UNIVERSITY OF SOUTHAMPTON
FACULTY OF MEDICINE,
HEALTH AND LIFE SCIENCES

SCHOOL OF MEDICINE

Non –invasive biomarkers in chronic liver disease

Dr Julie Parkes

Thesis for the Doctor of Philosophy

August 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES
SCHOOL OF MEDICINE

Doctor of Philosophy

Non –invasive biomarkers in chronic liver disease

By Julie Parkes

Liver fibrosis is the final common pathway following chronic insult to the liver. Progression to cirrhosis is mostly asymptomatic, but the end stages of disease result in clinical events such as ascites, bleeding, infection, hepatocellular cancer and death. Liver biopsy has been regarded as the gold-standard for assessing fibrosis and whilst it provides useful information it is hazardous for patients and subject to inaccuracy. Many serum markers of liver fibrosis have been evaluated as tests for severity of fibrosis on biopsy, as a surrogate for prediction of clinical outcomes. Their performance as direct predictors of clinical outcomes as the reference standard would be ideal.

This thesis focused on serum biomarkers and had two broad aims. Firstly, the evaluation of published literature on the diagnostic accuracy of serum markers in identification of fibrosis in the three main causes of chronic liver disease (CLD) in the UK (chronic Hepatitis C (CHC), Non-alcoholic fatty liver disease (NAFLD), and Alcoholic liver disease (ALD). Secondly, the exploration of the diagnostic accuracy of a particular panel of markers (Enhanced Liver Fibrosis (ELF) in the diagnosis of fibrosis severity in external independent populations of patients with CLD. Thirdly the evaluation of the performance of ELF in prediction of clinical outcomes.

The systematic reviews highlighted the breadth of serum markers, found that markers performed better at identifying serious fibrosis than milder disease, and marked the evolution from single markers to panels of marker. One such marker panel was ELF, whose components are part of the fibrotic process, which was derived and validated in a cohort of patients recruited in 1998-2000. The performance of ELF was externally validated in eight studies in patients with CHC, NAFLD, Primary Biliary Cirrhosis (PBC) and HCV-HIV co-infection. ELF maintained its performance in all studies with AUC values >0.80 and >0.85 in identification of significant fibrosis and cirrhosis respectively. It performed particularly well in NAFLD. In the final part of the thesis, two studies were conducted to explore the accuracy of ELF in predicting clinical outcomes. One study involved follow-up of the patients in the original ELF cohort for liver related outcomes and all-cause mortality. Analyses showed that baseline ELF score can predict liver outcomes and all-cause mortality, with those people having highest ELF scores being significantly more likely to have clinical outcomes than those with lower scores. A unit change in ELF was associated with a doubling of the risk of having a liver–related outcome at 6 years. In the second prognostic study 161 patients with PBC were followed up for 8 years. The results confirmed the findings in the first study. ELF was found to be better than other markers currently used for prognosis of clinical events in PBC.

Serum markers have a role in the assessment of patient with CLD both predicting fibrosis and clinical outcomes. More research is needed to assess performance of these markers in different settings, such as Primary Care.

Contents

List of Tables

List of Figures

Appendices

Acknowledgements

I would like to thank the following people, departments and organisations:

My supervisors Paul Roderick and William Rosenberg for giving me the opportunity to pursue this research. Their unfailing support, guidance, teaching and constant encouragement have allowed me to gain a huge amount from this thesis.

I would like to thank Professor Keith Palmer for his support and for his helpful comments following the transfer viva

All of the staff at Public Health Sciences and Medical Statistics. In particular Scott Harris, who bore my quest for statistical knowledge with patience and fortititude, was a source of great support, and who taught me a lot about statistics, for which I am very grateful.

I performed the systematic reviews and developed the clinical utility modelling with Dr Neil Guha. Our collaboration was truly synergistic and it was a joy to work with him during my thesis.

Mark Wheatley and Carol Gough helped me collect data from several of the centres in the prognostic study, and Mark Wheatley was the main researcher for one of the external validation studies. They approached the work with great enthusiasm and accompanied me on some of my travels through England and Europe.

Dr Richard Cross of iQur Ltd, who performed all the laboratory analysis for ELF markers for the external validation studies, taught me all I know of the laboratory science aspect of diagnostic tests with skill and good humour.

I would like to thank all the staff at the centres participating in the prognostic study for all their help in particular Elsbeth Henderson, Katrina O'Donnell, Linda Knowles, Wendy Smeeton, Susan Williams, Devina Mallon, Madeleine Thyssen, Annie Lorton, Jeanne Prosser. They made me very welcome and performed wonders with the vagaries of the NHS record systems.

I would like to thank the collaborators who made the external validation studies possible: Professor Will Irving; Dr S Ryder, Professor C Day. Dr G Aithal; Dr V Nobili; Dr M Mayo; Professor M Peters; Dr S Bialek.

I would like to thank the principal investigators of the prognostic study: Professor C. Day, Dr D. Mutimer, Dr J. Collier, Dr G. Alexander, Dr M. Lombard, Dr J. Ramage, Professor G. Dusheiko, Professor M. Pinzani, Dr S. Milani, Professor R. Hultkrantz, Dr K. Hagen, and Dr M. Barr, without whom the prognostic study would not have taken place. In addition I would like to thank all the GPs who took the time to reply to the morbidity questionnaire which provided such good follow up of the cohort.

Ellen Sampson at Bayer/Siemens has facilitated all of the studies from Bayer/Siemens and has provided a model of best practice of how industry and academia can work together to make advances in clinical science happen.

I would like to thank all of the patients who took part in all the studies involved in the thesis.

Finally grateful thanks are due to my family, in particular my daughters Molly, Anna, and Poppy who have been unconditionally patient and supportive during the thesis.

GLOSSARY

CLD	Chronic Liver Disease
HCV	Hepatitis C Virus
CHC	Chronic Hepatitis C
NAFLD	Non-Alcoholic Fatty Liver Disease
NASH	Non-Alcoholic Steatohepatitis
ALD	Alcoholic Liver Disease
HCC	Hepatocellular Carcinoma
PBC	Primary Biliary Cirrhosis
PSC	Primary Sclerosing Cholangitis
HHC	Hereditary Haemochromatosis
AIH	Auto-Immune Hepatitis
HA	Hyaluronic Acid
PIIINP	Amino Terminal Type III Procollagen Peptide
TIMP-1	Tissue Inhibitor of Metalloprotease 1
α_2M	α_2 Macroglobulin
MMP	Matrix MetalloProtease
ALP	Alkaline Phosphatase
AST	Aspartate Amino Transaminase
ALT	Alanine Transaminase
LFT	Liver Function Test
Pl	Platelets
TGF-	Transforming Growth Factor
IFN-	Interferon
HSC	Hepatic Stellate Cell
ELISA	Enzyme-Linked ImmunoSorbent Assay
ECM	Extracellular matrix
HES	Hospital Episode Statistics
ONS	Office of National Statistics

UAPMP	Unlinked anonymous prevalence and monitoring programme
ELFa	Original published European Liver fibrosis panel
ELF	Enhanced Liver Fibrosis (serum marker panel)
ROC	Receiver Operating Characteristic
SROC	Summative Receiver Operator Characteristic
AUC	Area Under the Curve
PPV	Positive Predictive Value
NPV	Negative Predictive Value
DOR	Diagnostic Odds Ratio
OR	Odds Ratio
HR	Hazard Ratio
LR	Likelihood Ratio
BMI	Body mass index
CI	Confidence interval
IQR	Inter quartile range
ACM	All-cause mortality

Some definitions pertaining to tests (Last dictionary epidemiology 2nd edition):

Reliability is the degree of stability shown when a measurement is repeated under identical conditions.

Reproducibility (repeatability) A test is repeatable if the results are identical or closely similar each time it is conducted.

Validity is the degree to which a measurement measures what it purports to measure.

CHAPTER 1
RATIONALE FOR THESIS

This chapter will provide the rationale behind the research presented in this thesis, and outline the structure and aims of the study.

1.1 Background of liver disease

Chronic Liver Disease (CLD) is a significant and current public health problem. The three main causes of CLD in the UK are Alcoholic Liver Disease (ALD), Chronic Hepatitis C (CHC), and Non-Alcoholic Fatty Liver Disease (NAFLD). The major risk factors for the development of CLD are excessive consumption of alcohol, injecting drug use (leading to CHC) and obesity (leading to NAFLD). These risk factors are rising rapidly and concurrently in the UK and are leading to an epidemic of CLD.

