Research

Nick Sheron, Mike Moore, Stacey Ansett, Camille Parsons and Adrian Bateman

Developing a 'traffic light' test with potential for rational early diagnosis of liver fibrosis and cirrhosis in the community

Abstract

Background
Liver disease develops silently and presents late, with often fatal complications.

Aim
To develop a 'traffic light' test for liver disease suitable for community use that could enhance assessment of liver risk and allow rational referral of more severe disease to specialist care.

Design and setting
Two cohorts from Southampton University Hospital Trust Liver Unit: model development and a validation cohort to evaluate prognosis.

Method
A total of 1038 consecutive liver patients (inpatient and outpatient) (development $n = 397$, validation $n = 641$) for whom the relevant blood tests had been performed, were followed for a mean of 46 months (range 13–89 months). Blood tests for: hyaluronic acid (HA), procollagen-3 N-terminal peptide (P3NP), and platelet count were combined in a diagnostic algorithm to stage liver disease.

Results
A simple clinical rule combined: HA, P3NP, and platelet count into a 'traffic light' algorithm, grading the results red — high risk, amber — intermediate risk, and green — low risk. In the validation cohort, no green subjects died or developed varices or ascites ($n = 202$); in the amber group, 9/267 (3.3%) died, 0/267 developed varices, and 2/267 (0.7%) developed ascites; in the red group, 24/172 died (14%), 24/172 (14%) developed varices, and 20/172 developed (11.6%) ascites. Survival was reduced in red ($P<0.001$) and amber ($P<0.012$) groups compared with green.

Conclusion
A simple blood test triages liver disease into three prognostic groups; used in the community, it could enhance the management of risk factors in primary care and rationalise secondary care referrals, including the many patients with fatty liver and relatively minor elevations in alanine transaminase.

Keywords
cirrhosis; diagnosis; early diagnosis; liver diseases; liver fibrosis; primary care.

INTRODUCTION

At a time when mortality for many diseases is falling, deaths from liver disease have increased fourfold since 1970 and doubled since 1993.[1] The majority of these deaths have been from alcohol-related disease as a result of increasing alcohol intake, but the increasing incidence of obesity, metabolic syndrome, and diabetes has increased the prevalence of other forms of fatty liver disease, and chronic viral hepatitis is also increasing. Hospital episodes for liver disease have increased by 8.3% each year from 1998 to 2008; in 2005, there were 43 694 episodes coded with liver disease as the primary diagnosis, and 6798 deaths — a case fatality rate of 15.5% per episode.[2] Unfortunately, liver disease develops silently and frequently presents with the late complications of cirrhosis: variceal haemorrhage, decompensated cirrhosis, or acute or chronic liver failure — all have a high mortality. The hospital mortality of cirrhosis has not changed for 30 years,[3] suggesting a significant rethink is desperately needed.

If liver deaths are to be reduced, then there is a need to address the major risk factors for liver disease: alcohol, obesity, and viral hepatitis,[2] but it is also necessary to detect liver disease before the development of cirrhosis, when lifestyle changes or specific treatment can prevent the progression of disease. Historically, the diagnosis of liver disease is reliant on a referral to specialist services, very often based on an elevated level of alanine transaminase (ALT). ALT and gamma glutamyltransferase are useful tests for inflammatory liver disease, and are elevated in around half of simple fatty livers but it is important to realise that they are of little help in predicting which patients have liver fibrosis or cirrhosis.[4]

This study used fibrosis markers (procollagen-3 N-terminal peptide [P3NP] and hyaluronic acid), along with routine liver function tests in a clinical algorithm in the liver clinic (Southampton Traffic Light — STL). The algorithm was developed for the ongoing Alcohol and Liver Disease Detection Study (ALDDeS), in which 10 000 subjects in 10 general practices were screened for alcohol use, and hazardous and harmful drinkers offered the 'traffic light' screening test.[5] Alongside this project, a number of competing and perhaps equally effective non-invasive diagnostic methodologies have been published.[6–10] This traffic light system is probably no more accurate than the better ones, but it was developed specifically for use in primary care, with the aim of being intuitively easy for patients to understand.

