
Article:
Sitch, A.J., Dinnes, J., Hewison, J. et al. (2024) Optimising research investment by simulating and evaluating monitoring strategies to inform a trial: a simulation of liver fibrosis monitoring. BMC Medical Research Methodology, 24 (1). 315. ISSN: 1471-2288

https://doi.org/10.1186/s12874-024-02425-w

BMC Medical Research Methodology

RESEARCH

Open Access

Optimising research investment by simulating and evaluating monitoring strategies to inform a trial: a simulation of liver fibrosis monitoring

Alice J. Sitch[1,2*], Jacqueline Dinnes[1,2], Jenny Hewison[3], Walter Gregory[4], Julie Parkes[5] and Jonathan J. Deeks[1,2]

Abstract

Background The aim of the study was to investigate the development of evidence-based monitoring strategies in a population with progressive or recurrent disease. A simulation study of monitoring strategies using a new biomarker (ELF) for the detection of liver cirrhosis in people with known liver fibrosis was undertaken alongside a randomised controlled trial (ELUCIDATE).

Methods Existing data and expert opinion were used to estimate the progression of disease and the performance of repeat testing with ELF. Knowledge of the true disease status in addition to the observed test results for a cohort of simulated patients allowed various monitoring strategies to be implemented, evaluated and validated against trial data.

Results Several monitoring strategies ranging in complexity were successfully modelled and compared regarding the timing of detection of disease, the duration of monitoring, and the predictive value of a positive test result. The results of sensitivity analysis showed the importance of accurate data to inform the simulation. Results of the simulation were similar to those from the trial.

Conclusion Monitoring data can be simulated and strategies compared given adequate knowledge of disease progression and test performance. Such exercises should be carried out to ensure optimal strategies are evaluated in trials thus reducing research waste. Monitoring data can be generated and monitoring strategies can be assessed if data is available on the monitoring test performance and the test variability. This work highlights the data necessary and the general method for evaluating the performance of monitoring strategies, allowing appropriate strategies to be selected for evaluation. Modelling work should be conducted prior to full scale investigation of monitoring strategies, allowing optimal monitoring strategies to be assessed.

Keywords Monitoring, Surveillance, Tests, Biomarkers, Liver disease

*Correspondence:
Alice J. Sitch
a.j.sitch@bham.ac.uk
[1]National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, Birmingham, UK
[2]Test and Prediction Group, Department of Applied Health Sciences, Public Health Building, University of Birmingham, Birmingham, UK
[3]Leeds Institute of Health Sciences, University of Leeds, Leeds, UK
[4]Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK
[5]Primary Care & Population Sciences Faculty of Medicine, University of Southampton, Southampton, UK

Background

Tests are used in healthcare to monitor, and subsequently manage, a variety of chronic conditions. The focus of this research is monitoring of progressive or recurrent conditions, where the aim of monitoring is to identify early signs of recurrence or progression prompting a change in management, typically initiation of treatment or further testing.

Although patient monitoring is a fundamental function of healthcare, incurring considerable cost to health care providers, the underlying methodology of monitoring is under researched [1, 2] and there is an increased need for monitoring strategies to be developed incorporating known likely progression of disease and the performance of the monitoring test. In a methodological review of guidelines for prostate specific antigen (PSA) monitoring to identify recurrence of prostate cancer, Dinnes et al. [3] identified a lack of a systematic approach in developing monitoring strategies, with monitoring intervals based on standard follow-up schedules and limited evidence of consensus for the thresholds used to initiate treatment.

Monitoring strategies are complex interventions combining a test, a schedule, a decision rule and further testing or treatment. Fundamentally, the frequency of testing, and a 'monitoring rule' indicating the value (or values) that would trigger a change in patient management should be stipulated. Monitoring rules can be simple, using a single value as a threshold (a 'snap-shot rule'), or more complex, where series of test results are required to initiate a change in management (a 'track-shot rule') [4].

To inform the methods used to investigate potential monitoring strategies, we reviewed monitoring-related methodology [5]. Most monitoring studies evaluated were concerned with the frequency of testing [4, 6–13]. Fewer studies investigated test thresholds [4, 12] or monitoring test decision rules [4, 6, 7, 13]. Stevens et al. proposed a general statistical model for monitoring data [14], that combines the true disease state, which can be modelled but never observed, with estimates of measurement error. The model is informed by existing data, and evidence from the literature, allowing monitoring data to be simulated and the potential effect of monitoring strategies to be estimated, prior to full-scale investigation [15].