The Chief Medical Officer's report (2001) highlighted the increase in mortality from cirrhosis over time especially in young adults[1]. Deaths from chronic liver disease have increased by eight times in men aged 35-44 and seven times in women over the past three decades. This increasing trend was maintained in the last decade with cirrhosis mortality doubling in Scottish men between 1987/91 and 1997/2001 and rising by 69% in men in England and Wales[2]. Mortality in women increased by almost half (44% England 46% Scotland). Age-standardised rates in England and Wales for men aged 45-64 is 26.7 per 100,000, and women 9.4 per 100,000 (2002). This rise is the steepest in Europe and differs from the decline seen in other European countries. The most recent mortality rate from cirrhosis and chronic liver disease is 151 per 1000,000 with 6,237 persons dying from CLD (http://www.nchod.nhs.uk). The rise in mortality is likely to be due to increasing incidence of CLD although increasing case fatality, changes in diagnostic labeling or more accurate death certification could be other explanations. It is estimated that at least 250,000 people have Hepatitis C in the UK, and due to the difficulties in ascertaining prevalence in vulnerable, traditionally hard to reach groups (such as injecting drug users) this figure is likely to be higher[3]. Recent modeling of future trends has predicted a dramatic rise in HCV related cirrhosis and deaths from hepatocellular cancer in the next 10 years[4]. Studies have suggested that 20-30% of the general population

may be affected by NAFLD[5] and the number of alcohol related deaths has risen from 4,144 in 1991 to 8,758 in 2006 in the UK. In 2006/07, there were 43,548 admissions for alcoholic liver which is a tripling of numbers since 1995/6.

There are effective treatments for people with CLD. Rapid developments in anti-viral therapeutics have led to accepted regimes for CHC using pegylated interferon and ribavirin in combination[6;7]. This leads to a viral clearance in up to 80% of patients. Lifestyle modifications in alcohol consumption and obesity may lead to improvement in liver fibrosis and prognosis in ALD and NAFLD[8-10]. Even in those people with compensated late stage disease (cirrhosis) it is important to identify people as surveillance for hepatocellular carcinoma (annual incidence 3-4% in patients with cirrhosis), screening and management of oesophageal varices and optimising health in preparation for transplant can be planned.

1.2 Evidence for the reduction in morbidity and mortality in CLD by earlier identification of cirrhosis

Much of this thesis is concerned with the earlier diagnosis of significant liver disease using serum markers, with the ultimate aim of reducing morbidity and mortality in patients. The evidence for the earlier diagnosis of cirrhosis reducing morbidity and delaying mortality is an important issue in the potential clinical use of biomarkers. The major treatable complications of cirrhosis are portal hypertension and hepatocellular cancer. Evidence shows that earlier detection translates into better survival and reduced morbidity.

Early detection of cirrhosis and instigation of prophylactic treatment of portal hypertension with beta-blockers and variceal banding has been shown to reduce morbidity (Relative risk reduction 48%) and mortality (by 16%) [11-13]. This requires screening to identify those patients with high risk of bleeding from varices. Practice varies but current UK guidelines recommend that in the absence of a standard method of risk stratification all patients with cirrhosis should be endoscoped[14]. However, currently many cases of cirrhosis are not identified until they present with variceal bleeding and at this stage there is an associated mortality of 25%. Earlier detection of cirrhosis before any compensation event and instigation of prophylaxis would translate into improved survival and less cost to the health service. In secondary prevention

once an episode of bleeding has occurred there has been increased survival following improved strategies over the past decades[15].

Hepatocellular cancer

Retrospective analyses have identified patient characteristics that are associated with better outcomes for tumour resection and liver transplantation in those with hepatocellular cancer (HCC). These include the presence of a single lesion less than 5 centimeters in diameter, or three or fewer lesions less than 3 centimetres in diameter[16]. However many HCC are identified at a time when the tumour has grown larger than 5 cm or when there are more than three 3 cm tumours ruling out successful transplantation. Major factors associated with increased survival in untreated patients with HCC are less severe underlying liver dysfunction and small size of tumour at detection. Between 50-90% of patients with HCC on a background of Child Pugh A cirrhosis -mild clinical disease (see Appendix 1b) - will survive 1 year untreated, compared to 20% Child Pugh C cirrhosis (severe clinical disease). Small HCCs have relatively long tumour doubling times and for those with <5cm survival time is 81-100% at 1 year and 17-21% at 3 years with no therapy[17;18] . Recent advances in treatment of HCC have improved prognosis[19].

More effective methods such as biomarkers for detecting cirrhosis in a higher proportion of cases would permit earlier instigation of tumour surveillance. This in turn may result in earlier detection and a greater proportion of patients being cured through resection or transplantation, translating into greater patient benefit and cost-effectiveness[20-22]. Some of these screening studies have been limited by lead-time and length-time bias[23] . This was accounted for in a later study which showed survival benefit from surveillance[24]. National guidelines at the present time recommend surveillance in patients with cirrhosis arising from CHC, Chronic Hepatitis B, ALD (if abstinent/likely to comply with treatment), Primary Biliary Cirrhosis (if male), and Haemochromatosis[25]. Surveillance has been supported by a recent Health Technology Assessment publication of a systematic review and economic analysis which recommended annual/6 monthly screening with ultrasound and alpha-feto protein, based on modeling that suggested this strategy would halve the number who die from HCC, and would be cost effective with £30,000-£40,000 per QALY[26].

While successful liver transplantation can extend life expectancy in all cases, limited supply of organs means that the identification of appropriate subjects and optimal timing of transplantation are essential.

It is therefore important to diagnose CLD and identify people who may benefit from treatment, to be able to monitor such treatment and to instigate surveillance for oesophageal varices and HCC in those with asymptomatic (or compensated) cirrhosis.

1.3 Diagnosis of Liver fibrosis

Liver fibrosis is the common pathological response of the liver to any injurious agent whether it is virus, alcohol or fat. The diagnosis of CLD has traditionally used the identification of liver fibrosis as a proxy for clinical outcome and this has been achieved using liver histology obtained via percutaneous or transjugular biopsy. These are invasive techniques and histology so derived is subject to sampling error, and inter- and intra-observer variability in interpretation making it a flawed reference standard. In the past few decades a search has been made for alternative robust non-invasive diagnostic tests which could identify the presence of liver fibrosis. Amongst these have been biomarkers which identify decline in liver function or detect accumulation of extra cellular matrix which occurs during fibrogenesis and collagen breakdown, radiological techniques (ultrasound "elastography" of the liver where a shear wave is propagated and its transit through the liver is monitored to detect reduction of elasticity due to fibrosis and microbubble techniques which measure differing transit times of radiological medium through the liver depending on the amount of fibrosis[27;28] and magnetic resonance imaging and spectroscopy[29]). Liver fibrosis acts as a proxy for important clinical outcomes such as decompensated cirrhosis or death. It has been suggested that direct assessment of alternative biomarkers in CLD against clinical outcomes would provide more meaningful performance indicators.

One such set of serum markers were identified in a rigorous multi-centre cohort study in more than one thousand people referred to hepatology for a liver biopsy in the investigation of CLD [30](see Chapter 4 for further detail). Access to these study data was allowed by the investigators and study funders, (Bayer Healthcare now Siemens Diagnostics). This thesis

describes the diagnostic and prognostic performance of this panel of markers in relation to liver histology and clinical outcomes.

1.4 Aims and Objectives

The main aims and objectives of the thesis are reported below

Aims

1. To evaluate the **diagnostic performance of serum markers** of liver fibrosis in chronic liver disease by;
    a. Conducting systematic literature reviews of the diagnostic performance of serum markers in the three main causes of CLD
    b. Assessment of diagnostic performance of one particular published panel of markers (European Liver Fibrosis-ELF markers)
    c. External validation of ELF in independent populations of patients with CLD

2. To evaluate the **prognostic performance** of serum markers for clinical outcomes by
    a. Follow up study of the ELF cohort of patients in 13 centres in England and continental Europe for significant liver related morbidity, liver transplantation and mortality
    b. Follow up of a cohort of patients with PBC, recruited in the USA, using ELF serum markers and comparison with biomarkers that are currently used in clinical practice.

1.5 Outline of thesis

Chapter 2 will provide the background of CLD, including the anatomy and physiology of the liver, the pathological processes involved in liver fibrosis, and key epidemiological data that provide evidence for the epidemic of CLD in the UK.

Chapter 3 presents three systematic reviews of the diagnostic performance of biomarkers for each of the major causes of CLD.

Chapter 4 reports evaluations of ELF in independent populations. Background data on the original ELF cohort study, in addition to the eight evaluation studies of the **diagnostic** performance of ELF markers in different populations of people with CLD are reported.

Chapter 5 presents a review of the literature to assess the prognostic performance of serum markers in predicting clinical outcomes in CLD. The **prognostic** performance of the ELF markers was assessed against clinical outcome measures by following up patients in two cohorts: (i) All those people in the original ELF cohort followed up for significant liver related morbidity, liver transplant and mortality. Evaluation of baseline ELF was conducted to assess the performance of the panel in the prediction of clinical outcome. An analysis to evaluate the performance of a novel serum marker model and limited comparison to other serum markers of liver fibrosis using baseline parameters are presented; (ii) a cohort of patients with PBC initially recruited to a clinical trial in which the prognostic performance of ELF markers was evaluated and compared to clinical scores currently used in clinical practice.

