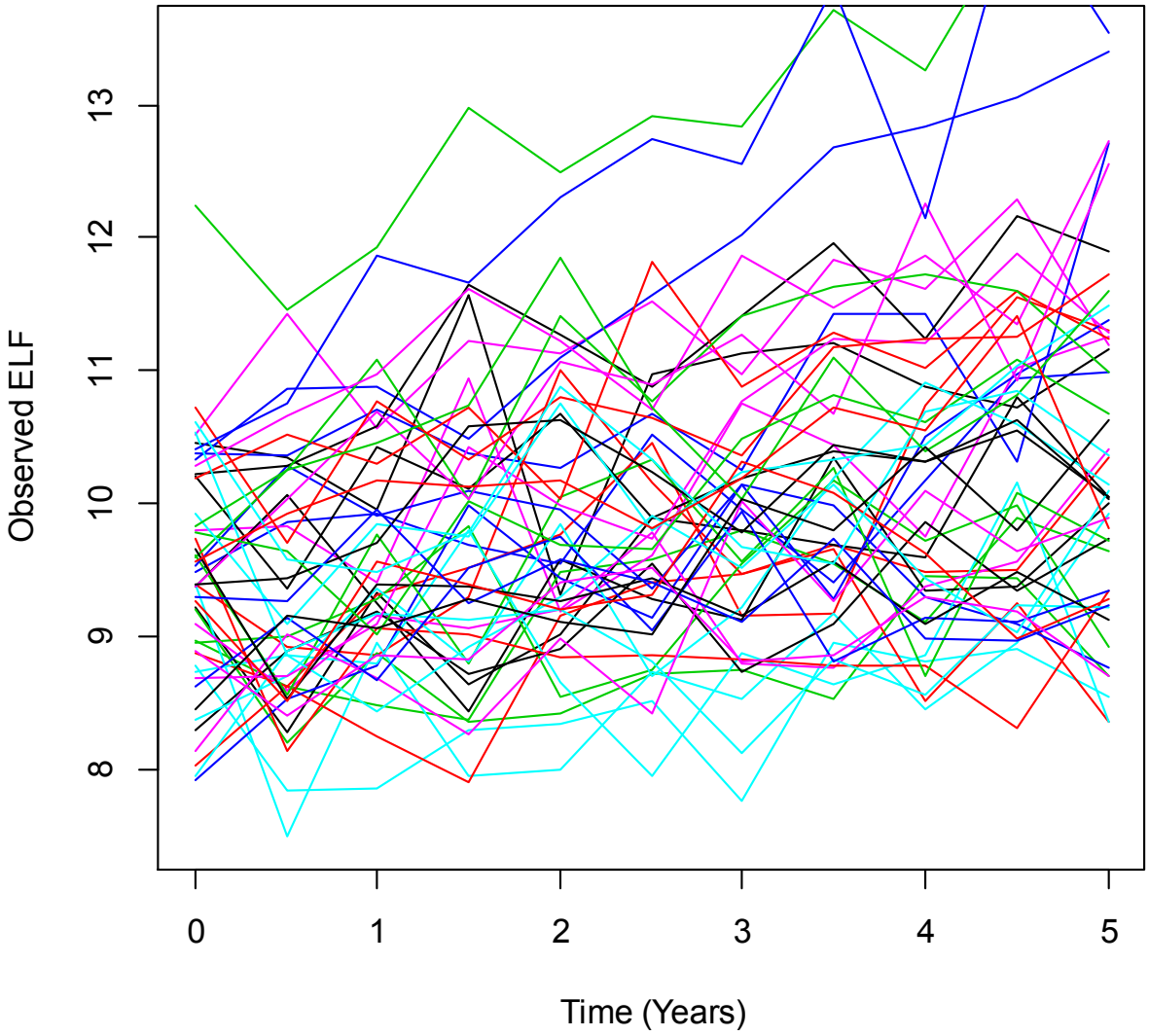
1) ELF value at each fibrosis stage

2) speed of disease progression

3) fibrosis stage at entry

UNIVERSITY OF BIRMINGHAM

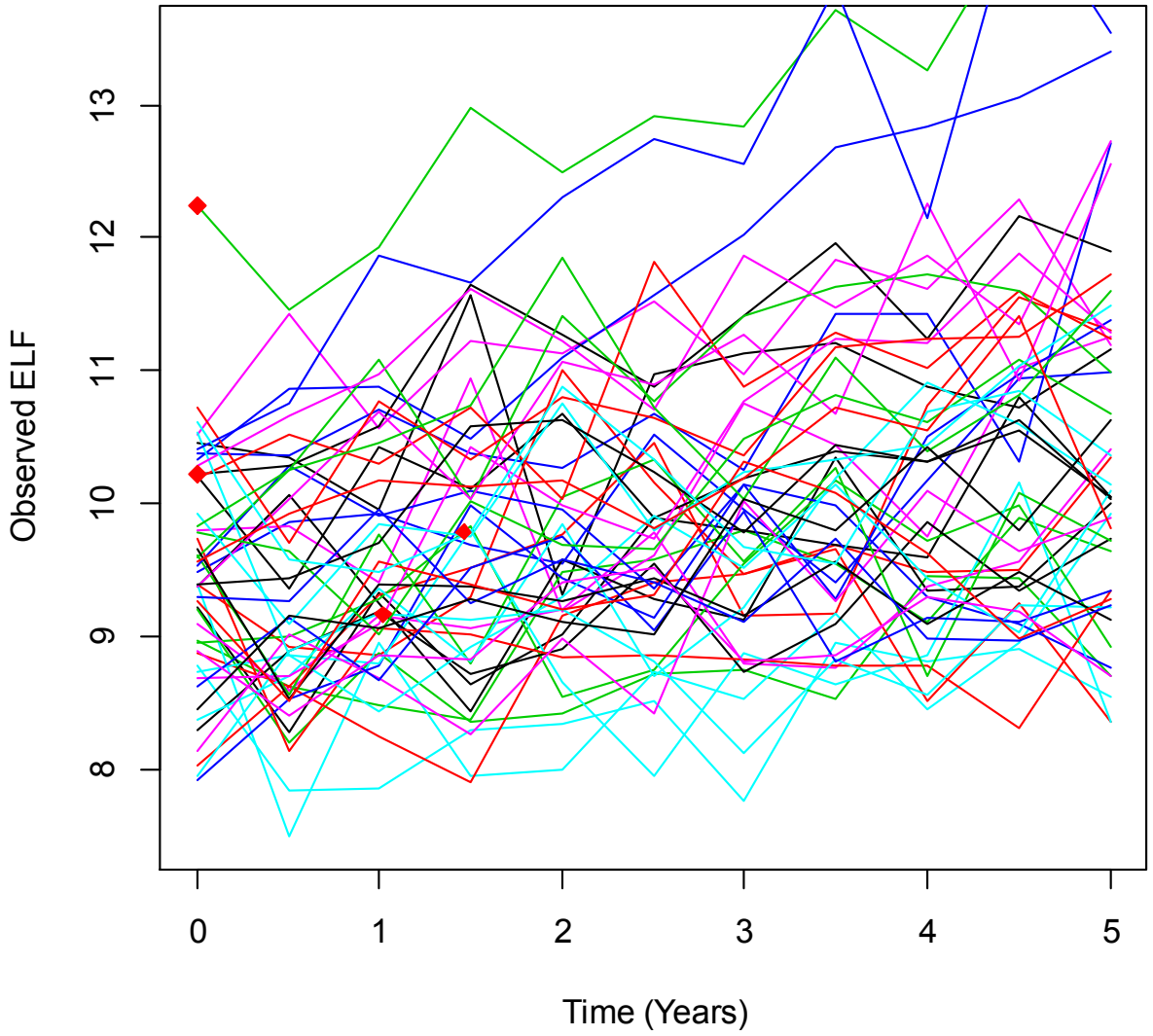
Modelling measurement error