Liver fibrosis can progress to liver cirrhosis and lead to complications such as portal hypertension and hepatocellular cancer. We evaluated monitoring strategies using the Enhanced Liver Fibrosis (ELF) biomarker to detect liver cirrhosis in people with known liver fibrosis. The addition of ELF to standard monitoring in this group of patients has been evaluated in a prospective multicentre randomised trial (the Enhanced Liver fibrosis (ELF) test to Uncover Cirrhosis as an Indication for Diagnosis and Action for Treatable Events (ELUCIDATE) trial [16]), allowing the model to be validated. In the ELUCIDATE trial participants suspected to have chronic liver disease were recruited and randomised to standard care or standard care with monitoring using the ELF biomarker. All participants received standard outpatient assessment every six months, with those randomised to receive ELF monitoring also being tested using the ELF biomarker at each outpatient observation. An ELF test above the threshold and/or diagnosis of cirrhosis on standard assessments triggered further investigation. The study looked to evaluate liver related outcomes, such as variceal bleeding, ascites, encephalopathy, hepatocellular carcinoma, transplantation and death.

The aim was to identify the optimal monitoring strategy, by varying decision rules (thresholds for a positive test results), or monitoring intervals (time between monitoring tests), or by introducing targeted retesting (re-application of the test).

Methods

The aim of the study was to investigate the development of evidence-based monitoring strategies in a population with progressive or recurrent disease. We simulated disease progression for a group of patients, and the test results they would have received, given existing evidence and opinion from an expert in chronic liver epidemiology (JP). Using the simulated data we were able to compare the performance of different monitoring strategies and compare our simulated results to the ELUCIDATE trial.

Simulation model

The true disease level (U_{it}) was simulated using a random intercept (α_{it}, ELF value at entry to the trial) and a random slope (β_{it}, change in ELF over time) for each participant, $U_{it} = \alpha_{it} + \beta_{it}$. The true underlying values were converted to observed values by the addition of measurement error ($Y_{it} = U_{it} + \omega_{it}$). Full details of the simulation method have previously been reported [5].

Data sources

Estimates of fibrosis progression rate (from a study of the natural history of liver fibrosis) [17], fibrosis stage at trial entry, measurement error and ELF score link to fibrosis stage were obtained from published literature and other data sources (see Table 1).

Evaluation of monitoring strategies

A simple reference strategy with a simple threshold and 6-monthly test intervals, akin to that evaluated in the ELUCIDATE trial, was used. Alternative strategies were evaluated by:

Table 1 Data used in simulation model

Estimate required	Data	Estimates used in model
Fibrosis progression rate	Poynard et al.: estimate of median fibrosis progression (units per year) 0.133 (95% CI 0.125, 0.143) [17].	Estimate calculated from Poynard et al.: [17] Estimated fibrosis progression (units per year) $\sim N\ (0.13,\ 0.17^2)$. Estimate after adjustment (to be used in sensitivity analyses): estimate of fibrosis progression was increased to reflect expert opinion $\sim N\ (0.27,\ 0.17^2)$.
ELF stage at entry to trial	Cross sectional data set: estimated proportion of patients in each stage [18]. Stage 0- 0.25; stage 1- 0.35; stage 2- 0.13; stage 3- 0.15; stage 4- 0.12.	The cross-sectional data set was used: Stage 0- 0.25; stage 1- 0.35; stage 2- 0.13; stage 3- 0.15; stage 4- 0.12.
Measurement error	Longitudinal data set: estimate of the standard deviation of measurement error of 0.81. Siemens: estimate of the standard deviation of total measurement error of 0.11 [19]. ELUCIDATE registration and randomisation data: estimate of standard deviation of total measurement error 0.47.	Estimate of measurement error obtained from ELUCIDATE was used: $\omega_{it} \sim N\ \left(0,\ 0.47^2\right)$.
ELF link to fibrosis stage	Cross sectional data set: estimates of ELF mean (SD) at each fibrosis stage [18]. Stage 0- 8.82 (0.87); stage 1- 9.18 (0.96); stage 3- 9.55 (1.00); stage 4- 11.32 (1.47).	After adjustment: measurement error is accounted for to give the true unobserved ELF values and modified to represent values for each stage: stage 0 $\sim N\left(8.63,\ 0.73^2\right)$; stage 1 $\sim N\left(9.00,\ 0.84^2\right)$; stage 2 $\sim N\left(9.36,\ 0.89^2\right)$; stage 3 $\sim N\left(9.91,\ 1.22^2\right)$; stage 4 $\sim N\left(10.80,\ 1.39^2\right)$.

- Retesting participants with a test value within one unit of the threshold.
- Changing the frequency of monitoring to every 12 months.
- Using alternative decision rules (absolute and relative increases from randomisation and last recorded ELF measure and rule using predictions from a linear regression model fitted using all available observed data points).

Number of simulations

Simulations were based on a cohort of 20,000 patients. With 20,000 test results, if one of the performance measures gave an estimate of 15% a corresponding 95% confidence interval would range from 14.5 to 15.5%; for an estimate of 1.5% a 95% confidence interval would range from 1.3 to 1.7%.