This study presents the STL test; examines whether a modified version (mSTL) created using logistic regression analysis would be an improvement; and

N Sheron, MD, FRCP; **S Ansett**, BM, BMedSci, Clinical and Experimental Sciences Academic Unit, Faculty of Medicine, University of Southampton. **A Bateman,** BSc, MD, FRCPath, University Hospital Southampton NHS Foundation Trust. **C Parsons,** BSc, MSc, Southampton Centre for Biomedical Research, University Hospital Southampton NHS Foundation Trust, Southampton. **M Moore**, MSc, MRCP, FRCGP, Primary Care Research Network South, Aldermoor Health Centre, Southampton.
Address for correspondence
Nick Sheron, University of Southampton Faculty of Medicine, Clinical and Experimental Sciences, University Hospital Southampton NHS Foundation Trust, Tremona Road, Southampton, SO16 6YD.
E-mail: Nick.Sheron@soton.ac.uk

How this fits in

Liver disease mortality in the UK has doubled over the last 15 years, and survival rates of liver admissions have not improved. This study's experience in Southampton is that more than 90% of first liver admissions have unsuspected liver disease, whereas clinic referrals are dominated by patients with minor elevations of liver enzymes who could be dealt with in primary care. The current diagnosis and management of liver disease in the UK leaves much to be desired. The study has developed a simple blood test which triages patients into red, amber, or green categories according to the degree of liver fibrosis. These categories predict survival and the development of liver complications and have the potential to rationalise the diagnosis and management of liver disease in primary care. Ongoing studies are addressing how best this new technology can be best used.

analyses how well both models (STL and mSTL) predicted prognosis in a separate validation cohort. The study aimed to examine the following clinical questions:

- Can a patient be informed that they have an increased likelihood of progressive fibrosis, and thus incentivised to modify lifestyle risk factors or undergo specific treatment?
- Can the presence of progressive liver fibrosis be ruled out for a period of time, even in the presence of an elevated ALT, thus avoiding unnecessary referral to secondary-care liver services?

METHOD

The study population comprised 1038 consecutive patients with suspected liver disease, in whom the routine full blood count, liver function tests, and analysis of the serum fibrosis markers HA or collagen P3NP had been performed as part of routine diagnosis in the NHS laboratory of Southampton University Hospital Trust (SUHT), between July 2003 and November 2009. P3NP (Orion Diagnostica, Espoo, Finland) and HA (Corgenix Inc, Broomfield, US) were assayed using commercial immunoassays. Results are given throughout the text as follows: HA and P3NP (µg/l), platelets (109/l), ALT (iu/l).

Ethical permission to prospectively study these patients was obtained in 2003; as no research procedures were involved, informed consent was not required. Subjects were subsequently identified through the SUHT biochemistry database; all subjects with suspected liver disease and with the relevant clinical and laboratory data were included, with no exclusions. Medical, endoscopy, radiology, and pathology records were analysed to provide clinical data, and the subjects are described in two groups; demographic and clinical data are given alongside outcome data in Table 1.

The Southampton Traffic Light

The STL is a clinically derived rule of thumb, based on the authors' experience using fibrosis markers in the liver clinic. To aid interpretation for the ALDDeS study, results were categorised into three grades: green, amber, and red, as follows:

- HA > 30 µg/l or P3NP >5.5 µg/l – score +1
- HA >75 µg/l – score +2
- platelet count <150 ×109/l – score +1
- total score: 0 = green, 1 = amber, 2 or more = red.

At the time the original algorithm was designed, the researchers did not have the benefit of the huge dataset reported here, and had no validation cohort. Earlier versions included the international normalised ratio (INR) and albumin but these were dropped, as interim analyses showed them to be of no discriminatory value for liver fibrosis. For the platelet cut-off, the normal range in Southampton was used, P3NP and HA cut-offs were informed by interim area under the receiver operator curve (AUROC) analyses, but the algorithm was a clinical interpretation as opposed to a scientific analysis.

It was anticipated that binary logistic regression analysis of the development cohort would produce a more accurate algorithm than the STL. In actual fact, although the mSTL proved slightly more accurate in terms of AUROC analysis, the difference was clinically insignificant and the authors have continued to use the STL, which is easier to calculate. In the analysis of the validation cohort, the results of the two algorithms, STL (clinically derived) and mSTL (logistic regression model), are presented side by side.

Development of a logistic regression model

In accordance with the standards for reporting of diagnostic accuracy (STARD),[11]

Table 1. Demographic details, diagnosis, and outcome in the derivation and validation cohorts

	Model derivation, *n* = 397	Validation, *n* = 641	Total
Sex, male:female	236:161	378:263	
Mean age, years (SD)	52 (13.4)	49 (14.3)	
Diagnosis			
Viral hepatitis, *n* (%)	43 (10.8)	122 (19)	165 (15.9)
Autoimmune/metabolic, *n* (%)	44 (11.1)	64 10)	108 (10.4)
Biliary disease, *n* (%)	3 (0.8)	18 (2.8)	21 (2)
Alcohol-related liver disease, *n* (%)	121 (30.5)	198 (30.9)	319 (30.7)
Non-alcohol fatty disease, *n* (%)	134 (33.8)	188 (29.3)	322 (31)
Miscellaneous, *n* (%)	18 (4.5)	28 (4.4)	46 (4.4)
Primary liver cancer, *n* (%)	0	3 (0.5)	3 (0.3)
Cryptogenic cirrhosis, *n* (%)	19 (4.8)	6 (0.9)	25 (2.4)
Drug related, *n* (%)	15 (3.8)	14 (2.2)	29 (2.8)
Total, *n* (%)	397	641	1038
Stage of liver disease			
No fibrosis, *n* (%)	72	Unknown	
Progressive fibrosis, *n* (%)	170	Unknown	
Cirrhosis, *n* (%)	155	Unknown	
Validation: mean (range), months	54 (15–89)	41 (13–89)	
Survival and complications			
Dead, *n* (%)	61 (15.4)	33 (5.1)	
Varices, *n* (%)	84 (21.2)	24 (3.7)	
Ascites, *n* (%)	90 (22.7)	22 (3.4)	
Fibrosis markers and ALT			
Mean HA, µg/l (SD)	408 (1367)	149 (1272)	
Mean P3NP, µg/l (SD)	9.7 (12.3)	6.0 (6.9)	
Mean ALT, iu/l (SD)	76.5 (106)	72.3 (189)	