Chapter 6 presents an overall discussion of the research presented in this thesis and the future work that has been identified.

CHAPTER 2

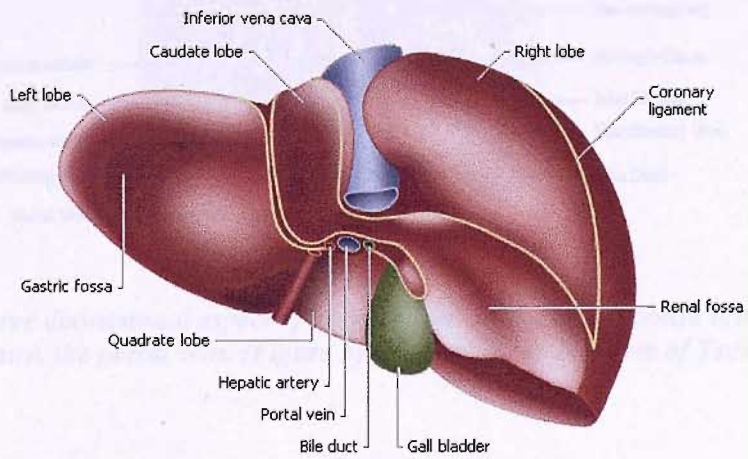
BACKGROUND OF CHRONIC LIVER DISEASE

In the first section of this chapter the anatomy and physiology of the liver will be described, and the pathophysiology of fibrosis will be outlined. The second section will evaluate the importance of liver fibrosis and provide an assessment of the performance of the liver biopsy as a diagnostic test for liver fibrosis. The third section will review the epidemiology of CLD in the UK, identify the major causes of CLD, and give an overview of their natural history.

2.1 The Liver and its function

The liver is situated in the upper part of the abdominal cavity against the diaphragm. It has two main lobes (the right being the larger), and two accessory lobes. The liver is supplied by two main blood vessels to its right lobe: the hepatic artery and the portal vein. The hepatic artery normally comes off the coeliac trunk. The portal vein brings venous blood from the spleen, pancreas, and intestines, so that the liver can process the nutrients and by products of food digestion. The hepatic veins drain directly into the inferior vena cava. (Figure 2.1)

Figure 2.1 Diagrammatic view of the liver

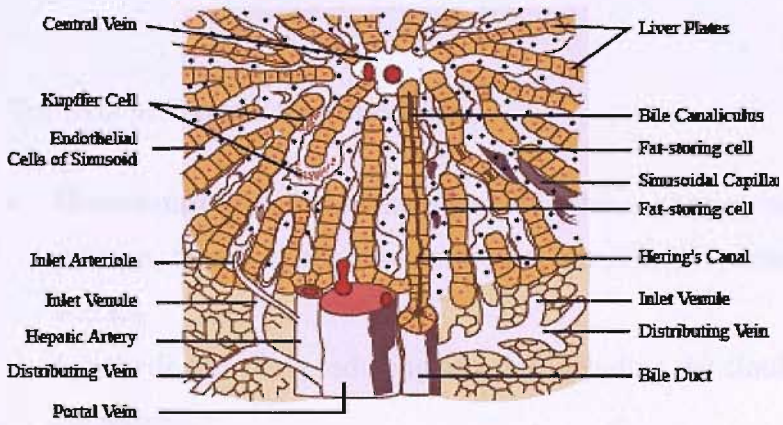


The macroscopic anatomy of the liver showing two main lobes and major vasculature

The liver is the largest solid organ in the body, weighing around 1.6 kgs in males and 1.3 kgs in women. It holds approximately 13% of the total blood supply at any given moment. The liver receives over 25% of the total resting cardiac output and is responsible for over 20% of the body's resting oxygen consumption.

The liver is organized into lobules which take the shape of polygonal prisms. Each lobule is typically hexagonal in cross section and is centered on the central vein which is a branch of the hepatic vein. Within each lobule, hepatocytes are arranged into hepatic cords separated by adjacent sinusoids. The fenestrated endothelium lining the sinusoids lies immediately adjacent to the cords, with no basement membrane and practically no intervening connective tissue, so that each hepatocyte is bathed on two faces by blood plasma. At the corners between adjacent lobules are the portal tracts. These are regions of connective tissue which include branches of the bile duct, the portal vein, and the hepatic artery (Figure 2.2).

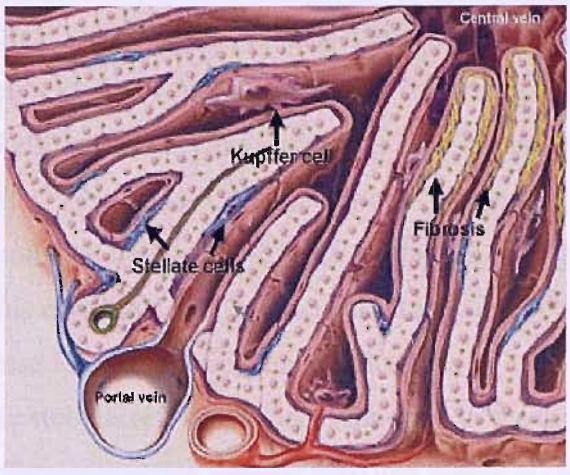
Figure 2.2 Schematic diagram of Liver Architecture



Three dimensional aspect of normal liver. In the upper centre is the central vein; in the lower centre, the portal vein. (Figure by Massachusetts Institute of Technology: OpenCourseWare)

The space between the endothelium and the hepatocytes is the space of Disse, containing a network of reticular fibres (collagen type III) which hold the hepatocytes together (figure 2.3).

Figure 2.3 Schematic diagram of sinusoidal architecture



Sinusoidal architecture and location of hepatic stellate cells. In normal liver cords of hepatocytes are surrounded by a fenestrated endothelial lining. In the intervening space of Disse are the hepatic stellate cells (blue). Kupfer cells (purple) are intra-sinusoidal and are shown as adherent to endothelial wall. Activation of stellate cells can lead to accumulation of extra cellular matrix (yellow bands to right of diagram)[31]

The liver performs 6 main functions:

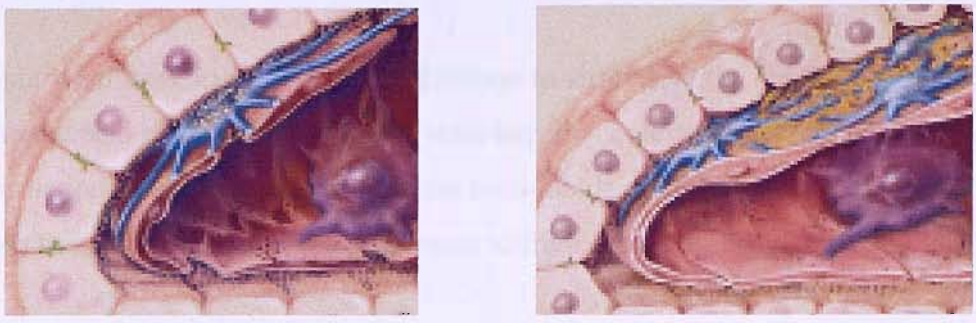
- **Homeostatic** function maintaining the balance of many nutrients producing glucose, proteins, fat and cholesterol hormones, vitamins, in particular the fat soluble ones (A, D, E and K),
- **Synthetic** function producing proteins including the clotting factors, bile acids, and cholesterol,
- **Excretory** function producing cholesterol, bile acids, phospholipids, bilirubin, drugs, toxins (e.g. pesticides, insecticides, heavy metals),
- **Storage** of vitamins and cholesterol
- **Filter** of toxins from the gut, nutrients such as amino acids, sugar and fat, bilirubin, bile acids, and drugs,
- **Immune defence** via the excretion of IgA and specialised macrophages (Kupfer cells).

2.2 Pathophysiology of liver fibrosis

Fibrosis is a wound healing response to injury and is a dynamic process involving matrix deposition, degradation and re-modelling. This has three main sequelae- firstly it distorts hepatic architecture and vasculature, secondly it disrupts normal hepatic functions, and lastly increases the likelihood of neoplastic changes. The molecular mechanisms of fibrosis are similar regardless of the nature of the insult/or injury sustained by the liver. Extracellular matrix (ECM) constituents include fibrillar collagens (collagen type I and III), proteoglycans and glycoproteins organised in a 3 dimensional network. These ECM proteins are deposited in the liver and cause loss of function and architecture of hepatocytes. They are produced by the activated hepatic stellate cell whose usual function in the quiescent state is to support the matrix which offers supportive cellular structure whilst allowing easy diffusion of solutes from the sinusoid to the hepatocyte in the space of Disse. Hepatic Stellate Cells exist in quiescent and activated states. In their quiescent state they store vitamin A and have the characteristic appearance of adipocytes. On activation they transform into myofibroblasts. Activation of the stellate cell occurs in two stages, initiation (where the stellate cell becomes sensitised to cytokines -such as transforming growth factor (TGF) β); and perpetuation, where the stellate cell proliferates and migrates in response to cytokines. There is therefore an accumulation of stellate cells in areas of injury and an increase in amount of matrix produced. Fibrillar matrix is laid down in this process, separating the hepatocytes from sinusoidal blood depriving them of nutrients, and reducing their function. The usual fenestration of the sinusoid is thus eliminated by the matrix leading to capillarisation of the sinusoid. The ECM becomes denser and reticulated and so more resistant to enzyme degradation than ECM in normal livers (Figure 2.4).