Identifying positive and negative results

Test results were positive or negative based on the simulated observed data and the decision rule used. The test result was true or false depending on the simulated underlying disease state. As it may be beneficial to identify patients prior to progression to cirrhosis, participants were classed as 'diseased' three months prior to the development of cirrhosis. As a positive test result changes patient management, patients with positive results are not subsequently monitored.

Measuring the performance of monitoring strategies

Performance of monitoring strategies was measured based on previous criteria [8]: the number of tests per person for the duration of monitoring, positive predictive value (PPV), and percentage of patients with delayed diagnosis (over 12 months). For ease of comparison of strategies, monitoring thresholds associated with a clinically acceptable overall PPV of 25% (i.e. across the duration of monitoring) were selected, to allow paired comparisons.

Sensitivity analyses

Sensitivity analyses were conducted to estimate the effect of variations in measurement error, within-individual variation and disease progression rates on the reference strategy with all other aspects of the strategy kept constant (including the threshold value). Analyses were repeated allowing the threshold to vary in order to maintain the overall PPV at 25%.

Comparison to ELUCIDATE data

To assess the accuracy of the simulation model, the mean and standard deviation of randomisation ELF values were compared between the ELUCIDATE data and simulated data sets. Analysis of variance was used to assess between-individual and within-individual variability of ELF values recorded for participants in the trial and the simulated results. Multilevel models were fitted using the simulated observed values and the ELUCIDATE trial data (for participants with two or more ELF measures post registration). In the ELUCIDATE trial, an ELF score of 9.5 or above was considered positive, and no further ELF measurements were usually taken. The ELUCIDATE and simulated data sets were therefore modified so each patient with an ELF measure of 9.5 or above did not have

any subsequent measures. The number of participants with a diagnosis of cirrhosis in the simulated and trial data (monitoring and standard care arm) was compared using the trial strategy threshold of 9.5. In the standard care arm of the ELUCIDATE study a diagnosis of cirrhosis could be based on: clinical judgement and various tests (liver biopsy, ultrasound scan, liver CT scan, MRI scan, gastroscopy or FibroScan); for the monitoring arm, a diagnosis of cirrhosis was made if a participant's ELF test value was over the threshold value, in addition to the methods of diagnosis available in the standard care arm.

The number of observation points used from the simulation model was capped to give a similar mean number of observations per person to the value seen in the ELUCIDATE data. Allowing more observations per person would introduce bias as patients with slower progressing disease will have more ELF measurements prior to a test result of 9.5 or above [20].

Results

For the simulated cohort of 20,000 patients, 5,314 (26.6%) would develop cirrhosis during a five-year trial.

Reference monitoring strategy

The ELF threshold required to maintain the overall PPV at 25% was 10.715. Due to prevalent cases of liver cirrhosis at the beginning of the monitoring period, the sensitivity and PPV calculated for the reference strategy were highest at the initial observation point, as was the percentage of tests with a positive result. The percentage of false negative results generally increased at each observation point. Over the duration of the monitoring strategy 7.64 tests per person (152,724 tests in total) were performed and 6.10% of all patients had a delay to diagnosis of over 12 months, see Supplementary Table S1.

Comparison of strategies

Table 2 reports results for the different monitoring strategies and Fig. 1 shows that none of the evaluated strategies were clearly superior to the reference case. The reduced monitoring frequency strategy (C) decreased the number of tests required per person (by 3.30) with a 0.15% increase in the percentage of patients with delay to diagnosis of over 12 months. All of the track-shot decision rules led to an increase in the percentage of patients with a delay to diagnosis (of between 1.58% and 11.09%), with concordant increases in the mean number of tests performed per person (ranging from 0.14 (strategy D) to 1.18 (strategy G)). The retest strategy (B) increased the number of tests performed (increase of 3.30 tests per person), and increased the percentage of patients with delay to diagnosis (absolute increase of 0.40%). The linear regression strategy (H) had a small impact on both the number of tests (0.12 fewer tests per person) and on delay to diagnosis (0.47% lower).

When using the reduced monitoring frequency strategy the number of tests required decreased by 3.30 tests per person and the percentage of patients with delay to diagnosis of over 12 months increased by 0.15% points (absolute increase) compared with the reference strategy.

The linear regression strategy used fewer tests (0.12 tests per person) and had a lower percentage of patients with delay to diagnosis (0.47%) when compared to the reference strategy, see Table 2; Fig. 1.

Sensitivity analyses

Table 3 demonstrates the effect on the reference strategy (strategy A) of increasing or decreasing various parameter estimates.