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

UNIVERSITY OF BIRMINGHAM

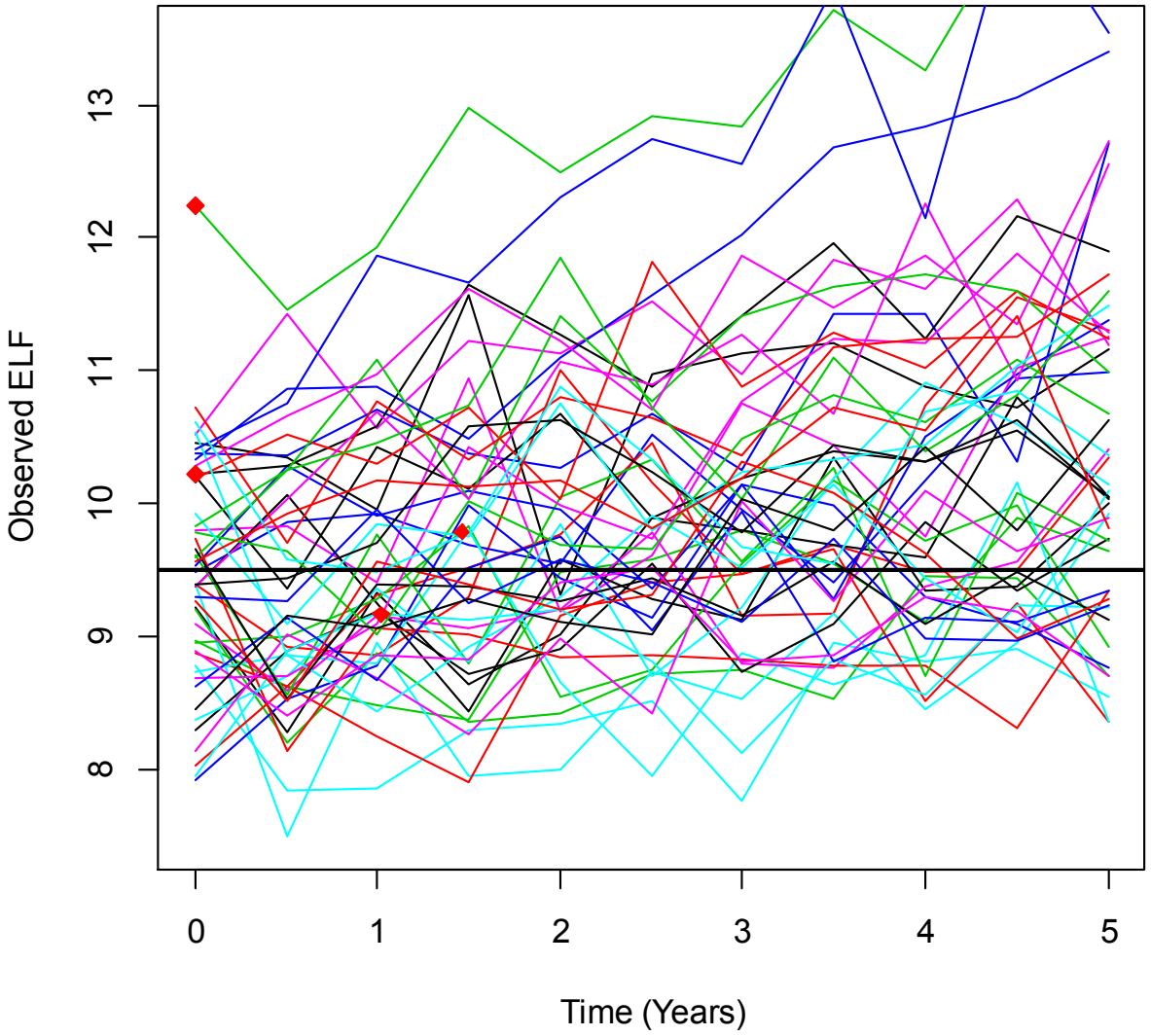
Identifying disease end points- compensated cirrhosis



- 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.