ALT = alanine transaminase. HA = hyaluronic acid. P3NP = procollagen-3 N-terminal peptide. SD = standard deviation.

the cohort was split into a model development cohort (397) and a validation cohort (641). The model derivation cohort comprised 397 subjects with objective evidence of the degree of liver fibrosis on liver biopsy within 2 years of the fibrosis markers (*n* = 334), or cirrhosis (*n* = 63), as evidenced by clinical pathological features together with evidence of portal hypertension, ascites, or liver morphology on imaging prior to the fibrosis markers. Biopsies were graded according to severity of fibrosis: no fibrosis (F0), fibrosis without cirrhosis (progressive fibrosis F1–3), or cirrhosis (F4), and the earliest stage of fibrosis, F1, was chosen as the cut-off for the analysis, because the study aim was to investigate the accuracy of the STL in diagnosing early disease.

This model was developed and internally validated in the development cohort (*n* = 397), using a logistic regression analysis and a 0.632 bootstrap sampling process (Table 2).[12,13] A sample of 397 subjects was taken, with replacement, from the 397 with a biopsy, HA, P3NP, and platelet values. As this sample was taken with replacement, it was possible for some subjects to be sampled multiple times and for others not to be sampled at all. A logistic regression model was then fitted to the sampled subjects, forcing HA, P3NP, and platelets into the model. This model was then applied to the subjects that were not sampled, with the AUROC of this validation model saved. This process was repeated a large number of times (*n* = 10 000) and the 2.5th, 50th, and 97.5th percentiles used as the validation AUROC and accompanying 95% probability interval. Findings were then validated against the key clinical outcomes, in a prognostic model, in a separate cohort of subjects who had undergone the test for routine diagnostic purposes and in whom the stage of fibrosis was unknown at the time of the test.

In the development dataset, there were 379/397 subjects with the full panel of possible variables being considered. A backwards-stepwise modelling approach was used to derive the model in the full set of subjects with a biopsy. Initial variables were: hyaluronic acid (HA), P3NP, albumin, international prothrombin ratio (INR), platelet count (PLT), bilirubin (Bili), alkaline phosphatase (ALP), and alanine transaminase (ALT).

The intermediate logistic regression model (mSTL) was as follows:

predicted probability (p) = exp(HA * 0.015 + P3NP * 0.447 – PLT * 0.005 + –0.611) / (1 + exp(HA * 0.015 + P3NP * 0.447 – PLT * 0.005 + –0.611))

Green/amber and amber red cut-off values were obtained from AUROC analysis. The red/amber cut off of 0.921 corresponded to 95% specificity (52% sensitivity) for any degree of fibrosis, and the green/amber cut-off of 0.616 corresponded to 90% sensitivity (54% specificity).

The equation above can also be written as:

log(P/1 – P) = HA * 0.015 + P3NP * 0.447 – PLT * 0.005 – 0.61.

Validation cohort

Clinical outcomes were analysed in a separate cohort of 641 subjects out of 1038 in total, in whom objective evidence of the stage of liver fibrosis was not available when the risk algorithm was performed. No data from this cohort were used in the development cohort. In the validation cohort, 53/641 were missing the full

Table 2. Initial and final covariates in the binary logistic regression analysis; the dependent variable was the presence of liver fibrosis (F0 versus F1–4)

	P-value	Odds ratio	95% CI
Step 1a			
ha1	0.018	1.012	1.002 to 1.023
P3NP1	0.000	1.637	1.291 to 2.076
ALB_first	0.731	0.983	0.891 to 1.085
INR_first	0.925	0.916	0.147 to 5.715
PLT_first	0.009	0.994	0.989 to0.998
TB_first	0.522	1.015	0.969 to 1.064
ALP_first	0.229	1.001	0.999 to 1.004
ALT_first	0.302	0.996	0.990 to 1.003
Constant	0.965	1.111	
Step 6a			
ha1	0.009	1.013	1.003 to 1.023
P3NP1	0.000	1.572	1.263 to 1.957
PLT_first	0.009	0.994	0.990 to 0.999
Constant	0.611	0.667	

ALB = Albumin. ALP = alkaline phosphatase. ALT = alanine transaminase. ha = hyaluronic acid. P3NP = procollagen-3 N-terminal peptide. INR = international normalised ratio. PLT = platelet count. TB = Total bilirubin.