Estimates of reduced test variability (decreased measurement error and between-individual variability) improved PPV (increases of 4.6% and 8.6%) and increased the number of tests required (0.73 and 0.91 tests per person) with decreased measurement error also increasing the percentage of patients with delay to diagnosis of more than 12 months (increase of 1.30%). Both increased and decreased between-individual variability reduced the percentage of patients with delay to diagnosis (0.72% and 2.12%). An increased rate of fibrosis progression led to increased PPV (4.2%) and percentage of patients with delay to diagnosis (1.52%) but decreased the number of tests required (0.64 tests per person).

The largest difference in PPV resulted from increased between-individual variability (8.8%); the largest difference in number of tests required from increased measurement error (1.84 tests per person); and the largest difference in the percentage of patients with delay to diagnosis of over 12 months from decreased between-individual variability (decrease of 2.12%).

Increasing the rate of fibrosis progression, produced similar results to the unadjusted estimate (see Supplementary materials Table S2 and Figure S1).

Comparison to ELUCIDATE data

The ELUCIDATE data contained 705 observations taken from 420 participants randomised to the ELF monitoring arm. After removing measurements following an ELF value of 9.5 or above for each individual, the simulated data set contained 66,320 observations for 20,000 participants. Analysis of the ELF value at the point of randomisation for each of the data sets showed similar results (see Table 4) with the mean value slightly lower for the ELUCIDATE data. The between-individual standard deviation was higher for the ELUCIDATE data than the simulated data (0.93 for the ELUCIDATE data compared with 0.76 for the simulated data). The within-individual

Table 2 Results of strategies A-H

Strategy	Monitoring strategy components					PPV	Tests*			Delay to diagnosis‡			Test performance			
	Decision rule	Threshold value	Interval (months)	Retest	Initial threshold	%	N	N pp†	Median (Q1, Q3)	N	% of all§	% of stage 4‖	TP pp¶	FP pp#	Positive n (%)	Sensitivity (%)
A	Simple threshold	10.715	6	FALSE	-	25	152,724	7.64	11 (3, 11)	1220	6.10	22.96	0.12	0.37	9883 (6.47)	22.13
B	Simple threshold	10.580	6	TRUE	-	25	218,974¥	10.95	12 (6, 15)	1300	6.50	24.46	0.12	0.35	9406 (6.11)	21.15
C	Simple threshold	10.550	12	FALSE	-	25	86,787	4.34	6 (2, 6)	1249	6.25	23.50	0.13	0.38	10,053 (11.58)	35.34
D	Absolute increase from initial value	1.295	6	FALSE	10.715	25	155,648	7.78	11 (4, 11)	1536	7.68	28.90	0.13	0.40	10,598 (6.81)	19.85
E	Absolute increase from last value	1.460	6	FALSE	10.715	25	172,363	8.62	11 (7, 11)	2085	10.42	39.24	0.08	0.24	6305 (3.66)	8.45
F	Relative increase from initial value	1.144	6	FALSE	10.715	25	156,460	7.82	11 (4, 11)	1630	8.15	30.67	0.13	0.38	10,266 (6.56)	18.12
G	Relative increase from last value	1.1795	6	FALSE	10.715	25	176,385	8.82	11 (9, 11)	2217	11.09	41.72	0.07	0.20	5338 (3.03)	6.68
H	Linear regression	10.495	6	FALSE	10.715	25	150,478	7.52	11 (2, 11)	1126	5.63	21.19	0.12	0.35	9342 (6.21)	21.58

*Tests over the duration of monitoring

† Mean number of tests per person over the duration of monitoring

‡ Patients with delayed diagnosis (delay from onset of disease to diagnosis of over 12 months)

§ % of all patients with delay to diagnosis

‖ % of patients that would reach cirrhosis within the trial period with delay to diagnosis

¶ TP pp is the mean number of true positive results per person over the duration of monitoring

FP pp is the mean number of false positive results per person over the duration of monitoring

¥ 218,974 tests were carried out to generate 153,971 results due to retests being used

A is the simple threshold strategy; B is the retest strategy; C is the decreased monitoring frequency strategy; D is the absolute increase from initial value strategy; E is the absolute increase from last value strategy; F is the relative increase from initial value strategy; G is the relative increase from last value strategy; H is the linear regression strategy

standard deviation was similar for the ELUCIDATE data and simulated data (0.53 and 0.51 respectively).

The ELUCIDATE data modelled consisted of 429 observations from 153 participants; each had a minimum of 2 and a maximum of 6 ELF observations and the average number of observations per person was 2.8. The model fitted to simulated data used 26,429 observation points for 9,608 simulated participants and the mean number of observations was 2.8.

Modelling of the ELUCIDATE data estimated an increase in ELF per year of 0.31 (95% CI (0.22, 0.39); p-value < 0.001), see Table 4. Modelling the simulated data estimated an increase in ELF per year to be comparable at 0.24 (95% CI (0.23, 0.26); p-value < 0.001); for the simulated data with adjusted fibrosis progression the increase was 0.28 (95% CI (0.27, 0.30); p-value < 0.001.