UNIVERSITY OF BIRMINGHAM

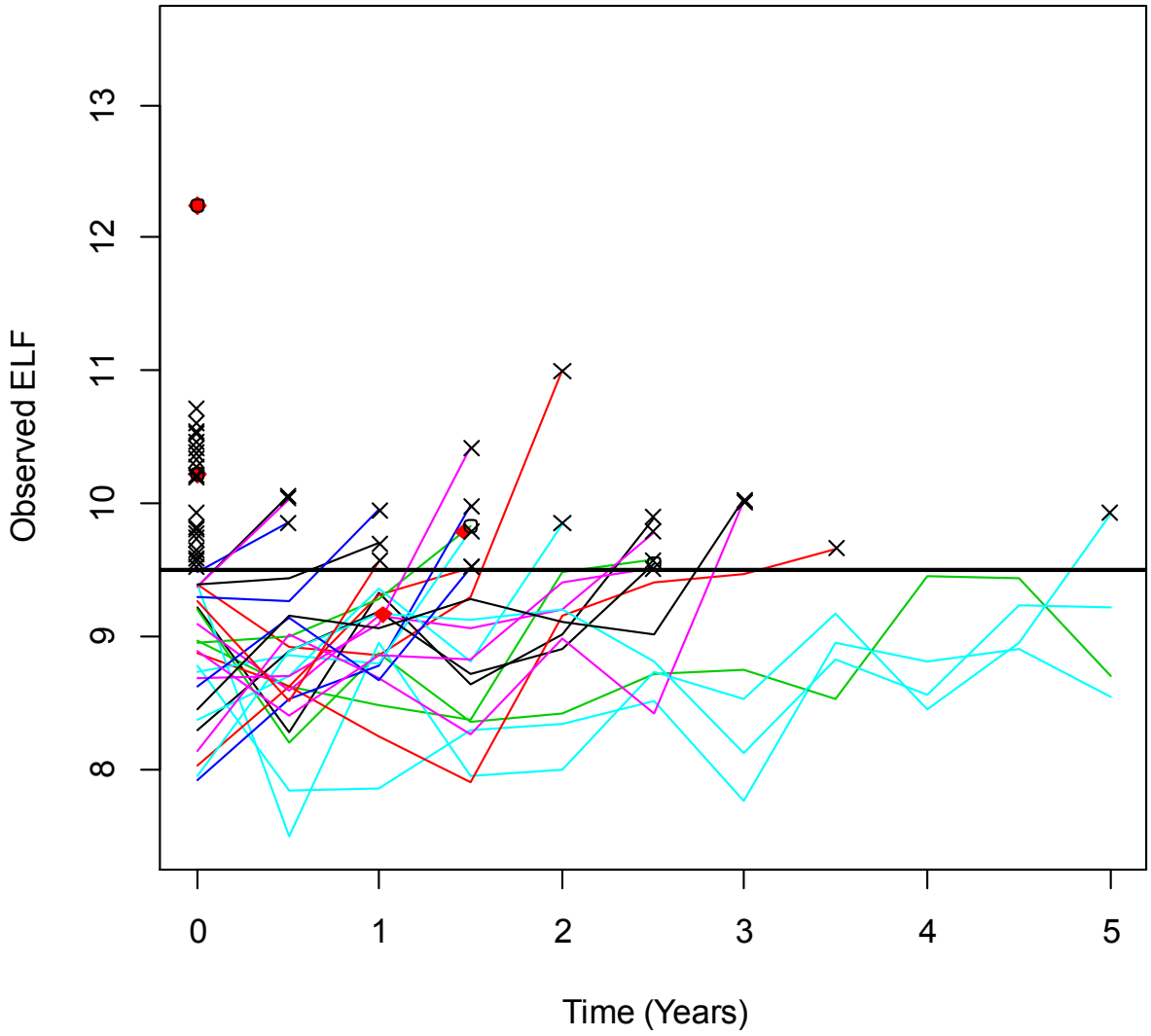
Implementing a monitoring strategy



- 6 monthly observations
- Threshold of 9.5

UNIVERSITY OF BIRMINGHAM

Implementing a monitoring strategy



- 6 monthly observations
- Threshold of 9.5
- o= TP, x=FP

UNIVERSITY OF BIRMINGHAM

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%

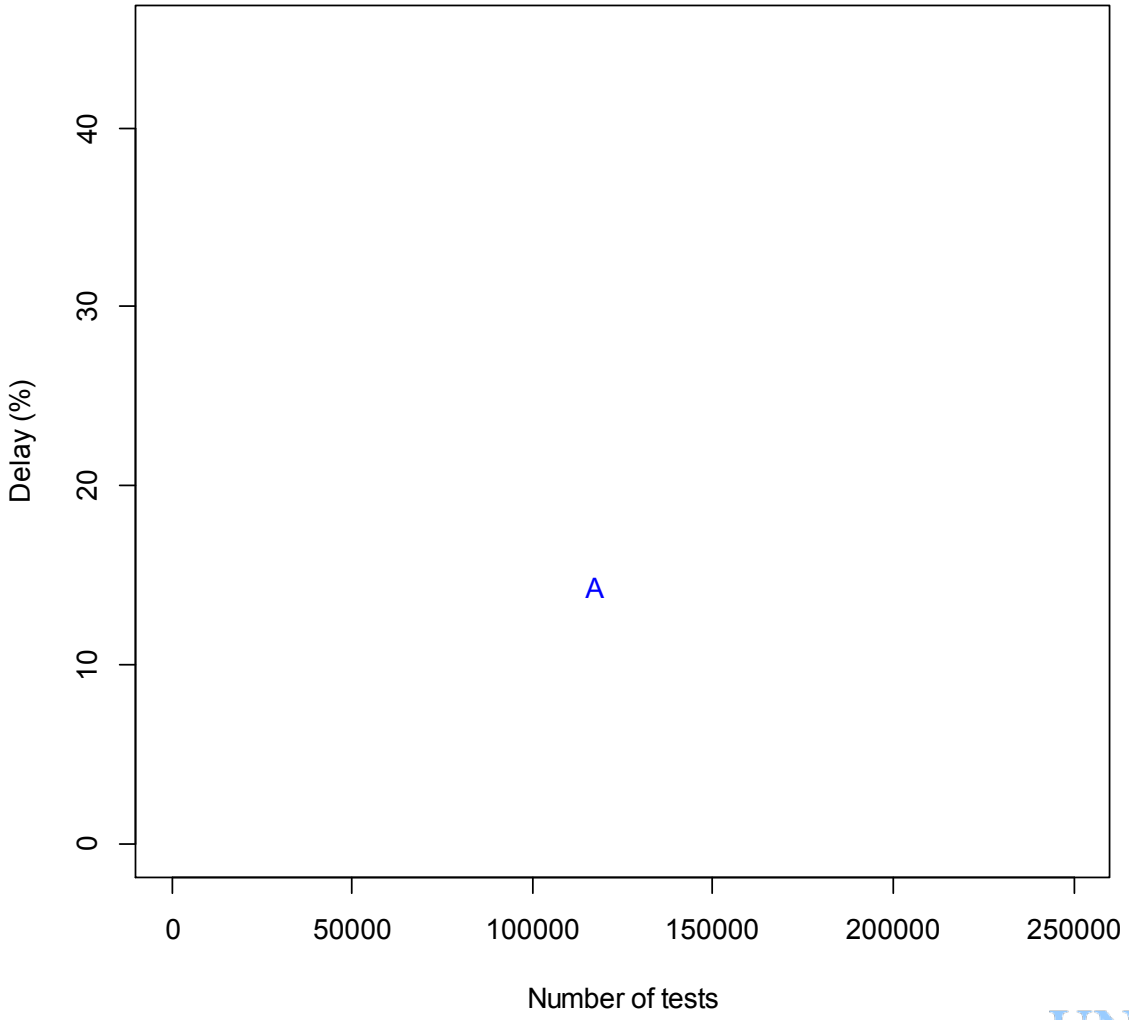
UNIVERSITY OF BIRMINGHAM

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



A