However this is a dynamic process and there are compensatory mechanisms involving inhibitors and modulators of ECM production and degradation. For example, matrix metalloproteinases are enzymes that are proteolytic and are capable of degrading matrix. These enzymes are inhibited by the binding of Tissue Inhibitors of Metalloproteinases (TIMPs). MMP2 may have a role in the apoptosis of stellate cells and TIMP1 inhibits MMP2 thus blocking apoptosis and perpetuating ECM production. The ebb and flow of matrix deposition is thus in part a balance between TIMPs and activated Metalloproteinases.

Figure 2.4 Schematic diagrams of the cellular aspects of fibrosis



(a) Normal sinusoidal architecture with a stellate cell (blue) with foot processes that encircle the sinusoid. (b) During liver injury stellate cells multiply and are surrounded by fibrillar matrix, contributing to closure of endothelial fenestration and loss of hepatocyte microvilli[32].

In viral hepatitis fibrosis begins around the portal areas and gradually extends into the lobules towards the central veins, with septa formation and then bridging fibrosis. Finally extensive fibrosis links portal and central areas and this is the cirrhosis stage of the fibrotic process. In some forms of fibrosis such as that due to alcohol, early matrix deposition is more pronounced in the peri-central region but progresses to produce bridging between lobules and cirrhosis. Cirrhosis is characterized by the formation of regenerative nodules of liver parenchyma that are separated by and encapsulated in fibrotic septa and is associated with major blood flow and structural changes. Fibrosis has been shown to be a reversible process with the removal of the insult/primary cause of damage[33] and there is emerging evidence that cirrhosis itself may have an element of reversibility[34].

In different aetiologies of CLD the processes of fibrosis may differ in the initial way in which the liver responds to the viral/metabolic/immunological insult and the development of the fibrotic processes. For example in ALD liver damage occurs through several interrelated pathways. Oxidative stress plays a pivotal role, promoting hepatocyte necrosis and apoptosis, which are exaggerated in the alcoholic who is deficient in antioxidants such as glutathione and vitamin E. Free radicals initiate lipid peroxidation, which causes inflammation which triggers fibrosis where collagen is deposited in a typical perivenular and pericellular pattern. Inflammation is also incited by acetaldehyde which, when bound covalently to cellular

proteins. Histologically, the earliest changes in alcoholic hepatitis are located predominantly around the central vein[35;36].

Primary Biliary Cirrhosis (PBC) is different to viral hepatitis, NAFLD and ALD in that it is characterised by slowly progressive intra-hepatic cholestasis because of immunological destruction of epithelial cells lining the intra-lobular and septal bile ducts. Such cholestasis leads to loss of bile ducts and progresses to fibrosis and biliary cirrhosis.

2.3 Clinical aspects of fibrosis

Fibrosis of the liver in itself is clinically asymptomatic, but as fibrosis develops, the functioning of hepatocytes becomes compromised, cell death occurs, and the liver's synthetic, metabolic and filtering roles are compromised. Serum levels of liver enzymes such as ALT and AST may increase, albumin may decrease and bilirubin increases as the liver's capacity to process of bilirubin is reduced. Development of cirrhosis leads to a further compromise of function and jaundice (due to rising bilirubin), hepatic encephalopathy (due to reduced metabolism by the liver of toxic products such as ammonia from protein breakdown), and coagulopathy due to decrease in production of clotting factors become clinically apparent. Disruption of portal blood flow by the extensive fibrotic processes in the liver causes resistance and increased flow in the portal system leading to portal hypertension.

Portal Hypertension is defined as a raised portal pressure greater than 15mm Hg. The major complications of portal hypertension are ascites (accumulation of fluid in abdominal cavity), development of varices at sites of porto-systemic anastamosis, gastrointestinal haemorrhage from varices in the oesophagus and rectum and renal dysfunction. Prospective studies have shown that 90% of patients with cirrhosis will develop varices and that a third of them will bleed, carrying a short term mortality of 50% in the group with the most severe liver dysfunction. Up to 70% of patients who do not receive treatment die within 1 year of the initial bleeding episode. Risk factors for bleeding have been identified and include portal pressure values (>12mmHg), location and size of varix (larger ones and in fundus of stomach), degree of liver failure and presence of ascites.

Encephalopathy is usually due to porto-systemic shunts as part of fibrotic progress or as a consequence of therapy (TIPS insertion). Sub-acute encephalopathy may only be apparent

using psychometric testing and has been reported to be as frequent as 84% in some cirrhosis studies[37;38] . Acute deterioration may be precipitated by infection, dehydration, haemorrhage, protein ingestion or drugs.

Hepatocellular carcinoma (HCC) is a clinical complication of cirrhosis with 80% occurring in the cirrhotic liver[39]. The annual risk of developing HCC in viral hepatitis is estimated at 3-8% and in the UK about 1500 people die each year from HCC. 30% of explanted livers at the time of transplantation for CHC have undetected HCC[40]. HCCs develop as small nodules with estimated doubling times varying between 1-19 months with a median of 6 months.

The natural history of cirrhosis and time to decompensation varies depending on the cause of CLD. Overall, if there is no decompensation the 10-year survival is approximately 75%.The risk of decompensation is roughly 4-5% per year in a patient with Child's cirrhosis, and after decompensation, the probability of 5 year survival without transplant is 35-50%.

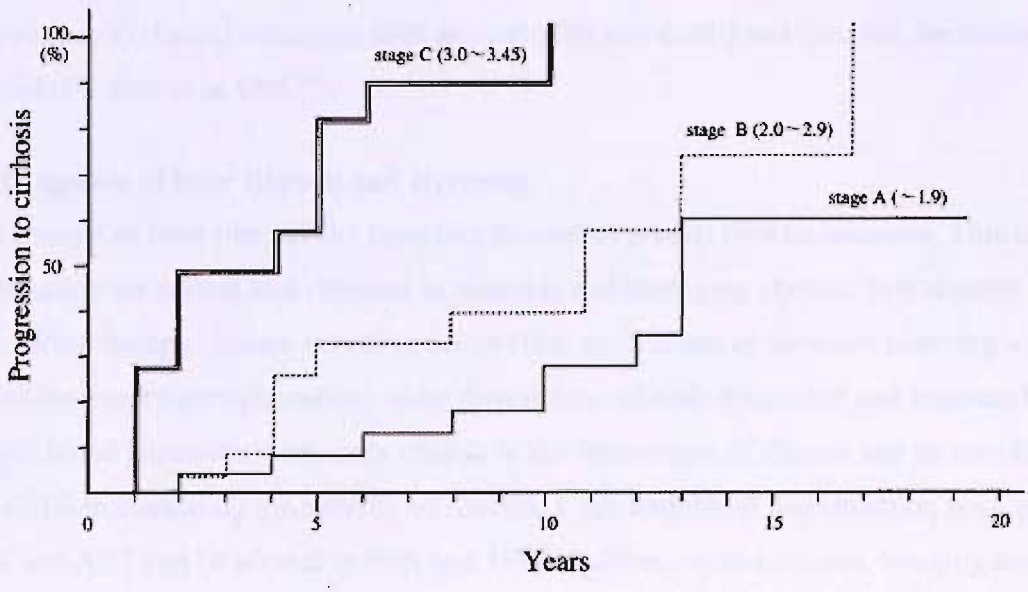
2.4 Disease progression

Cirrhosis is the end stage of liver fibrosis. There is a progression to cirrhosis through mild, moderate and severe fibrosis stages. These have been measured histologically in samples derived from liver biopsy. Initially in the context of clinical trials pathologists have staged them according to various scales, assigning ordinal categorical variables to the different stages that are described on the basis of their visual appearance. Examples of such scales are Scheuer (5 stages), METAVIR (5 stages), Ishak (7 stages), Knodell[41-43] (Appendix 1) .More often histological scores that are used in clinical practice are locally derived, and generally assign biopsies to mild, moderate, or severe fibrosis. These histological classification systems have led to the widespread belief that progression in fibrosis is linear and that the amount of fibrosis in Stage 1 is half as much as the fibrosis in Stage 2. However it has been shown that fibrosis progression is not linear between histological stages, or over time, varies between individuals, and between different causes of CLD[44;45].

Factors that have been shown to contribute to the more rapid progression in CHC are alcohol use, male sex, acquisition of infection >40 years of age, long duration of infection, immunosuppression, lack of response to anti-viral therapy, co-infection with other blood borne viruses, and older age[46].