Comparison of the cirrhosis outcomes for the simulated data and ELUCIDATE data (monitoring arm) showed a comparable percentage of positive results; there was a lower percentage of participants with a cirrhosis diagnosis in the standard care arm. However, the percentage of diagnoses of cirrhosis at the first testing time point was greater for the simulated data than the trial, see Table 5.

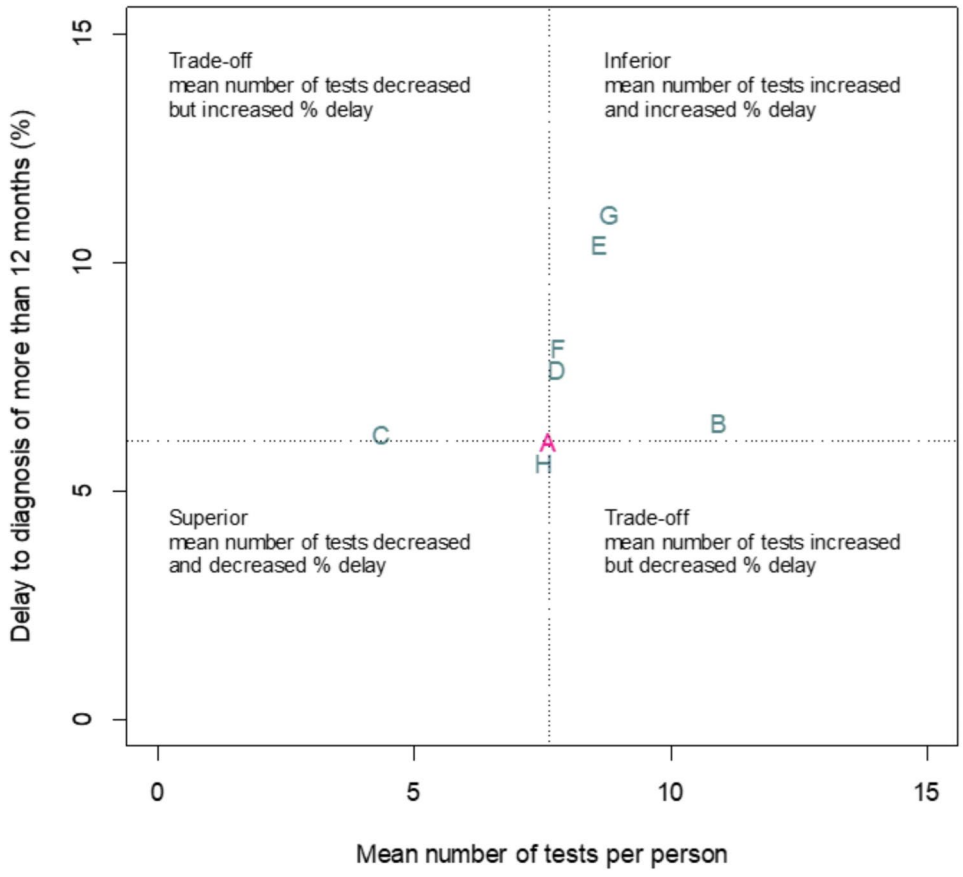

Fig. 1 Performance of various monitoring strategies on simulated monitoring data with PPV of 25%. A is the simple threshold strategy; B is the retest strategy; C is the decreased monitoring frequency strategy; D is the absolute increase from initial value strategy; E is the absolute increase from last value strategy; F is the relative increase from initial value strategy; G is the relative increase from last value strategy; H is the linear regression strategy

Discussion

Analysis of the simulated and trial data showed similar results. The monitoring arm of the trial detected cirrhosis in 64.2% of patients compared with just 4.5% in the standard care arm and the model predicted this would be 70.7%. The difference in detection between the trial arms suggests the strategy has a high false positive rate. With sufficient time this modelling exercise could have been used to modify the strategy.

When a monitoring strategy is introduced, cases will be identified from a prevalent population where a large proportion of patients will have high ELF values, hence the difference in results at the initial monitoring time point compared with others. The increasing percentage of false negative results at subsequent monitoring points suggests the simple threshold should be reduced to account for the patients that have false negative results using the original threshold.

The simple threshold strategy outperformed all strategies that incorporated participant's previous measurements; the strategies using absolute and relative increase from last recorded value decision rules performing particularly poorly. Strategies using absolute and relative changes from the initial recorded value consider differences in ELF across the entire monitoring period, and so are better at detecting true change over measurement error. This result is related to the index of individuality (II), or the ratio of within-individual and between-individual variation; the higher the II the higher the within-individual variation is in comparison to between individual variation. Tests with higher II values perform better with constant thresholds as an individual can have results spanning a wide range of the possible results for a group of people. Tests with decision rules that compare a result to a previous result are more beneficial if the II value is lower, as an individual will have tests results spanning only part of the possible range of results [21].