Basic threshold strategy

Threshold value 10.47

Observations every 6 months

No retesting

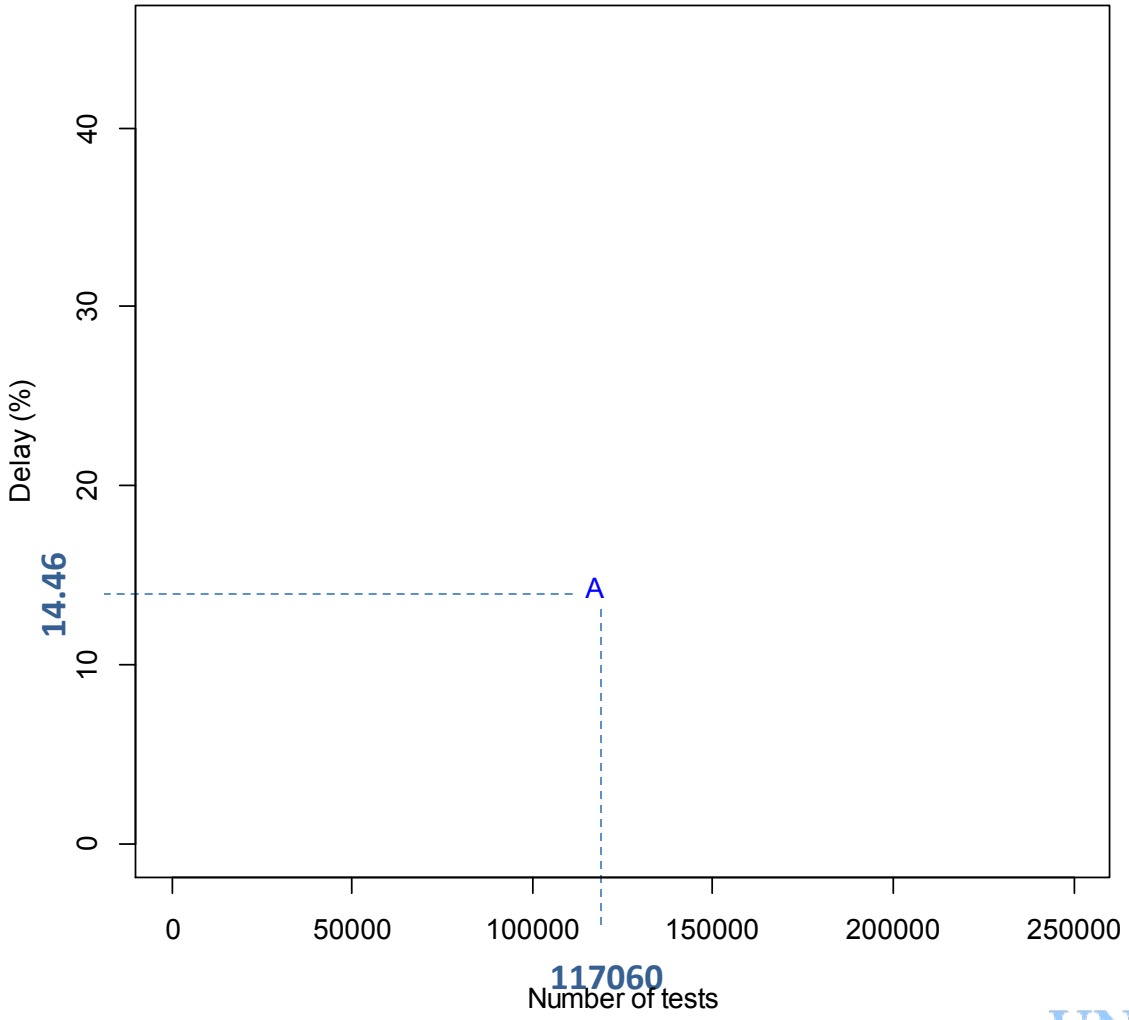
PPV 25%

Number of tests for 20,000 patients over 5 years

UNIVERSITY OF BIRMINGHAM

Basic threshold strategy

% patients with delay in diagnosis of 12 months or more



A

Basic threshold strategy

Threshold value 10.47

Observations every 6 months

No retesting

PPV 25%

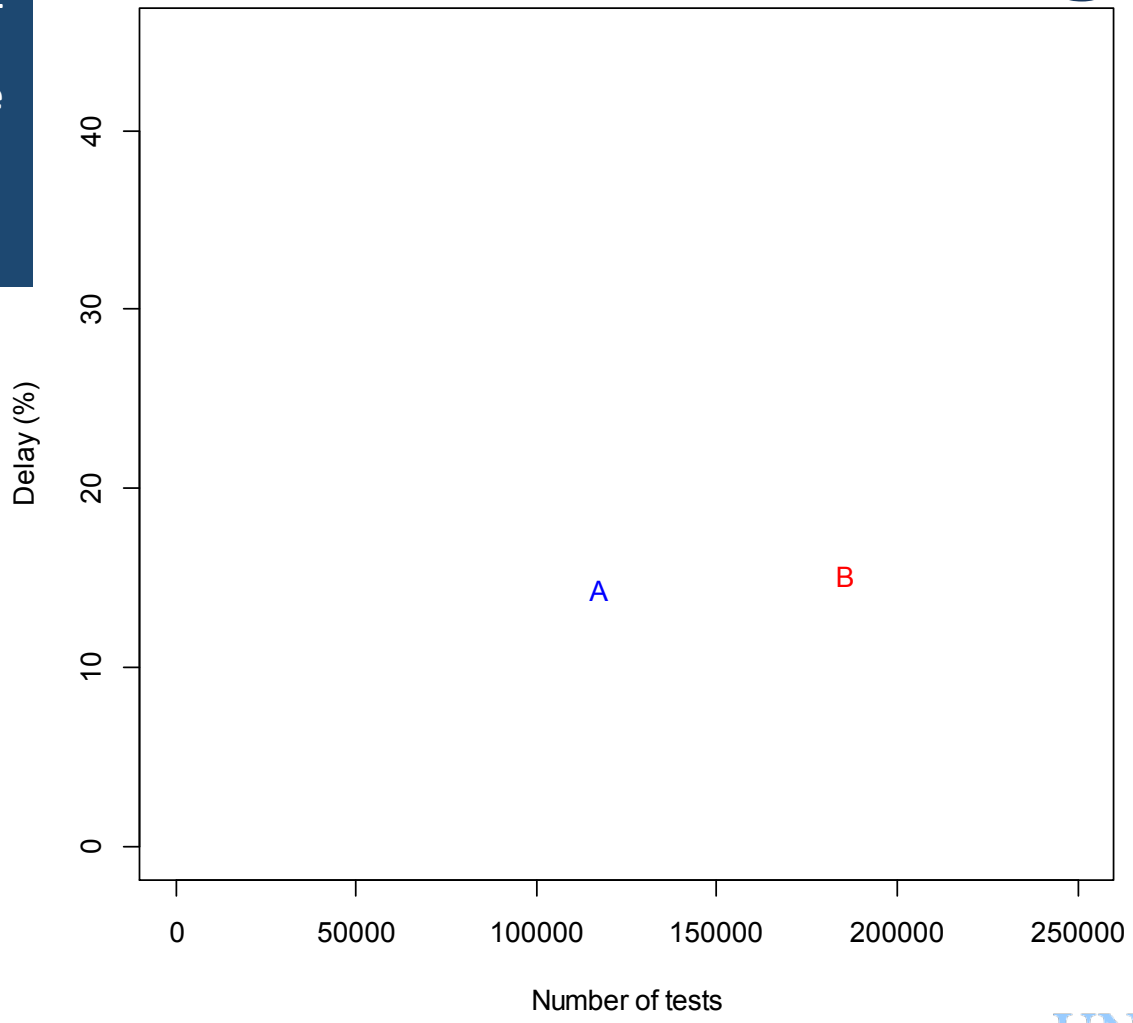
Number of tests for 20,000 patients over 5 years