Table 3. Results of the STL and mSTL traffic light algorithms in the model development cohort according to the level of fibrosis/cirrhosis

	Green	Amber	Red	Total
mSTL				
No fibrosis	44	25	2	71
Mild fibrosis	17	49	17	83
Moderate fibrosis	6	23	10	39
Severe fibrosis	6	14	19	39
Cirrhosis	1	19	127	147
All fibrosis + cirrhosis	30	105	173	308
Totals	74	130	175	379
STL				
No fibrosis	32	32	8	72
Mild fibrosis	22	47	21	90
Moderate fibrosis	5	22	14	41
Severe fibrosis	3	17	19	39
Cirrhosis	2	18	135	155
All fibrosis + cirrhosis	32	104	189	325
Totals	64	136	197	397

computed tomography (CT), or clinical examination was recorded (Table 1). All investigations were part of routine NHS diagnosis and so the incidence of varices and ascites is likely to be an underestimate, but mortality data are comprehensive, as all deaths are recorded by NSTS. Survival and time to varices/ascites was measured from the time of the fibrosis marker test, and censored for dead patients from the day of death.

RESULTS

Model derivation
The AUROC analysis for the mSTL regression model was 0.87 (95% confidence interval [CI] = 0.83 to 0.91, F0 versus F1–4) for any fibrosis and 0.88 (95% CI = 0.85 to 0.92, F1–2 versus F3–4) for severe fibrosis. The values are categorised into three bands to aid clinical decision, with a resultant drop in the AUROC value. For the banded logistic regression model, the mSTL, the AUROC was 0.85 (95% CI = 0.81 to 0.90) for any fibrosis and 0.84 (95% CI = 0.80 to 0.88) for severe fibrosis.

The 'easy to calculate' version of the algorithm, the STL, performed as well as the complex regression model. In the derivation cohort, the AUROC analyses for the STL were 0.78 (95% CI = 0.72 to 0.83) for any degree of fibrosis (F0 versus F1–4) and 0.81 (95% CI = 0.77 to 0.86) for severe fibrosis (F0–2 versus F3–4). The breakdown for various grades of severity for the two traffic light scores mSTL and STL is presented in Table 3.

Predicted values in the community
The study data originated from a secondary care population with a high prevalence of fibrosis and cirrhosis, and prevalence affects positive (PPV) and negative (NPV) predictive values, but the researchers were specifically interested in how the test might perform in a community population where the prevalence of fibrosis or cirrhosis is unknown. This modelling was done in the development cohort, because it is only in this cohort that it is possible to correlate the traffic light tests' data with the stage of liver fibrosis on liver biopsy. A wide range of estimates of liver fibrosis and cirrhosis for a community sample were used to illustrate the effect on PPV and NPV. PPV and NPV, as well as estimated values (ePPV and eNPV) are given in Table 4.

In the hospital setting, a red STL had a high PPV for both fibrosis (0.96) and cirrhosis (0.69), whereas a green test had only a moderate NPV for fibrosis (0.50), and a very good NPV for cirrhosis (0.97). For

dataset. The period of validation was as follows: validation cohort, mean 41 months (range 13–89 months); entire cohort, mean 46 months (range 13–89 months). Follow-up time and Kaplan–Meir survival curves were calculated from the day of the fibrosis marker test.

Mortality data were obtained from the NHS Strategic Tracing Service (NSTS); other data were extracted from the SUHT computer-based records and medical notes. The date at which oesophageal varices were first found at endoscopy, or ascites first demonstrated on ultrasound,

Table 4. Positive (PPV) and negative predictive values (NPV) of the simple traffic light grading for any fibrosis or cirrhosis in the model development cohort[a]