With the re-testing strategy, measurement error in both the initial and retest result leads to some people with a positive result on their initial test (as with the reference strategy) having a negative result when the mean of the initial and retest measurements is used. An increase in negative results from a retest strategy will have a small effect on the percentage of participants with delay to diagnosis of over 12 months.

Table 3 Results of using the reference strategy when changing estimates required for data simulation keeping the threshold consistent and changing the threshold to fix PPV at 25%

Change in data simulation	Threshold	PPV (%)	Number of tests per person*	Delay† (%)	Develop cirrhosis‡*n* (%)
None	10.715	25.0	7.64	6.10	5314 (26.6)
Decreased§ measurement error	10.715	29.6	8.37	7.40	5248 (26.2)
		(+4.6)	*(+0.73)*	*(+1.30)*	*(-66 (0.33))*
	10.450	25.0	7.70	5.73	
			(+0.06)	*(-0.37)*	
Increasedǁ measurement error	10.715	17.6	5.84	3.85	5421 (27.1)
		(-7.4)	*(-1.80)*	*(-2.25)*	*(+107 (0.54))*
	11.365	25.0	7.65	7.05	
			(+0.01)	*(+0.95)*	
Decrease§ between-individual variability	10.715	33.6	8.55	3.98	5139 (25.7)
		(+8.6)	*(+0.91)*	*(-2.12)*	*(-175 (0.88))*
	10.463	25.0	7.84	2.28	
			(+0.20)	*(-3.82)*	
Increasedǁ between-individual variability	10.715	16.5	5.80	5.38	5272 (26.4)
		(-8.8)	*(-1.84)*	*(-0.72)*	*(-42 (0.21))*
	11.905	25.0	8.26	9.95	
			(+0.62)	*(+3.85)*	
Decreased§ fibrosis progression rate	10.715	22.7	7.90	4.95	4440 (22.2)
		(-2.3)	*(+0.26)*	*(-1.15)*	*(-874 (4.37))*
	10.860	25.0	8.29	5.68	
			(+0.65)	*(-0.42)*	
Increasedǁ fibrosis progression rate	10.715	29.2	7.00	7.62	7689 (38.4)
		(+4.2)	*(-0.64)*	*(+1.52)*	*(+2375 (11.88))*
	10.460	25.0	6.21	5.63	
			(-1.43)	*(-0.47)*	

*Number of tests per person over the duration of monitoring; †% of all patients with delayed diagnosis (delay from onset of disease to diagnosis of over 12 months);‡Patients that would go on to develop cirrhosis in the monitoring duration if no intervention were received; §Decrease is halving the estimate used in the original simulation; ǁIncrease is doubling the estimate used in the original simulation

Table 4 Results of analysis of randomisation ELF, analysis of variance for ELF measurements at all time points and multilevel modelling

	ELUCIDATE data	Simulated data	Simulated data with adjusted fibrosis progression
Randomisation point ELF			
ELF mean (SD)	9.57 (1.21)	9.71 (1.15)	9.83 (1.20)
Analysis of variance			
Between-individual SD	0.93	0.76	0.82
Within-individual SD	0.53	0.51	0.52
Multilevel modelling			
Years	0.31 (0.22, 0.39)	0.24 (0.23, 0.26)	0.28 (0.27, 0.30)
Constant	8.73 (8.63, 8.82)	8.84 (8.83, 8.85)	8.86 (8.84, 8.87)
Between-individual SD	0.43 (0.36, 0.51)	0.42 (0.41, 0.43)	0.42 (0.41, 0.43)
Within-individual SD	0.48 (0.44, 0.52)	0.47 (0.46, 0.47)	0.46 (0.46, 0.47)

Table 5 Comparison of outcomes for trial and simulated data

Outcome	RCT ELF arm	RCT Standard Care arm	Model ELF
Diagnosis of cirrhosis during trial	281/438 (64.2%)	20/440 (4.5%)	14,132/20,000 (70.7%)
Diagnosis of cirrhosis after 1st measurement	84/438 (29.9%)	20/440 (4.5%)	3740/20,000 (18.7%)

Reducing the test frequency led to a large decrease in the number of tests per person and a small increase in the percentage of people with delay to diagnosis. In some circumstances, for the substantial decrease in the number of tests required and therefore the resource used, the increased harm to patients may be acceptable.

The linear regression strategy showed a reduction in both the number of tests required and the percentage of patients with delay to diagnosis. The linear regression method utilised all available data and some allowance was made for the fluctuation in results due to measurement error. This modest improvement in monitoring strategy performance may not merit the extra complexity involved.