UNIVERSITY OF BIRMINGHAM

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



B

Basic threshold strategy

Threshold value 10.355

Observations every 6 months

Retesting

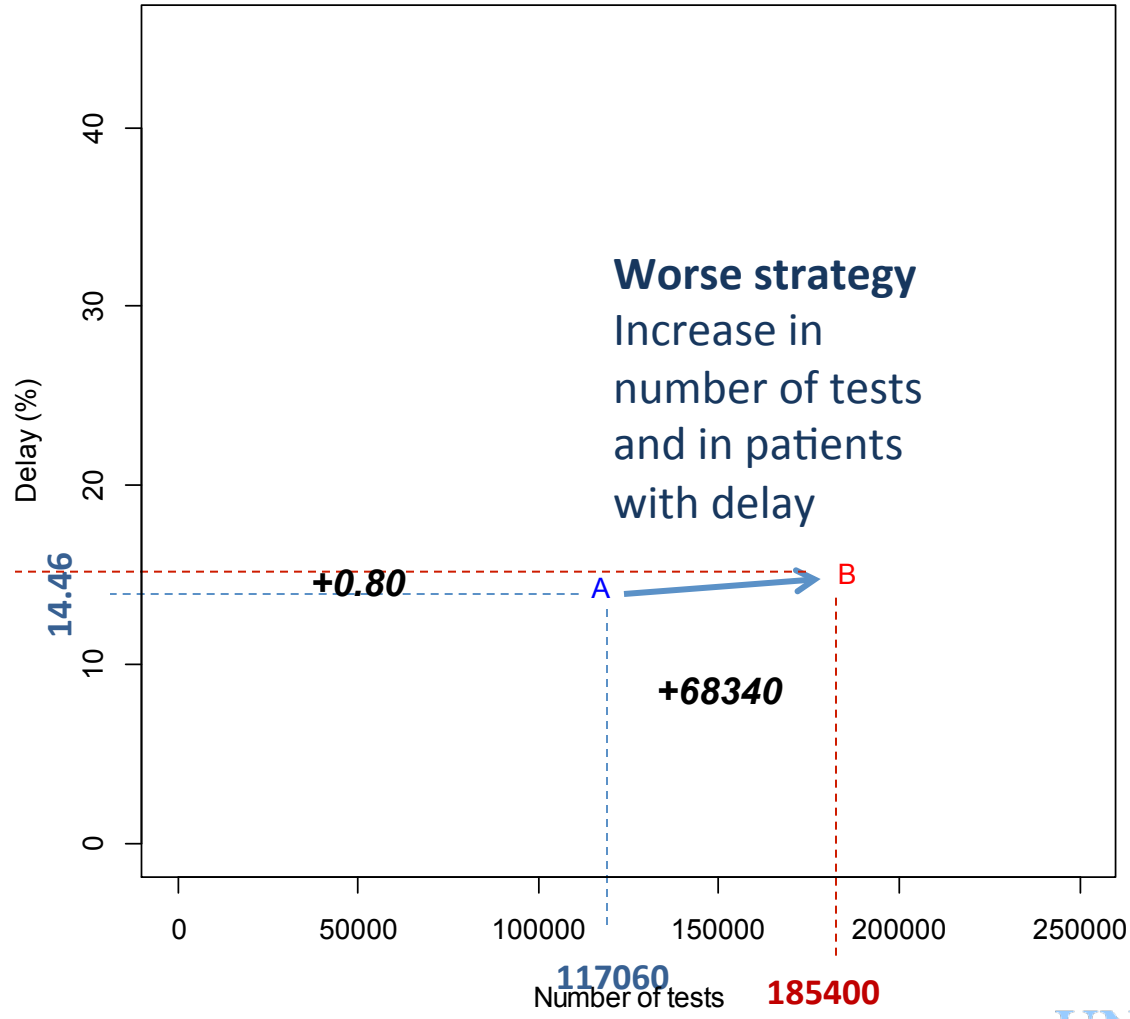
PPV 25%

Number of tests for 20,000 patients over 5 years

UNIVERSITY OF BIRMINGHAM

Addition of retesting

% patients with delay in diagnosis of 12 months or more



Number of tests for 20,000 patients over 5 years

B

Basic threshold strategy

Threshold value 10.355

Observations every 6 months

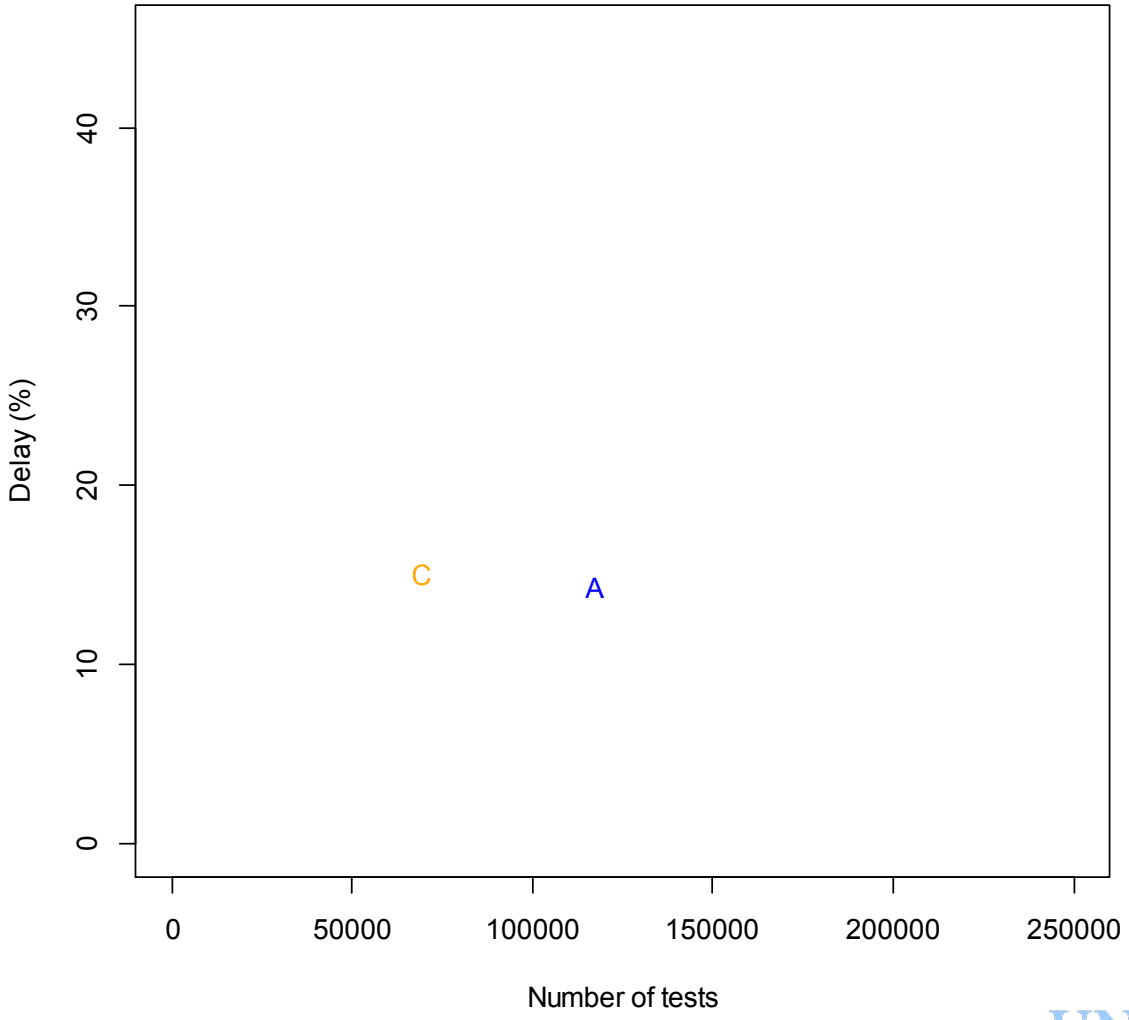
Retesting

PPV 25%

UNIVERSITY OF BIRMINGHAM

Decreasing observation frequency

% patients with delay in diagnosis of 12 months or more



C

Basic threshold strategy

Threshold value 10.31

Observations every 12 months

No retesting

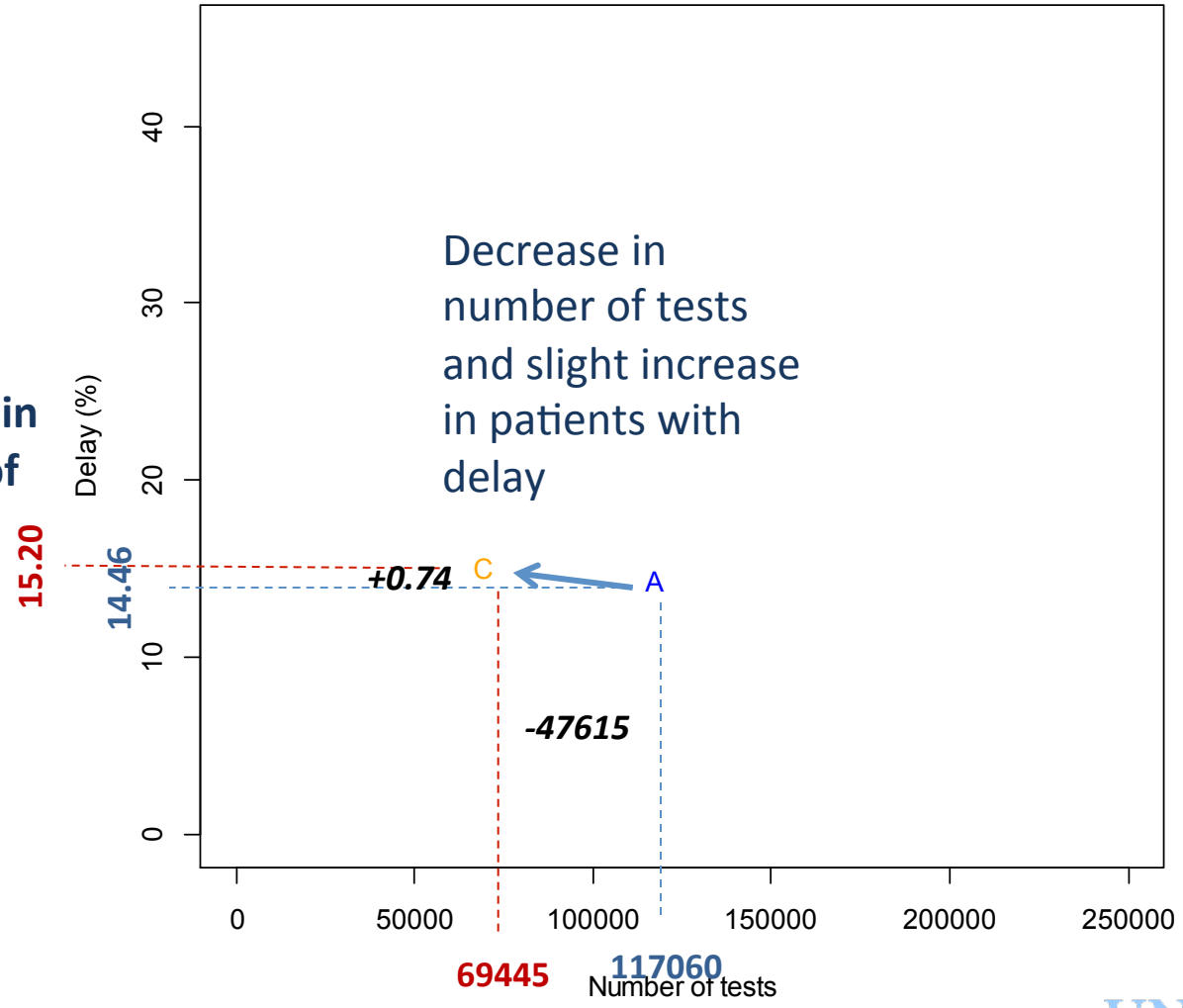
PPV 25%