Setting	Parameter	Predictive value	Fibrosis	Cirrhosis
STL		**Simple TL**		
Hospital setting prior probability: cirrhosis 39%, fibrosis 82%	Red	PPV	0.96	0.69
	Red/amber	PPV	0.88	0.46
	Green	NPV	0.50	0.97
Low prevalence community prior probability: cirrhosis 4%, fibrosis 8%	Red	Est PPV	0.31	0.13
	Red/amber	Est PPV	0.12	0.05
	Green	Est NPV	0.98	1.00
Higher prevalence community prior probability: cirrhosis 10%, fibrosis 20%	Red	Est PPV	0.57	0.27
	Red/amber	Est PPV	0.29	0.13
	Green	Est NPV	0.95	0.99
mSTL		**Modified TL**		
Hospital setting prior probability: cirrhosis 39%, fibrosis 82%	Red	PPV	0.99	0.73
	Red/amber	PPV	0.91	0.48
	Green	NPV	0.59	0.99
Low prevalence community prior probability: cirrhosis 4%, fibrosis 8%	Red	Est PPV	0.63	0.15
	Red/amber	Est PPV	0.17	0.04
	Green	Est NPV	0.99	1.00
Higher prevalence community prior probability: cirrhosis 10%, fibrosis 20%	Red	Est PPV	0.84	0.31
	Red/amber	Est PPV	0.38	0.13
	Green	Est NPV	0.96	1.00

[a]As the prevalence of fibrosis/cirrhosis in a primary care population with suspected liver disease was not known, estimated PPV (Est PPV) and NPV (Est NPV) values were calculated for a range of prevalences: fibrosis (8–20%) and cirrhosis (4–10%). These ranges of risk might be expected in patients with harmful alcohol consumption or known fatty liver and steatohepatitis.

the community population, the estimated PPV of a red STL for fibrosis dropped to 0.31–0.57, and of a red/amber STL to 0.12–0.29; the estimated NPV of a green test

Table 5. STL and mSTL outcomes for survival and development of varices and ascites in the validation cohort[a]

		Validation			
STL		**Green**	**Amber**	**Red**	**Total**
Survival	Alive	202	258	148	608
	Dead	0	9	24	33
Varices	No	202	267	148	617
	Yes	0	0	24	24
Ascites	No	202	265	152	619
	Yes	0	2	20	22
Total		202	267	172	641
mSTL					
Dead = 1	Alive	231	235	92	558
	Dead	1	5	24	30
Varices	No	232	237	98	567
	Yes	0	3	18	21
Ascites	No	231	238	99	568
	Yes	1	2	17	20
Total		232	240	116	588

[a]The numbers in the STL cohort are higher because it is still possible to use the simple algorithm to calculate the STL grade with a missing blood result.

for fibrosis was 0.95–98 and for cirrhosis 0.99–1.00. The predictive values for the mSTL were marginally better.

Validation cohort
The red mSTL group had a very poor survival and a high development of complications with varices and ascites; the green STL group had an excellent survival and almost no liver complications, and the intermediate amber group had a slightly reduced survival and a low rate of liver complications (Table 5).

Kaplan–Meier plots for survival and for the development of varices and ascites are given for the validation (FU) cohort (Figure 1); compared with the green group, survival was significantly diminished in the red (Mantel–Cox, $P<0.001$), and amber groups (Mantel–Cox, $P=0.012$). Significantly higher numbers of patients with a red grade developed varices (Mantel–Cox, $P<0.001$) and ascites (Mantel–Cox, $P<0.001$), but there was no difference between green and amber grades (Figure 1) over the time course of the study.

In terms of AUROC analysis for survival, the respective data in the validation cohort were for the mSTL 0.85 (95% CI = 0.78 to 0.91); in the validation cohort, the STL was associated with slightly lower AUROC values, 0.78 (95% CI = 0.72 to 0.83), but the clinical outcomes were as good as for the mSTL (Figure 1 and Table 5).

Alanine transaminase, cirrhosis, and survival
ALT was lower in patients who developed varices (mean 42 iu/l versus 76 iu/l, $P=0.03$) or ascites (mean 49 iu/l versus 76 iu/l, $P = 0.004$); these figures are from both cohorts combined, but the same trend was seen in the creation and validation cohorts separately. ALT was essentially the same in subjects with either no fibrosis, fibrosis, or cirrhosis (73 iu/l, 82 iu/l, and 67 iu/l respectively, creation cohort), or in subjects who died (mean 74 iu/l versus 73 iu/l). In the study population, a high ALT was not a useful discriminator for the severity of liver disease.