Studies have also shown that the degree of necro-inflammation at the first biopsy can predict future fibrosis. Higher stage fibrosis on initial/index biopsy is associated with progressively larger numbers who develop cirrhosis and an inverse relationship can be shown between stage of fibrosis and time of progression to cirrhosis. For example in CHC, 60% of Stage 1 (on index biopsy) progress to cirrhosis by 19 years; all patients with stages 2 and 3 on index biopsy did so by 17 and 10 years respectively[47-49] (Figure 2.5).

Figure 2.5 Progression to cirrhosis in 70 patients with CHC over time



Stage A= little or no portal fibrosis Stage B= portal/periportal fibrosis with or without portal-portal bridging fibrosis Stage C= septal fibrosis with regions of incomplete nodular regeneration. Progression to cirrhosis is more rapid in more severe stages of fibrosis[47].

In non alcoholic steatohepatitis (NASH) risk factors for progression of fibrosis include obesity, insulin resistance, and older age[50]. Data from a retrospective cohort of biopsy proven NAFLD showed that the presence of baseline fibrosis predicted liver related death[51]. The rate of progression is similar in HCV and HBV, more rapid in ALD in those people who continue to drink heavily and fastest in HCV-HIV co-infection. Progression of fibrosis occurs through a common pathway but there are many factors that determine the speed and extent of the

fibrosis. Recent studies have shown that estimates of the progression rate to cirrhosis may be affected by recruitment of patients –those in community setting having slower rates of progression to cirrhosis (20 years) whilst those in a tertiary hepatology centre having quicker progress[52]. Authors suggested that the most likely explanation was recruitment and selection bias.

A recent systematic review of the natural history and prognostic indicators of survival in cirrhosis showed that the one year and two year and final cumulative survival were 78% and 75% and 61% respectively. Compensated cirrhosis had a better prognosis than decompensated cirrhosis. Therefore assessment of the stage of fibrosis has been used as a surrogate measures for prognosis (clinical outcomes such as morbidity and death) and the need for therapy in individuals such as in CHC[53] .

2.5 Diagnosis of liver fibrosis and cirrhosis

The amount of liver fibrosis can therefore be used to predict clinical outcome. This is useful information for patient and clinician in planning and managing chronic liver disease. It can help direct therapy, initiate surveillance for HCC and varices or facilitate planning a smooth transition toward transplantation. How then is liver fibrosis diagnosed and measured? Routine simple blood parameters may only change in the later stages of disease and so are often unhelpful in evaluating the severity of fibrosis. Even traditional liver function tests such as AST and ALT can be normal in 10% and 35% of patients with cirrhosis. Imaging modalities such as ultra-sound are unable to detect fibrosis with acceptable accuracy[54].

Liver biopsy

Since the first liver biopsy was performed in 1883 by Paul Erlich, this has been the method of choice to examine liver fibrosis. Biopsy has strengths in addition to the information it gives on fibrosis severity. It may provide information on diagnosis- for example iron in hepatic parenchyma suggest haemachromatosis, it can provide information on underlying pathological processes, and lastly it can provide additional information that can help in prognosis such as inflammation in CHC which is suggestive of future fibrosis. However this technique has limitations. The biopsy may be obtained percutaneously either blind or with ultrasound

guidance (now the most common), by the transjugular route if there is a problem with clotting (often found in end stage liver disease) or at laparoscopy. It is an invasive technique with a recognised morbidity and mortality. Large studies have found that the incidence of pain (30%), morbidity (0.3%) and mortality (0.003%) is not insubstantial[55].

Sampling error is an inherent problem and not surprising when one considers that a biopsy specimen measures1/50,000th of the entire organ and is used in diseases that may not be homogenously distributed. Post mortem studies have shown that in cirrhotic livers a single biopsy will identify cirrhosis in 16/20 cases, and only rises to 100% when 3 biopsies have been taken[56]. When right and left lobes are sampled at laparoscopy investigators have found a 33 % difference in reported stage of fibrosis[57]. Other authors found a 45% difference in scores between 2 samples from the right lobe[58].

Additionally, there are inter- and intra-observer differences in reporting of histological features by pathologists. This variability is lower for the diagnosis of cirrhosis but some studies have shown kappa scores as low as 0.59[59]. Studies have also showed that the level of experience (specialization, duration and location of practice) had more influence on agreement than the characteristics of the specimen (length fibrosis class number). Studies on the quality of biopsy have shown that the bigger the biopsy length the less is the sampling variability with biopsies 25mm being the optimum length. Even at this length the sensitivity is 75%[60]. Smaller biopsies-both length and width have also been shown in other studies to tend to underestimate the severity of fibrosis[61]. This study recommended 20mm length 1.4mm wide with at least 12 complete portal tracts for specimens to be considered to be accurate. Reduction in the number of classes within the histological scale improves the quality of agreement. The cost effectiveness of liver biopsy has not been widely reported although it incurs direct (cost of procedure analysis and hospital time) about $1500-2000) and indirect costs (time off work).Direct costs vary between hospital and countries.

The pathological classification of fibrosis severity into stages artificially represents fibrosis as an ordinal categorical variable with a linear quantum progression in severity from 0 to 4 or 6. This does not accurately reflect the dynamic biological process of fibrosis and may constrain

any test that is measured as a continuous variable which more captures the processes underlying fibrosis. By its very nature the biopsy can give only a static picture of the liver architecture whereas fibrosis is a dynamic process.

2.6 Epidemiology of Chronic Liver disease

The mechanisms involved in the development of liver fibrosis, and its association with prognosis has been outlined above. The wider picture of the current prevalence of CLD and estimates of the future burden are needed to confirm that it is a disease of importance to patients, clinicians, and the health service.

2.6.1 Background

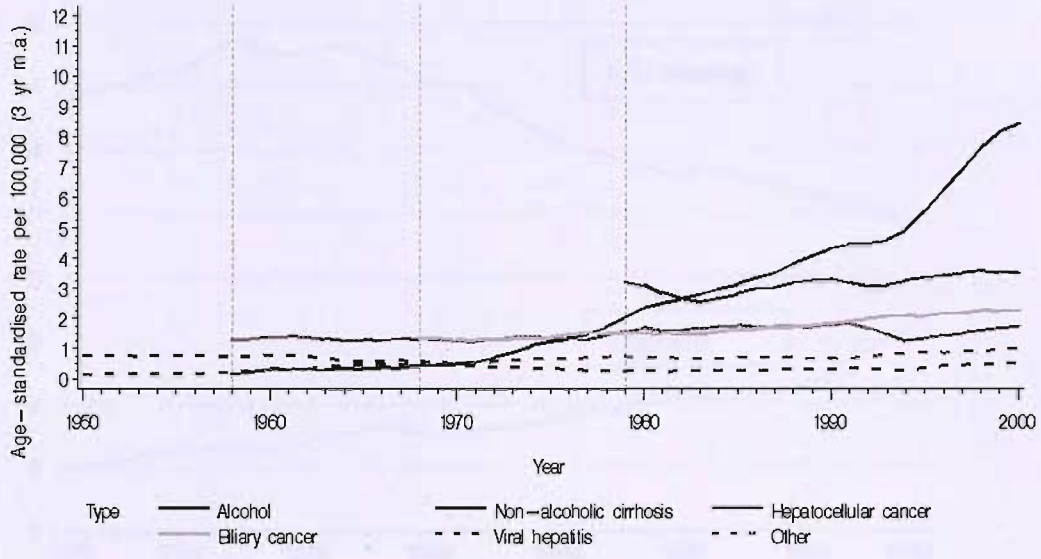
Whatever the insult to the liver for example, viral, alcohol, and obesity, it responds by attempting to heal the wound as described above. The resultant chronic liver disease (CLD) often remains clinically asymptomatic for many years making estimates of the true incidence and prevalence of CLD difficult to ascertain. Other problems in determining the burden of CLD include under-ascertainment of CLD using death certification as there is still stigma attached to some CLD especially that associated with lifestyle choices such as ALD or CHC acquired via injecting drug use.

A pragmatic difficulty in investigating the epidemiology of CLD is the lack of routine computerised standardized data collection; the Health Protection Agency collects data on laboratory results of positive viral hepatitis tests but this only gives information on those tested and is often incomplete. Where datasets exist, the validity of many routine data (e.g. ONS mortality and HES data) is unknown.

2.6.2 Trends in CLD

Best estimates using a variety of sources including routine datasets, have shown that CLD is a serious public health problem encompassing not only physical and psychological morbidity and mortality, but also incurring significant societal costs. From national mortality data CLD is the 5th commonest cause of death in middle-age in the UK[62].