Number of tests for 20,000 patients over 5 years

UNIVERSITY OF BIRMINGHAM

Decreasing observation frequency

% patients with delay in diagnosis of 12 months or more



Number of tests for 20,000 patients over 5 years

C

Basic threshold strategy

Threshold value 10.31

Observations every 12 months

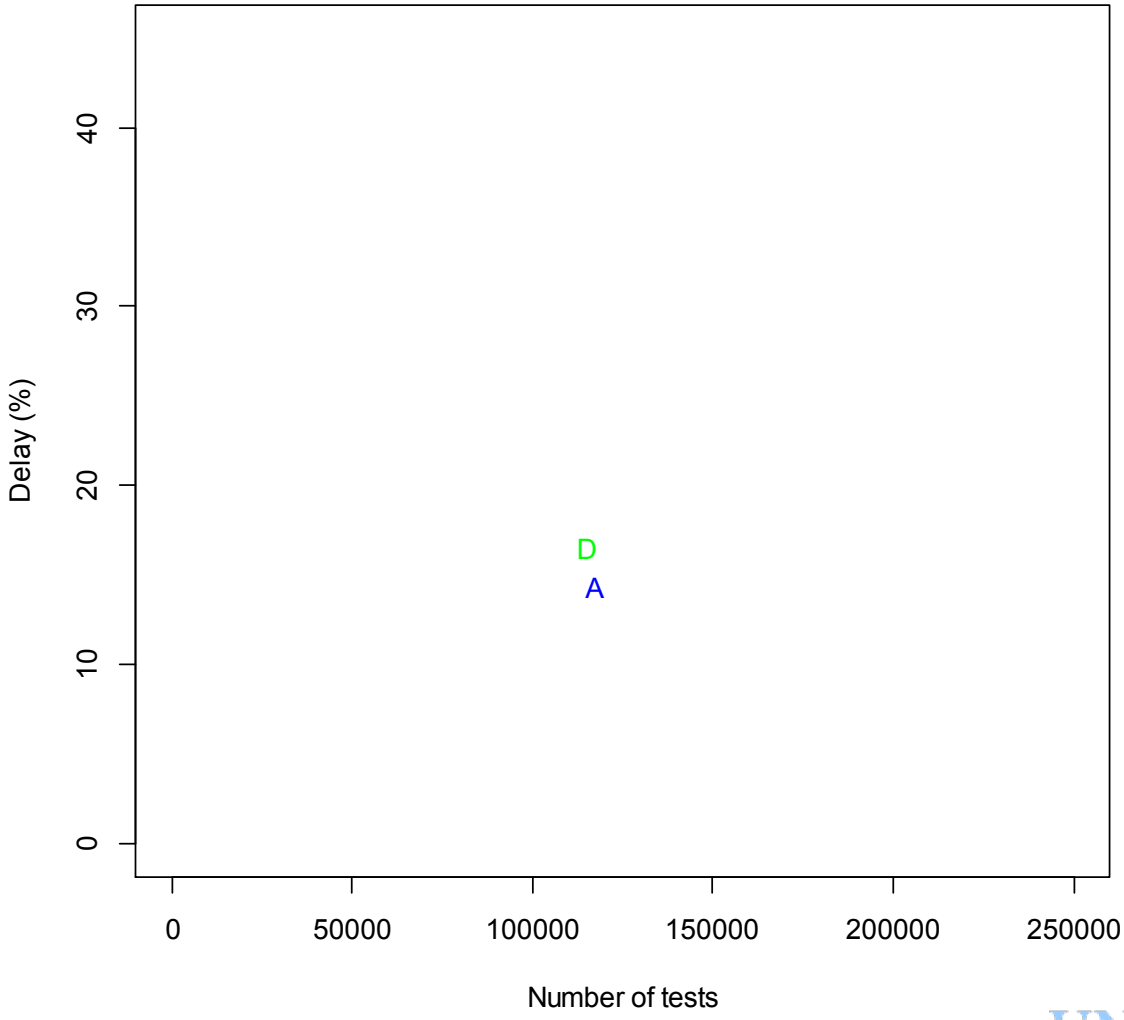
No retesting

PPV 25%

UNIVERSITY OF BIRMINGHAM

Absolute increase from start value

% patients with delay in diagnosis of 12 months or more



D

Absolute increase from start value strategy

Trigger value 1.065

Observations every 6 months

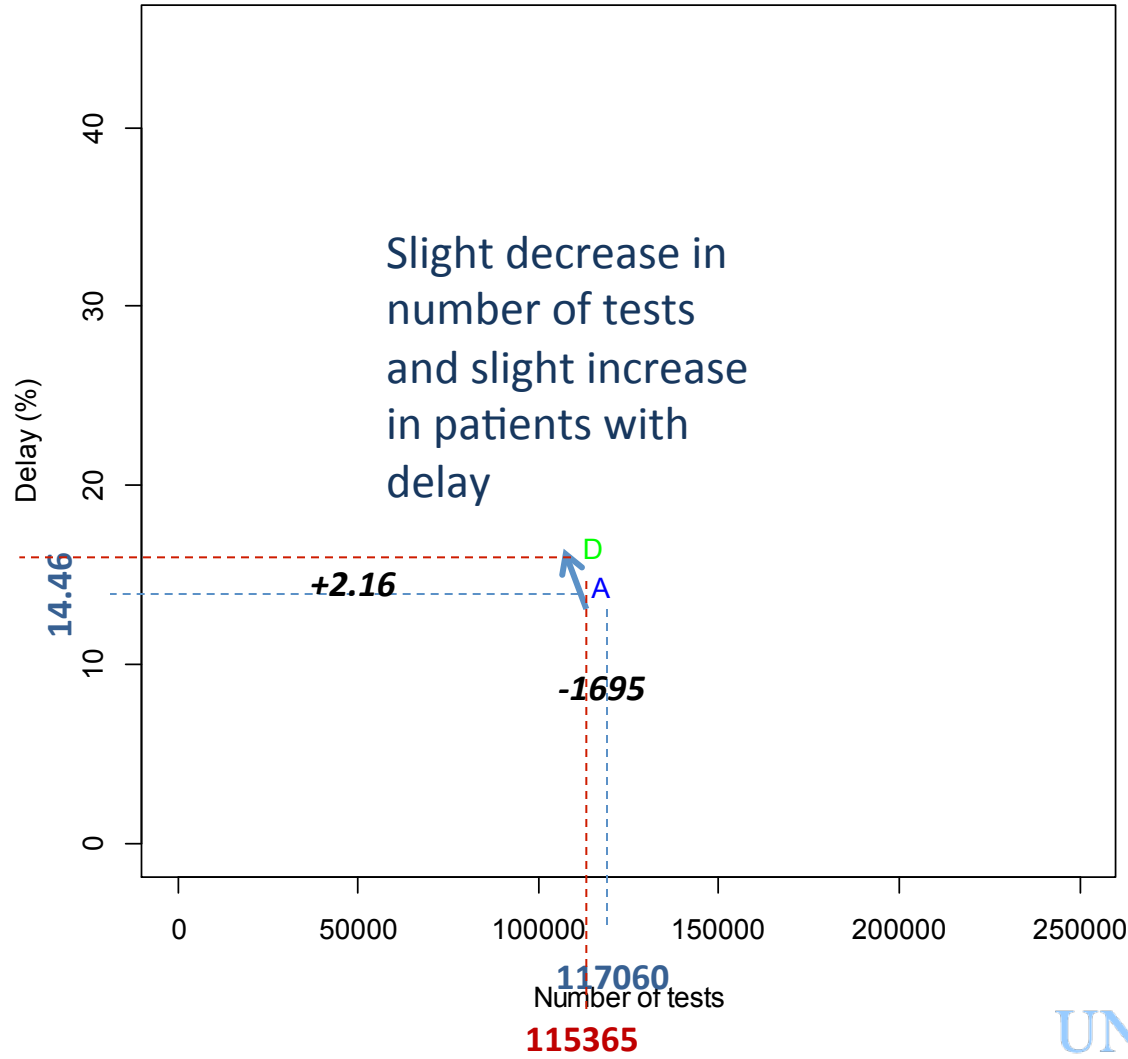
No retesting

PPV 25%

Number of tests for 20,000 patients over 5 years

UNIVERSITY OF BIRMINGHAM

Absolute increase from start value



% patients with delay in diagnosis of 12 months or more