DISCUSSION
Summary
This study has developed a 'traffic light' grading system, which is a simple-to-apply method of estimating the risk of liver fibrosis and cirrhosis in a clinic population. The score was validated in a second cohort from routine clinical data and shown to predict clinically relevant outcomes and mortality. Although the score will have lower predictive
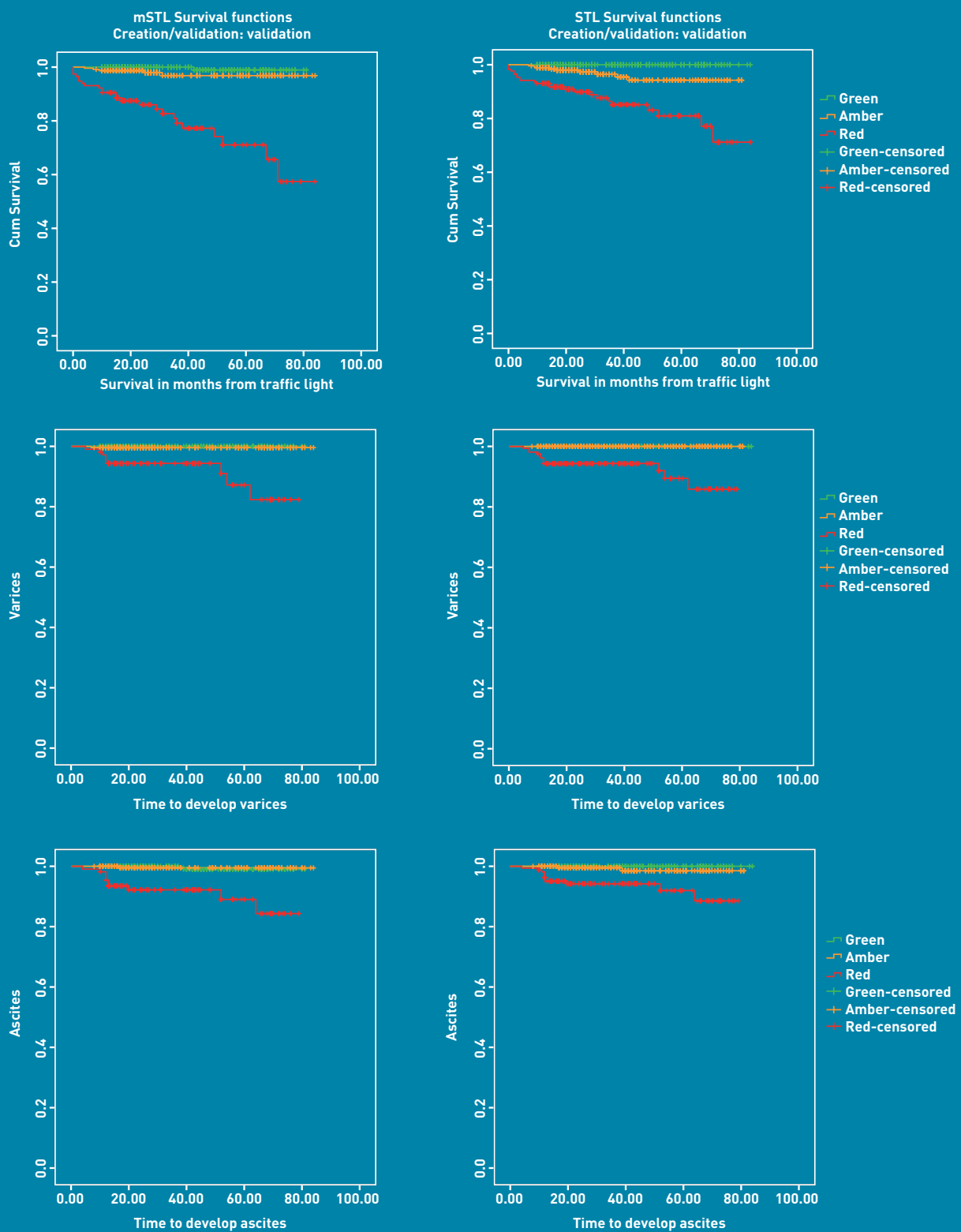
Figure 1. Kaplan Meier plots for survival and the subsequent development of varices and ascites in the validation cohort graded using the mSTL and the STL. For the complex mSTL logistical regression model, the difference between amber/red grades was significant (P<0.001) for survival, development of varices and ascites, and for survival between the green/amber groups (P = 0.05) but not varices or ascites. Using the 'easy calculate' STL, mortality in the amber group was increased and the red group decreased. The difference between amber / red grades was significant (P<0.001) for all three outcomes, and for survival (P≤0.012) between the green/amber groups, but not for ascites or varices.

Figure 1. Kaplan–Meier plots for survival and for the development of varices and ascites are given for the validation (FU) cohort

value in lower-risk community populations, it has clinical utility and provides valuable risk data to aid clinical management. A more complex score derived from the same dataset using logistic regression had marginally improved performance in terms of AUROC, but was not better at predicting relevant clinical outcomes. In practice, the STL is easier to calculate and the algorithm is still used routinely in the Southampton liver clinic.

Strengths and limitations

The traffic light score is simple to use, and in the follow-up cohort had good predictive value for clinical outcomes and survival. The score was derived using readily available clinical and biochemical data and commercially available fibrosis markers. Not all subjects in the derivation cohort had biopsy-proven cirrhosis but the authors believe the presence of ascites or varices is a sufficiently robust marker of cirrhosis to include these data. When only patients with biopsy data were included, the model was essentially the same. Only patients in whom liver biopsy results were available, or in whom there was strong clinical evidence of fibrosis/cirrhosis, were included in the development cohort, and hence those without these were included in the cohort used for survival analysis. Therefore, the follow-up cohort represents a slightly different population (ascites or varices excluded). This, in turn, will influence prognosis; as expected compared to the derivation cohort, members of the follow-up cohort have a better prognosis. Hence, the survival curves will underestimate true test performance in this clinic population.