Increased measurement error results in more false positive results. Between-individual variability will affect

the underlying ELF values possible at each fibrosis stage. Providing ELF is truly related to fibrosis stage, smaller between-individual variability means that ELF is more likely to correctly identify fibrosis stage with fewer false positive results. With fewer false positives, PPV will increase and the number of tests required will increase as the number of patients correctly staying in the monitoring programme will increase.

With an increased fibrosis progression rate more patients will develop compensated cirrhosis, higher prevalence increases PPV, and patients will have positive results earlier in the strategy, requiring fewer tests to be performed. If patients have increased fibrosis progression rate there is increased potential for patients to have undetected disease for over 12 months.

Limitations

Estimates from data sources were used to inform the simulation model. The suitability of data was discussed with a clinical expert and, where necessary, estimates were adjusted and sensitivity analyses performed; however, the quality and suitability of data used is a limitation. We were limited to being advised by a single clinical expert, involving additional experts may have changed the estimates used.

The ELUCIDATE trial data contained repeated observations from 153 participants with many participants having only two observations; more observations per person would allow better estimation of the error terms and changes over time. The ability of the data to estimate the true progression of ELF is limited as those with values over 9.5 cease to have measurements recorded meaning further progression cannot be assessed. Patients with lower measures continued monitoring and contributed more data to the model; however they were potentially in a better health state.

Modelling a monitoring strategy prior to starting a prospective study is time consuming, requires expert opinion, and estimates and data need to be available. A simplified version of this approach may be needed for some clinical scenarios.

Further work

A greater variety of strategies could be evaluated with multiple components assessed simultaneously. More complex decision rules and frequencies could be explored, for example varying thresholds by monitoring time points or non-constant testing frequencies.

The model could also show lifetime progression for a time-matched cohort of patients with fibrosis; this would indicate strategy performance in practice rather than in a trial setting.

Well-designed studies of biological variability would mean accurate estimates of variability would be available enabling accurate monitoring data to be generated and analysed.

Further comparison of the ELUCIDATE data and the simulated data can be performed when additional outcomes are collected from the trial participants.

Further work could also investigate the value of this approach if there is limited data available to support the simulation. This work could also identify ranges of values that could be used in the absence of appropriate data and/or estimates.

Conclusions

Simulation can be used to evaluate candidate monitoring strategies enabling appropriate strategies to be selected for full scale evaluation.

To generate monitoring data there has to be available evidence on the natural history of the disease and the performance of the monitoring test (measurement error and test accuracy)—this evidence can be from existing data sets, reviewing the literature or potentially expert opinion. If the data informing the simulation model is inaccurate the results obtained from evaluation of strategies will not reflect the truth. Inaccurate estimates will affect results in a complex way. The results of sensitivity analyses highlighted the importance of accurate estimates of test performance and progression.

Comparison of the trial data and the simulated data provided similar results. Bias in monitoring data, particularly concerning the number of recorded results, should be considered when analysing as those contributing more monitoring data points are generally different to those contributing few.

Abbreviations

ELF Enhanced Liver Fibrosis biomarker
RCT Randomised Controlled Trial
PPV Positive Predictive Value

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-024-02425-w.

Supplementary Material 1

Acknowledgements

The authors are grateful for the input from members of the methodology work-stream for the Biomarker Pipeline project: William Rosenberg, Douglas G Altman, Christopher McCabe, Paul Baxter and Roberta Longo.

Author contributions

AS performed the analysis and wrote the manuscript. AS, JD, JH, WG, JP and JJD contributed to the design of the study. All authors read and approved the final manuscript.

Funding

This publication presents independent research commissioned by the National Institute for Health and Care Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10101). This paper presents independent research supported by the NIHR Birmingham Biomedical

Research Centre (AJS, JD and JJD). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Data availability
This study primarily uses simulated data. Code to generate data is available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
Julie Parkes has been paid for providing lectures by Siemens on the topic of the ELF marker. Julie Parkes reported receiving support from the Speakers' Bureau for Siemens Healthcare Diagnostics, outside the submitted work, and is married to William Rosenberg. William Rosenberg was an inventor of the enhanced liver fibrosis (ELF) test when he was an employee of the University of Southampton. His rights were transferred to Siemens Healthcare Diagnostics Ltd by the University of Southampton. He does not receive any payment in relation to sales of the test by the manufacturer Siemens Healthcare Diagnostics. He has received grant support and speaker fees from Siemens Healthcare Diagnostics and is a director of iQur Ltd (Southampton, UK), a company that provides ELF testing. In the context of this NIHR-funded study, all ELF testing provided by iQur Ltd was performed on a not-for-profit, cost-recovery basis. During the period of the study Jonathan J Deeks was a panel member of the HTA Commissioning Board. Jenny Hewison was a panel member of the National Institute for Health Research (NIHR) Clinical Trials Unit Standing Advisory Committee and Subpanel Chair, NIHR Programme Grants for Applied Research. Walter Gregory was the ELUCIDATE Trial Director and Principal Statistician. Alice Sitch and Jacqueline Dinnes report no competing interests.