Number of tests for 20,000 patients over 5 years

D

Absolute increase from start value strategy

Trigger value 1.065

Observations every 6 months

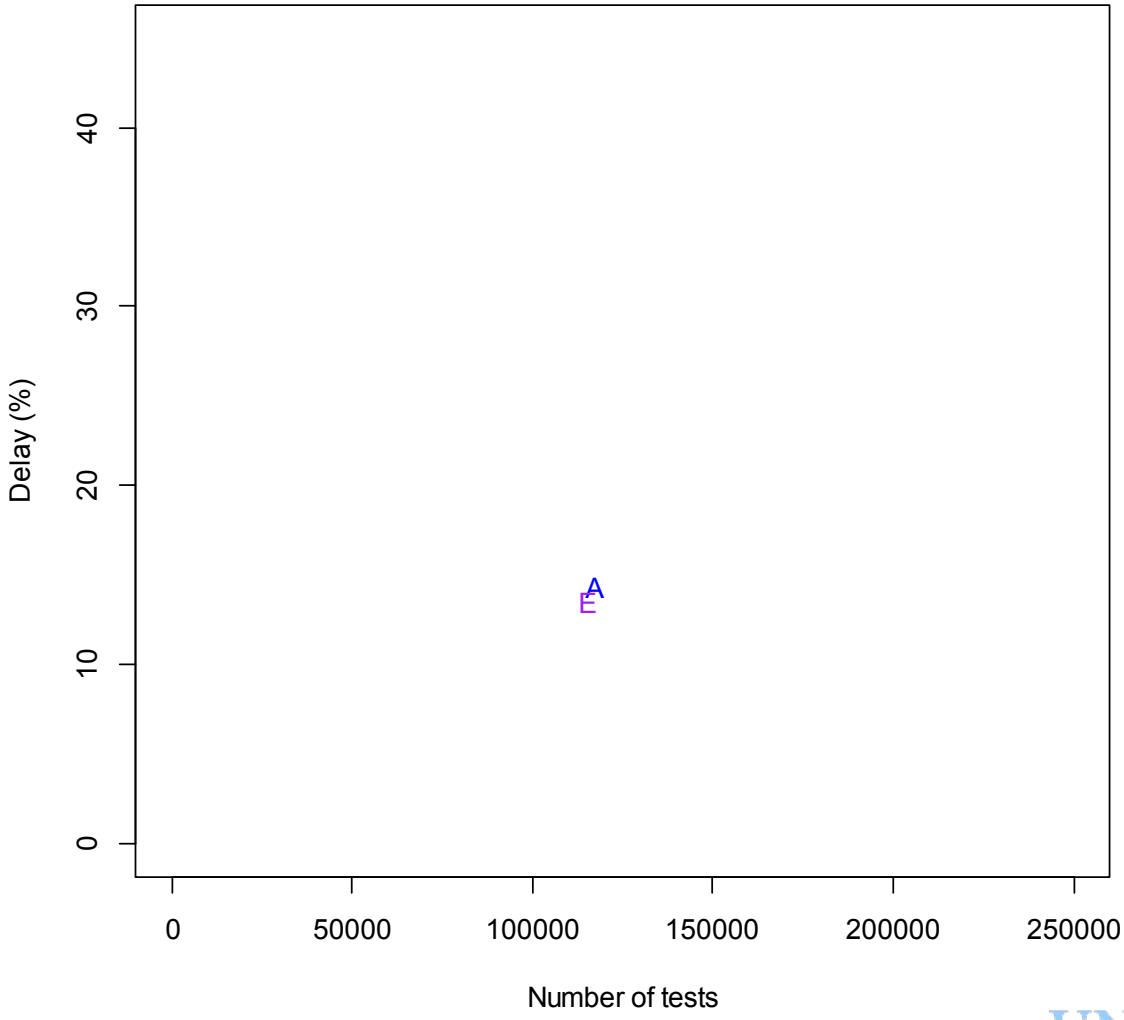
No retesting

PPV 25%

UNIVERSITY OF BIRMINGHAM

Linear regression

% patients with delay in diagnosis of 12 months or more



E

Linear regression strategy

Trigger value 10.26

Observations every 6 months

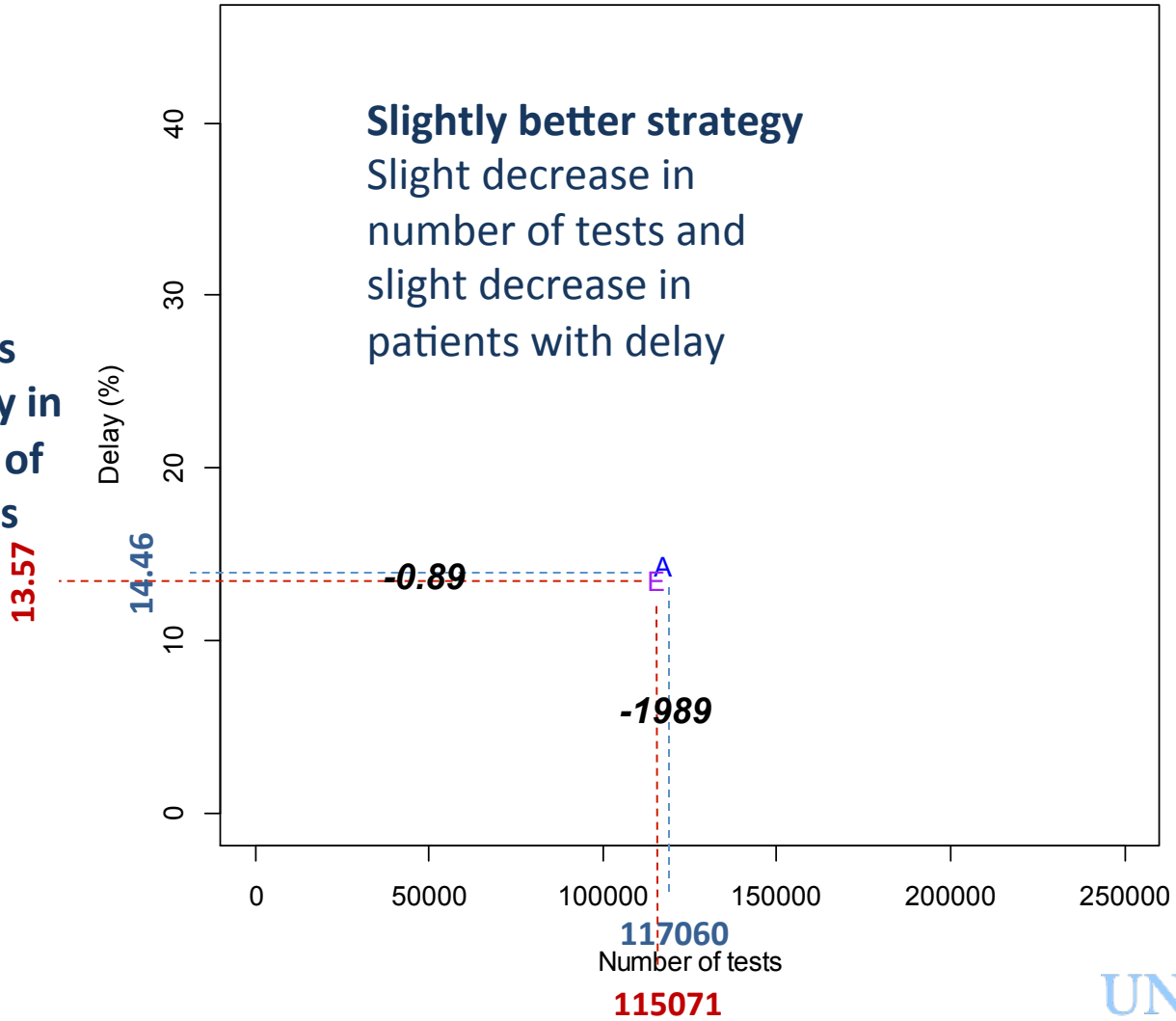
No retesting

PPV 25%

Number of tests for 20,000 patients over 5 years

UNIVERSITY OF BIRMINGHAM

Linear regression



% patients with delay in diagnosis of 12 months or more

Number of tests for 20,000 patients over 5 years

E

Linear regression strategy

Trigger value 10.26

Observations every 6 months

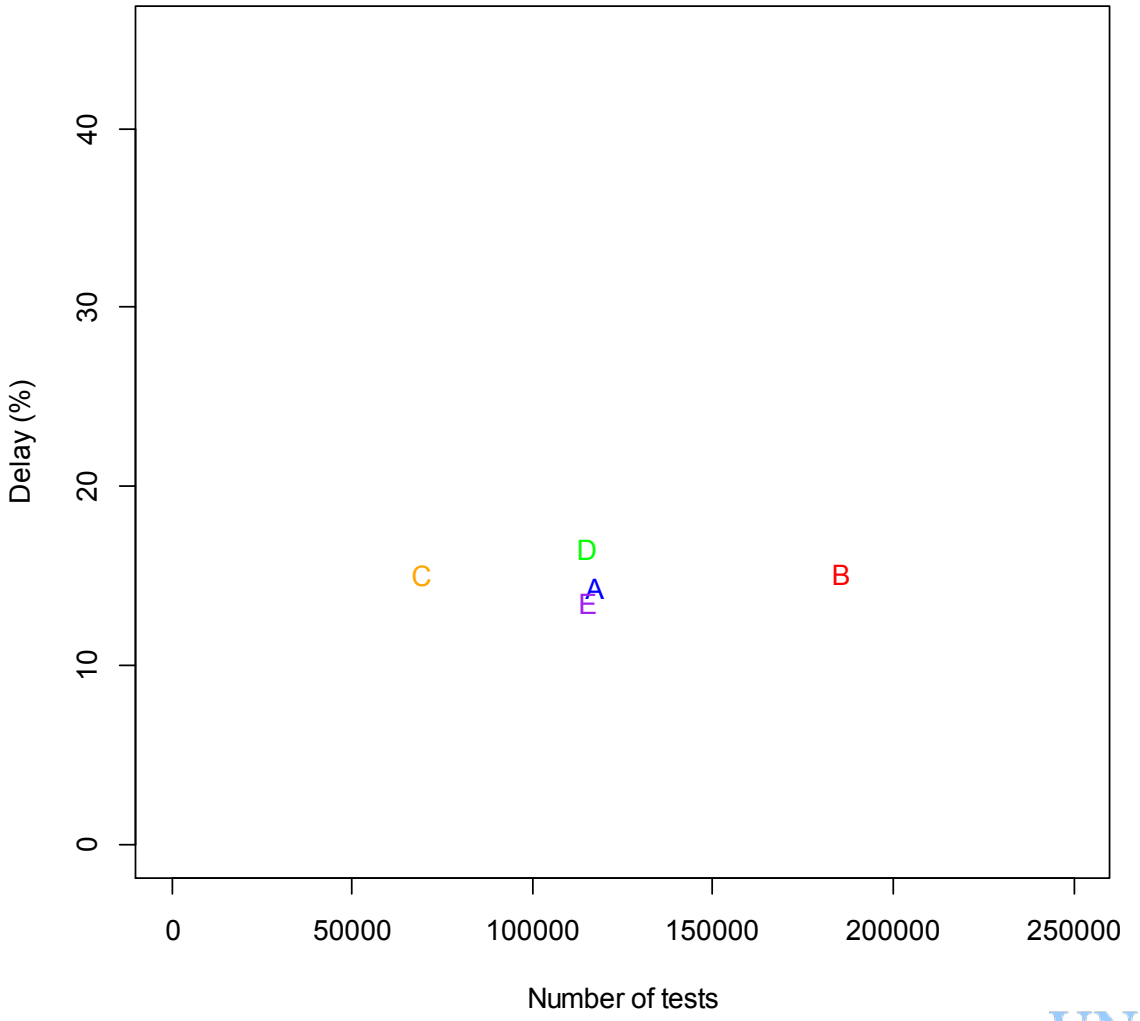
No retesting

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

UNIVERSITY OF BIRMINGHAM

Data required

Step	Data required	Quality
Natural history	Progression of fibrosis	Questionable
Natural history	Variability at presentation	Questionable
Test performance	Measurement error in ELF	Contradictory
Test performance	Accuracy of ELF	Fair

Planned validation by comparison of the predicted ELF values with those observed in the ELF arm of ELUCIDATE

UNIVERSITY OF BIRMINGHAM

Predicting impact on patient outcomes

Step	Data required	Quality
Natural history	Progression of fibrosis	Questionable
Natural history	Variability at presentation	Questionable
Test performance	Measurement error in ELF	Contradictory
Test performance	Accuracy of ELF	Fair
Natural history	Occurrence of liver related events	
Treatment	Effectiveness of interventions	

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

UNIVERSITY OF BIRMINGHAM