The score and modified score were derived in a clinic population with a high risk of fibrosis and cirrhosis (82%); the score may not perform as well in a lower-risk community sample. It is important to appreciate that a study derived in a high-prevalence population, in this instance the clinic population, may perform less well when applied to a different population, due to spectrum bias. The prevalence of fibrosis and cirrhosis in the community is unknown and so the researchers modelled a wide range of prevalence for fibrosis 8–20%, and cirrhosis 4–10%. A French study of FibroTest (known as FibroSure in the US, is a patented biomarker test that uses the results of six blood serum tests to generate a score that is correlated with the degree of liver damage in people with a variety of liver disease) and elastography in 7463 normal subjects found that 2.8% had evidence of liver fibrosis.[14] In the ALDDeS study, around 14% of hazardous drinkers had a red traffic light.[15] The study considers their range of estimates for fibrosis of 8–20% is reasonable for a high-risk community group. It is widely recognised that predictive values will be lower, but spectrum effects may also change sensitivity and specificity.[16] The validation sample used for prognostic validation had a poor outcome, reflecting the underlying severity of liver disease in this population. Since it is unlikely that liver biopsies to confirm the test result will be widely adopted, it will only be possible to ascertain the performance of these tests in community samples by validation of cohorts to determine outcomes. Despite the provisos, the study considers the results to be sufficiently useful to inform initial clinical management and lifestyle advice.

Comparison with existing literature

The various methodologies for non-invasive diagnosis of liver fibrosis and cirrhosis fall into two groups, various combinations of blood tests including serum markers of fibrosis as detailed in this study, and imaging modalities including elastography. All have been subject to intensive investigation, including systematic reviews and meta-analyses,[7,9,17–20] with a consensus emerging that in liver disease of varied aetiologies, the detection of cirrhosis and severe fibrosis is accurate, but the detection of early stages of fibrosis considerably less so.

A number of studies have examined clinical outcomes, and various tests have been shown to predict survival; a study of FibroTest in 537 subjects with hepatitis C found that it was able to predict survival with an AUROC of 0.76, compared with an AUROC of 0.66 for histological staging.[20] Another study compared FibroTest (AUROC for survival = 0.69) with Fibrometer A (0.69), Hepascore (0.69), histological staging (0.69), Pugh (0.62), FIB4 (0.64), AST to platelet ratio index (APRI) (0.56), and Forns[24] (0.43) in 218 subjects with alcohol-related liver disease followed for up to 11.8 years.[22] Further studies have used the AST/platelet count ratio and the commercial ELF test (a combination of HA, P3NP, and tissue inhibitor metalloproteinase [TIMP]), which predicted survival in 457 subjects with a range of liver diseases, with an AUROC of 0.87.[23] All these outcome data are essentially similar to the findings of the present study and the study would conclude that the various non-invasive tests including the STL, mSTL, FibroTest, ELF and APRI appear to have similar accuracy in predicting outcomes, and, given the accuracy of cirrhosis prediction in general, are probably equally

Discuss this article
Contribute and read comments about this article on the Discussion Forum: http://www.rcgp.org.uk/bjgp-discuss

likely to be able to define a population at risk of variceal haemorrhage.

Implications for practice and research

There are several tests incorporating biomarkers or other measures which are showing promise in the secondary care environment as indicators of liver fibrosis/cirrhosis, with prognostic relevance. None have yet been validated in the primary care context, where issues of both disease prevalence and performance bias may reduce their predictive value. These tests are not suitable for screening in low-risk individuals but they may now have a place in the management of patients, with evidence of some non-specified liver disease or high liver risk, when making decisions about further tests or referral. The STL test results may be used to recommend lifestyle changes and to guide referral for further investigation, with a red test indicating a probability of significant underlying liver disease, and the possibility of cirrhosis. While the PPV of the test in general practice may not be as good as predicted from the development cohorts — an amber test is associated with approximately 50% likelihood of liver fibrosis in the study's secondary care setting, but perhaps only 10–30% in a primary care setting depending on the degree of liver risk — NPVs are still likely to be high. The study would therefore advocate that tests such as the STL may gain a place in the rational management of liver risk when a period of watchful waiting is appropriate because the risk of fibrosis is so low. This should allow appropriate investigation of those at higher risk, by liver specialists, and delivery of lifestyle interventions in those managed in the community.

REFERENCES

1.   Jewell J, Sheron N. Trends in European liver death rates: implications for alcohol policy. *Clin Med* 2010; **10(3):** 259–263.