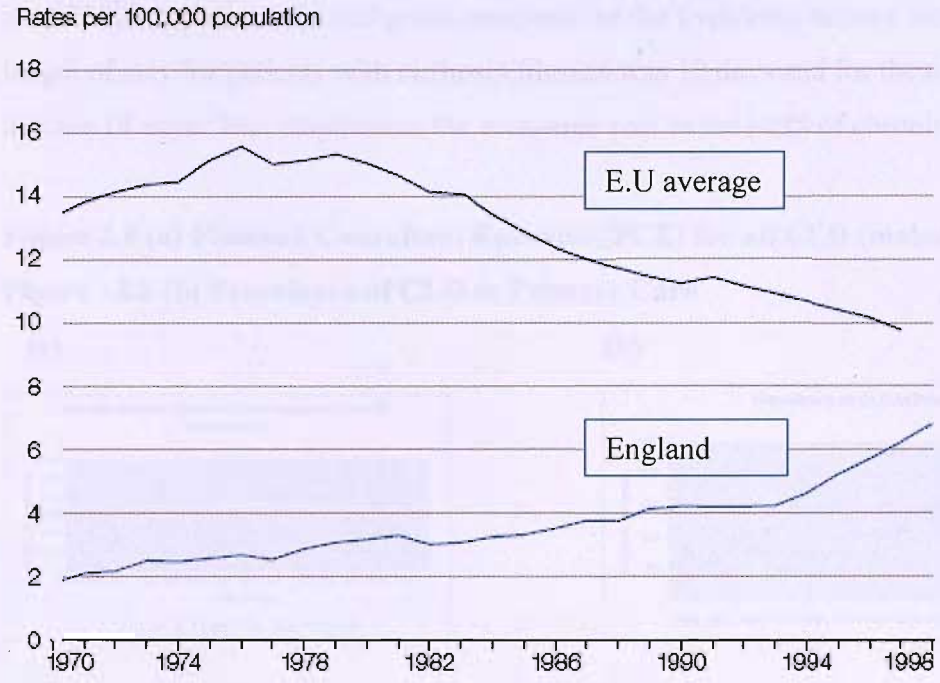
Figure 2.6 Age standardised mortality rates England 1950-2000



Age standardised mortality rate per 100,000 population over past 60 years showing a rise in CLD especially ALD in the last 3 decades.

Morbidity and mortality in England have risen rapidly especially over the last two decades, and continue to rise whilst death rates from CLD in Europe have declined in the same time frame. (See Figures 2.6 and 2.7) This upward trend in the UK has been confirmed in two studies of cirrhosis in the last two years[2, 63].

Figure 2.7 Average age standardised death rates from chronic liver disease England & European Union



Age standardised mortality rate per 100,000 populations in England and European Union showing the continuing increase in CLD mortality in England and fall in the EU
(From World Health Organisation (2001))

The age standardised mortality rate in 2004 was 151 per million population (males)- see Table 1.

Table 2.1 Age standardised mortality rates for CLD for England 2001-2004

Year	Age standardised mortality rate per 1,000,000 males	Age standardized mortality rate per 1,000,000 females
2001	139	77
2002	144	79
2003	157	81
2004	151	83

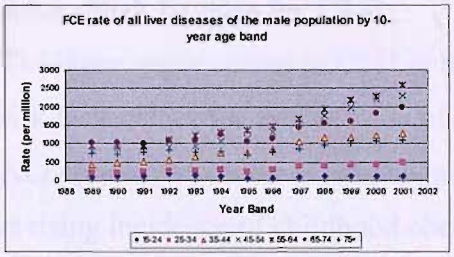
There has been a steady rise in morbidity as measured by finished consultant episodes over the past 10 years (Hospital Episode Statistics data) (see Figure 2.8a). In 2004/05 there were 34,323 admissions for chronic liver diseases (312,884 bed days) and almost 4000 admissions for fibrosis and cirrhosis of the liver (25,000 bed days). The mean age of patients was 53 years

and mean length of stay was 6 days. There were 4,344 admissions for chronic viral hepatitis, most of which (81%) were CHC, 23,000 for ALD (45% of total CLD admissions) and just over 5000 admissions for malignant neoplasm of the liver/intra-hepatic bile ducts. The mean length of stay for patients with cirrhosis/fibrosis was 12 days and for those with alcoholic liver disease 14 days. This emphasises the economic cost to the NHS of chronic liver disease.

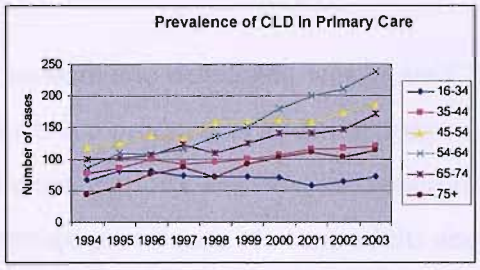
Figure 2.8 (a) Finished Consultant Episodes (FCE) for all CLD (males)

Figure 2.8 (b) Prevalence of CLD in Primary Care

(a)



(b)



Prevalence of morbidity of CLD is increasing shown by increase in FCE (The time spent under the care of a particular consultant is one episode of care) of all CLD ICD codes (a)
Data derived from Mediplus dataset which collates data on Read coded information on primary care consultations shows an increase in prevalence identified in primary care (b).

These figures are likely to be an underestimate of the burden of current disease due to the limitations of the datasets. In addition, they may represent the tip of a disease iceberg as there is a lengthy latency period for liver disease, patients with hepatitis C infection or excess alcohol consumption are traditionally hard to reach, and trends in some of the major risk factors show an increase in prevalence.

In the UK consumption of alcohol has increased especially amongst young females. Also the pattern of drinking has changed with more binge drinking which is associated with increased risk of liver damage. The number of alcohol-related deaths in England and Wales has risen throughout the 1980's and 1990's with alcohol related mortality rates in 2003 being 11.6 per 100,000. There has been a similar rise in the burden of ALD with a doubling of admissions from 1995/96 to 2006/7. (Hospital Episode Statistics data http://www.hes.gov.uk).

Non-alcoholic fatty liver disease (NAFLD) is emerging as one of the commonest causes of abnormal liver function tests and in the western world the estimated prevalence is reported to be as high as 30 %[64-66]. The prevalence of NAFLD is expected to rise in developed countries given the epidemic of its major underlying determinant- obesity, in addition to the increasing ascertainment of this condition. Recent estimates predict over half of the UK population will be obese by 2050[67]. Estimates of the prevalence of steatosis are 10-20% of the general population, and 75% of those with obesity.

2.6.3 Risk Factors for CLD

The three major causes of CLD in the UK (and western and developed world) are CHC (main risk factor since the introduction of screening for blood products is injecting drug use (IDU)) ALD (main risk factor is chronic heavy alcohol consumption), and NAFLD (main risk factor is rising incidence of childhood obesity and increased prevalence of obese adults and diabetes type II).

All of these risk factors are increasing especially amongst the younger age groups and the present epidemic of CLD is likely to get worse over the coming decades. There is a continuing incident population of at-risk injecting drug users as about one third of young injecting drug users share needles and equipment exposing themselves to infection with Hepatitis C.(Unlinked Anonymous Prevalence and Monitoring Programme (UAPMP)[68]
(Figure 2.9 (a-b).

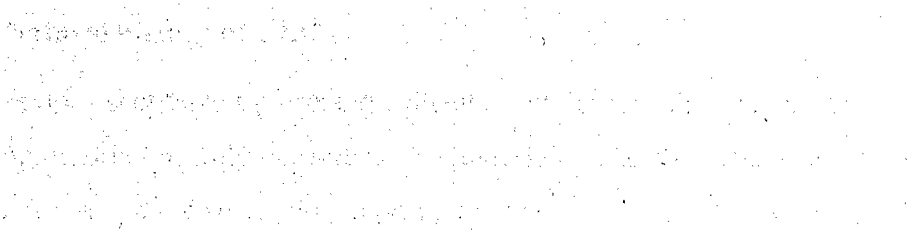
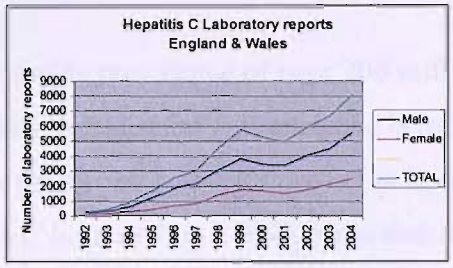
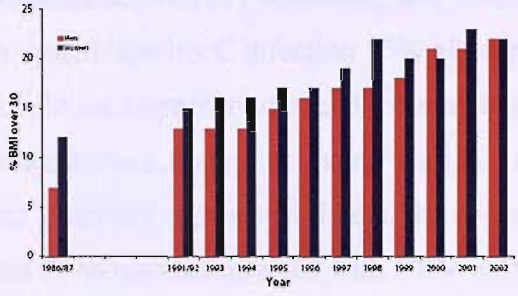


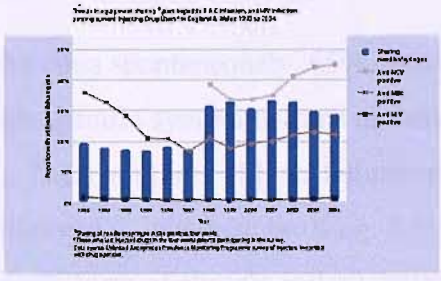
Figure 2.9 Trends in risk factors for CLD



(Health Protection Agency 2005)



(British Heart Foundation 2004)



(UAPMP Annual report 2004)

Trends in risk factors for CLD are all increasing- (a) Alcohol consumption (b) Laboratory reports of Hepatitis C (c) Obesity prevalence (d) injecting drug users sharing needles

There is therefore a potential for a marked increase in chronic liver disease (CLD) in the coming decades. Patients with complications of severe CLD such as varices and hepatic failure present complex medical problems for the health service, with expensive treatments and long stays including use of intensive care. The need and demand for services including transplant and the economic and societal burden, is likely to increase in the next two decades due to the rise in risk factors over the past 30-40 years.