Received: 27 March 2024 / Accepted: 26 November 2024
Published online: 20 December 2024

References

1. Glasziou PP, Aronson JK. In: Glasziou PP, Irwig L, Aronson JK, editors. An introduction to monitoring therapeutic interventions in clinical practice. Oxford: Blackwell Publishing; 2008. pp. 3–14.
2. Glasziou P. How much monitoring? British Journal of General Practice 2007(May):350 – 51. https://doi.org/10.1016/j.jacc.2006.10.081.6
3. Dinnes J, Hewison J, Altman DG, et al. The basis for monitoring strategies in clinical guidelines: a case study of prostate-specific antigen for monitoring in prostate cancer. Can Med Assoc J. 2012;184(2):169–77. https://doi.org/10.1503/cmaj.110600.
4. Buclin T, Telenti A, Perera R, et al. Development and validation of decision rules to guide frequency of monitoring CD4 cell count in HIV-1 infection before starting antiretroviral therapy. PLoS ONE. 2011;6(4):e18578–78. https://doi.org/10.1371/journal.pone.0018578.
5. Selby PJ, Banks RE, Gregory W, et al. Methods for the evaluation of biomarkers in patients with kidney and liver diseases: multicentre research programme including ELUCIDATE RCT. Programme Grants Appl Res. 2018. https://doi.org/10.3310/pgfar06030.
6. Bellera CA, Hanley JA, Joseph L, et al. Detecting trends in noisy data series: application to biomarker series. Am J Epidemiol. 2008;167(9):1130–9. https://doi.org/10.1093/aje/kwn003.
7. Bellera C, Hanley J, Joseph L, et al. A statistical evaluation of rules for biochemical failure after radiotherapy in men treated for prostate cancer. Int J Radiat Oncol Biol Phys. 2009;75(5):1357–63. https://doi.org/10.1016/j.ijrobp.2009.01.013.
8. Li H, Gatsonis C. Dynamic optimal strategy for monitoring disease recurrence. Sci China Math. 2012;55(8):1565–82. https://doi.org/10.1007/s11425-012-4475-y.
9. Oke JL, Stevens RJ, Gaitskell K, et al. Establishing an evidence base for frequency of monitoring glycated haemoglobin levels in patients with type 2 diabetes: projections of effectiveness from a regression model. Diabet Med. 2012;29(2):266–71. https://doi.org/10.1111/j.1464-5491.2011.03412.x.
10. Sölétormos G, Schiøler V. Description of a computer program to assess cancer antigen 15.3, carcinoembryonic antigen, and tissue polypeptide antigen information during monitoring of metastatic breast cancer. Clin Chem. 2000;46(8 Pt 1):1106–13.
11. Takahashi O, Glasziou PP, Perera R et al. Lipid re-screening: what is the best measure and interval? Heart 2010;96(6):448 – 52. https://doi.org/10.1136/hrt.2009.172619
12. Takahashi O, Glasziou PP, Perera R, et al. Blood pressure re-screening for healthy adults: what is the best measure and interval? J Hum Hypertens. 2012;26(9):540–6. https://doi.org/10.1038/jhh.2011.72.
13. Sölétormos G, Hyltoft Petersen P, Dombernowsky P. Progression criteria for cancer antigen 15.3 and carcinoembryonic antigen in metastatic breast cancer compared by computer simulation of marker data. Clin Chem. 2000;46(7):939–49.
14. Stevens RJ, Oke J, Perera R. Statistical models for the control phase of clinical monitoring. Stat Methods Med Res. 2010;19(4):394–414. https://doi.org/10.1177/0962280209359886.
15. MRC. A framework for development and evaluation of RCTs for complex interventiond to improve health, 2000.
16. ISRCTN. ISRCTN Register.
17. Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. Lancet. 1997;349(9055):825–32. https://doi.org/10.1016/S0140-6736(96)07642-8.
18. Rosenberg WMC, Voelker M, Thiel R, et al. Serum markers detect the presence of liver fibrosis: a cohort study. Gastroenterology. 2004;127(6):1704–13. https://doi.org/10.1053/j.gastro.2004.08.052.
19. Siemens. ELF Test.
20. Bellera C, Hanley J, Joseph L, et al. Hierarchical changepoint models for biochemical markers illustrated by tracking postradiotherapy prostate-specific antigen series in men with prostate cancer. Ann Epidemiol. 2008;18(4):270–82. https://doi.org/10.1016/j.annepidem.2007.10.006.
21. Fraser CG. Biological Variation: from principles to practice. AACC; 2001.

Publisher's note