2.   British Association for the Study of the Liver (BASL), British Society of Gastroenterology (BSG) (Liver Section). *A time to act: improving liver health and outcomes in liver disease.* National Plan for Liver Services UK 2009. http://eprints.soton.ac.uk/194669/1/National_Liver_Plan_2009.pdf (accessed 20 Jul 2012).

3.   Roberts SE, Goldacre MJ, Yeates D. Trends in mortality after hospital admission for liver cirrhosis in an English population from 1968 to 1999. *Gut* 2005; **54(11):** 1615–1621.

4.   Verrill C, Smith S, Sheron N. Are the opportunities to prevent alcohol related liver deaths in the UK in primary or secondary care? A retrospective clinical review and prospective interview study. *Subst Abuse Treat Prev Policy* 2006; **1:** 16.

5.   Sheron N, Moore M, Roderick W, *et al.* OP01 Alcohol and liver disease detection study ALDDeS: early results and implications. *Gut* 2011; **59(suppl 2):** 1.

6.   Leroy V, Halfon P, Bacq Y, *et al.* Diagnostic accuracy, reproducibility and robustness of fibrosis blood tests in chronic hepatitis C: a meta-analysis with individual data. *Clin Biochem* 2008; **41(16–17):** 1368–1376.

7.   Musso G, Gambino R, Cassader M, Pagano G. Meta-analysis: natural history of non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of non-invasive tests for liver disease severity. *Ann Med* 2011; **43(8):** 617–649.

8.   Forestier J, Dumortier J, Guillaud O, *et al.* Noninvasive diagnosis and prognosis of liver cirrhosis: a comparison of biological scores, elastometry, and metabolic liver function tests. *Eur J Gastroenterol Hepatol* 2010; **22(5):** 532–540.

9.   Poynard T, Ngo Y, Munteanu M, *et al.* Biomarkers of liver injury for hepatitis clinical trials: a meta-analysis of longitudinal studies. *Antivir Ther* 2010; **15(4):** 617–631.

10. Lin ZH, Xin YN, Dong QJ, *et al.* Performance of the aspartate aminotransferase-to-platelet ratio index for the staging of hepatitis C-related fibrosis: an updated meta-analysis. *Hepatology* 2011; **53(3):** 726–736.

11. Bossuyt PM, Reitsma JB, Bruns DE, *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003; **49(1):** 7–18.

12. Efron B. Estimating the error rate of a prediction rule: some improvements on cross-validation. *J Am Stat Assoc* 1983; **78:** 316–331.

13. Efron B, Tibshirani R. Improvements on cross-validation: the .632+ bootstrap method. *J Am Stat Assoc* 1997; **92:** 548–560.

14. Poynard T, Lebray P, Ingiliz P, *et al.* Prevalence of liver fibrosis and risk factors in a general population using non-invasive biomarkers (FibroTest). *BMC Gastroenterol* 2010; **10:** 40.

15. Sheron N, Moore M, Roderick P, *et al.* ALDDeS Alcohol and Liver Disease Detection Study. *Gut* 2010; **59(Suppl 2):** A1.

16. Willis BH. Spectrum bias — why clinicians need to be cautious when applying diagnostic test studies. *Fam Pract* 2008; **25(5):** 390–396.

17. Tsochatzis EA, Gurusamy KS, Ntaoula S, *et al.* Elastography for the diagnosis of severity of fibrosis in chronic liver disease: A meta-analysis of diagnostic accuracy. *J Hepatol* 2011; **54(4):** 650–659.

18. Poynard T, Ngo Y, Munteanu M, *et al.* Biomarkers of liver injury for hepatitis clinical trials: a meta-analysis of longitudinal studies. *Antivir Ther* 2010; 15(4): 617–631.

19. Stebbing J, Farouk L, Panos G, *et al.* A meta-analysis of transient elastography for the detection of hepatic fibrosis. *J Clin Gastroenterol* 2010; **44(3):** 214–219.

20. Estep JM, Birerdinc A, Younossi Z. Non-invasive diagnostic tests for non-alcoholic fatty liver disease. *Curr Mol Med* 2010; **10(2):** 166–172.

21. Ngo Y, Munteanu M, Messous D, *et al.* A prospective analysis of the prognostic value of biomarkers (FibroTest) in patients with chronic hepatitis C. *Clin Chem* 2006; **52(10):** 1887–1896.

22. Naveau S, Gaude G, Asnacios A, *et al.* Diagnostic and prognostic values of noninvasive biomarkers of fibrosis in patients with alcoholic liver disease. *Hepatology* 2009; **49(1):** 97–105.

23. Parkes J, Roderick P, Harris S, *et al.* Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. *Gut* 2010; **59(9):** 1245–1251.

24. Forns X, Ampurdanes S, Llovet JM, *et al.* Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. *Hepatology* 2002; 36(4 Pt 1): 986–992.