2.6.4 Natural history of CLD

The previous section has presented evidence that the prevalence of CLD in the UK population is rising. In order to fully understand the underlying causes of this increase and to predict future trends in the disease, it is important to know how each of the causes of CLD develops

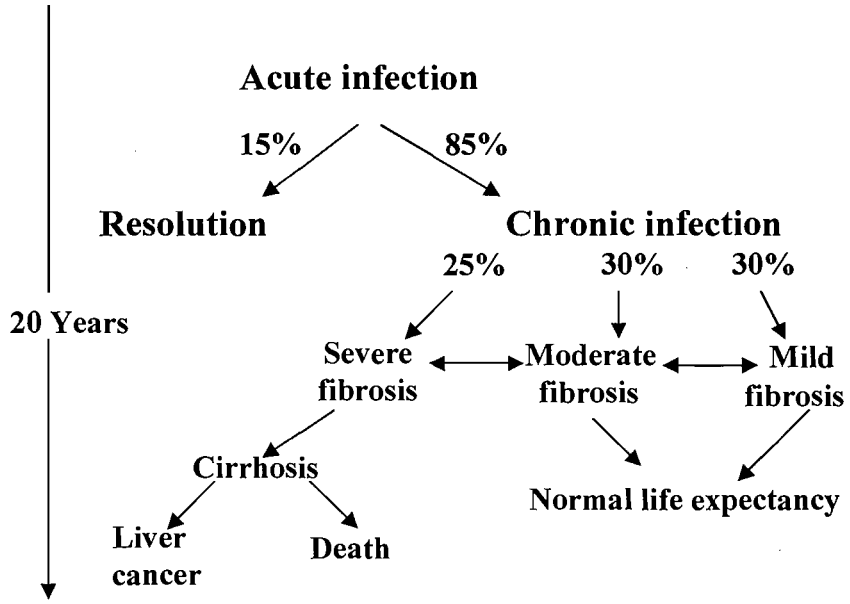
over time. The next section will provide an overview of the natural history of the major causes of CLD - CHC, ALD, and NAFLD.

Natural history of CHC

CHC is a major healthcare problem with a worldwide prevalence of over 200 million, 66 million of whom are at risk of developing serious liver disease[69]. Prevalence is increasing, and estimates of the future burden of CHC in many developed countries predict at least a three to five fold rise in cirrhosis by 2020[70-72]. Hepatitis C is caused by a RNA virus that is transmitted via blood. Common routes of infection include injecting drug use, transfusion of blood products before 1991 screening, and tattooing.

In acute Hepatitis C infection 15% of people clear the virus spontaneously. 65% of infected people are asymptomatic, and the remainder have non-specific symptoms such as fatigue, poor concentration, abdominal pain, "flu-like" symptoms. The symptoms of liver failure are rare and generally appear very late in the course of the disease e.g. jaundice, bruising. It is thought that 85 % patients infected with HCV will develop some form of chronic inflammation of the liver. Of these patients, about 20 % will develop cirrhosis of the liver after about 20 years of infection[73;74] (figure 2.10). When acquired in older age, the disease may progress more rapidly. Ingestion of alcohol has been clearly associated with an increased rate of hepatic inflammation[75]. CHC patients who drink alcohol excessively often have an acceleration of their disease. Diagnosis is confirmed by measuring virus in serum using polymerase chain reaction method. Treatment at present is with combination therapy with pegylated interferon and ribavirin for 24 or 48 weeks. Response is dependent on the genotype of the virus but the most favourable in Genotype 2 or 3 is 80-90% of people have a sustained response and clear the virus. In Genotype 1 40-50% have a sustained viral response (SVR). There have been recent changes in national guidelines following the publication of the cost effectiveness of treating mild disease[76;77]. It is now no longer necessary to demonstrate that patients have moderate fibrosis in order to offer anti-viral therapy[78].

Figure 2.10 Clinical Course of Hepatitis C Infection



Most people will develop chronic infection with Hepatitis C. Of these ~one third will progress to cirrhosis.

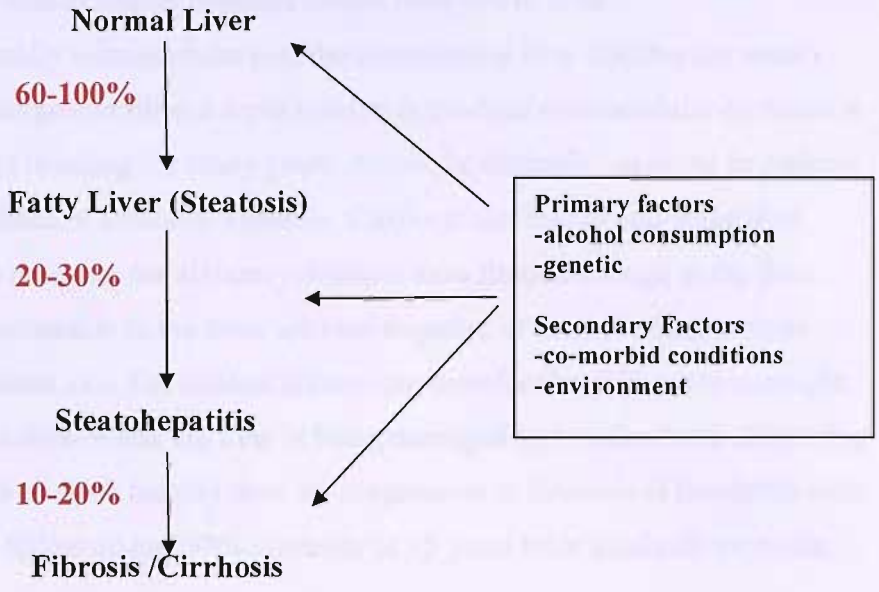
Natural history of ALD

About 15-20% of all heavy drinkers get CLD. Reasons for this are not clear but genetic and other environmental factors have been implicated.

ALD has three main pathological pictures:

Steatosis. This is the accumulation of fat within hepatocytes, and is reversible if the patient stops drinking. However, fatty liver can lead to steatohepatitis where the steatosis is accompanied by inflammation. This could progress to fibrosis and cirrhosis. (Figure 2.11) Patients with fatty liver are typically either asymptomatic or present with nonspecific symptoms that do not suggest acute liver disease. Laboratory tests are not diagnostic of fatty liver. Aminotransferases and alkaline phosphatase may be completely normal or only mildly deranged. With abstinence, morphologic changes of fatty liver usually revert to normal.

Figure 2.11 Progression of ALD in heavy drinkers



The majority of long term heavy drinkers will develop fatty liver, but only ~10-35% develop hepatitis, and ~8-20% will progress to cirrhosis. Genetic and environmental factors are thought to play role in ALD development[35].

Fatty liver has generally been considered to be a benign condition without risk of degeneration to a more ominous pathologic finding such as alcoholic hepatitis or fibrosis. However, although the short-term prognosis in patients with alcoholic steatosis is excellent, it has been found with longer follow-up that cirrhosis develops more commonly in alcohol abusers with fatty liver changes than in those with normal liver histology. Morphologic features predictive of progression to fibrosis and/or cirrhosis include severe steatosis, giant mitochondria, and the presence of mixed macro/microvesicular steatosis

Alcoholic Hepatitis can be acute or chronic and a typical clinical presentation is that of a chronic drinker who has had a recent episode of exceptionally heavy consumption of alcohol. Alcoholic hepatitis can range from a mild hepatitis, with abnormal laboratory tests being the only indication of disease, to severe liver dysfunction with complications such as jaundice, hepatic encephalopathy, ascites, bleeding esophageal varices (abnormal blood clotting and coma. Alcoholic hepatitis is reversible if the patient stops drinking, but it usually takes several months to resolve. Alcoholic hepatitis can lead to liver fibrosis and cirrhosis, and very

frequently occurs in alcoholics who already have cirrhosis of the liver or severe fibrosis. 30-day mortality in patients with alcoholic hepatitis ranges from 0% to 50%.

Alcoholic cirrhosis is usually micronodular (i.e. the regenerating liver nodules are small) although nodules may enlarge and fibrous septa resolve to produce macronodular cirrhosis in patients who have stopped drinking for many years. Alcoholic cirrhosis can occur in patients who have never had evidence of alcoholic hepatitis. Cirrhosis can lead to end-stage liver disease. It is important to note that not all heavy drinkers have fibrotic change in the liver, although they are more vulnerable to the other adverse sequelae of alcohol abuse such as societal, emotional, accidents etc. The clinical picture can therefore be difficult to untangle. The presence of fibrosis indicates that the liver is being damaged and studies have shown that if the individual continues to drink heavily then the progression to cirrhosis is inevitable with 71% mortality at 5 years follow up and 90% mortality at 15 years from alcoholic cirrhosis.

Long-term survival in patients with alcoholic hepatitis who discontinue alcohol is significantly better than in those who continue to drink, although it remains considerably below that of an age-matched population. Three-year survival approaches 90% in abstainers, whereas it is less than 70% in active drinkers. Morphologic changes of alcoholic hepatitis may be completely reversible in a small proportion of patients with strict abstinence. Although discontinuation of excessive alcohol consumption improves survival in patients who have alcoholic hepatitis with or without cirrhosis, even complete abstinence does not restore a normal life expectancy.

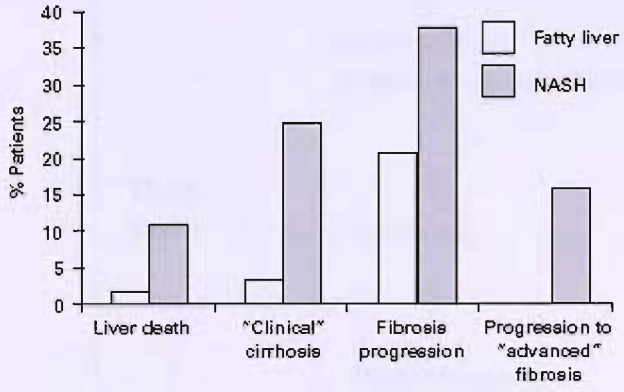
For those patients with decompensated alcoholic cirrhosis who undergo transplantation, survival is comparable to that of patients with other causes of liver disease. Acute alcoholic hepatitis, no matter how severe, is an absolute contraindication for liver transplantation. Most transplant centers currently require patients with a history of alcohol abuse to have documented abstinence of at least 6 months before undergoing transplantation. This requirement theoretically has the dual advantage of predicting long-term sobriety and allowing recovery of liver function from acute alcoholic hepatitis. This "6-month abstinence rule" may not have much prognostic significance in predicting recidivism, however. Alcohol use of any quantity after transplantation for alcohol-related liver disease approaches 50% during the first 5 years, and abuse occurs in up to 15% of patients.

NAFLD

Nonalcoholic fatty liver disease (NAFLD) is an increasingly recognized form of chronic liver disease. It encompasses a spectrum of conditions associated with lipid deposition in hepatocytes. It ranges from steatosis (simple fatty liver), to nonalcoholic steatohepatitis (NASH; fatty changes with inflammation and hepatocellular injury or fibrosis), to advanced fibrosis and cirrhosis. Studies suggest that although simple fatty liver is a benign condition, NASH may progress to fibrosis and lead to end-stage liver disease. The disease is mostly silent, and often discovered through incidentally elevated liver enzyme levels. It is strongly associated with obesity and insulin resistance and is currently considered by many as the hepatic component of the metabolic syndrome. The pathophysiology is not completely understood but accumulated triglycerides resulting from excess free fatty acids move into the liver. These then trigger oxidative stress, cytokine ("adipokines") and endotoxins which then leads to necroinflammation which can in turn move on to fibrosis. Adipokines include adiponectin, TNF-α and leptin. Adiponectin is anti-fibrotic and opposes fatty acid oxidation and reduced the TNF-α activity. Patients with fibrosis have been shown to have low adiponectin and high TNF α (which promotes insulin resistance and inflammation)[79]. Leptin is involved in the fibrogenesis pathways promoting fibrosis[80].

NASH cirrhosis is now one of the leading indications for liver transplantation in the United States. Because NAFLD resembles alcoholic liver disease but occurs in people who drink little or no alcohol, excessive daily alcohol consumption (more than 20 g/day in women and 30 g/day in men within the last 5 years) must be ruled out before making the diagnosis (350 ml [12 oz] of beer, 120 ml [4oz] of wine and 45 ml [1.5 oz] of spirits each contain 10 g of alcohol).

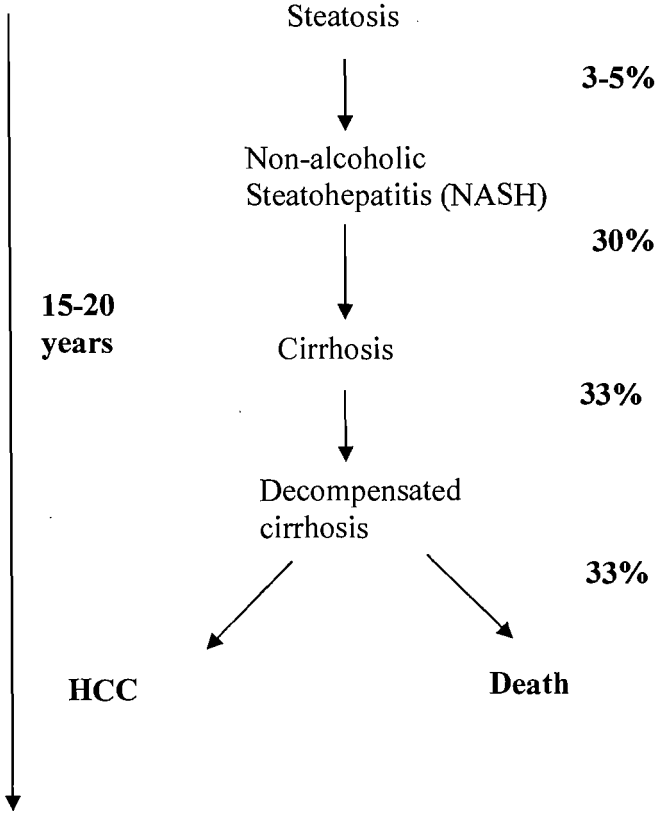
Figure 2.12 Natural history of Non-alcoholic fatty liver and NASH. Follow up 1-15 years[81]



Most of the advanced fibrosis occurs in those patients with NASH, although steatosis is not always a benign condition.

The natural history of NAFLD/NASH is not well defined, but it does seem that it is varied, some patients progress at varying rates to cirrhosis, some remain stable at the same histological stage and grade and some have regression of disease or exist in variations of the above e.g. isolated portal fibrosis (Figure 2.12). The long term prognosis depends on the stage of disease at presentation. In one study 25% of NASH with or without fibrosis on the index biopsy developed clinical evidence of cirrhosis with 11-33% dying a liver related death[82]. Recent studies indicate that NASH can result in the development of fibrosis in up to 30% of patients[83;84]. Once cirrhosis develops prognosis is poor with outcomes with Child Pugh Class B and C similar to those reported for CHC (details of clinical prognostic scales are reported in appendix 2). However cardiovascular mortality may be greater in cirrhotic patients with NASH[85;86] . Currently most people regard simple steatosis as a relatively benign condition, although NASH may develop and this latter group need to be more carefully monitored as progression of fibrosis is more common. Some studies have estimated that ~1-2 % of those with steatosis may progress to cirrhosis (Figure 2.13).

Figure 2.13 Progression of disease in NAFLD



A small proportion of people with simple steatosis progress to NASH and a proportion of these people will progress to cirrhosis and clinical outcomes.

2.6.5 Conclusion

The natural history of each cause of CLD varies depending on whether the original insult is infective, metabolic or toxic. Not all those people subject to the insult develop CLD although exact mechanisms for this remain unclear in all three causes.

The development of cirrhosis takes about 10-20 years in all cases. Once cirrhosis has developed the clinical course appears to be broadly similar with signs of clinical decompensation the same whatever the original insult. The mortality for each CLD cause may be complicated by cause specific non- liver related reasons, such as accidents for alcoholics in ALD, overdose and self harming for injecting drug users in CHC, and increased cardiovascular deaths for obese/metabolic syndrome patients in NAFLD.

CHAPTER 3
SYSTEMATIC REVIEWS OF SERUM MARKERS IN ASSESSING LIVER FIBROSIS

In this chapter the evidence is presented describing the diagnostic performance of non-invasive markers of liver fibrosis in the three main chronic liver diseases (CHC, NAFLD/NASH, and ALD). Three systematic reviews were conducted using a common methodology. Any differences in the methodology relating to the cause of CLD studied have been detailed. Search strategies are presented in Appendix 2. The results for each systematic review are reported separately, and a common discussion is presented for all three reviews.

3.1 Introduction

Chapter 2 has shown that liver fibrosis may be used as a surrogate for clinical outcome. It also identified limitations of using liver biopsy as a reference standard for liver fibrosis. Clinicians and patients require accurate information about the degree of liver fibrosis to assess disease severity, in order to guide management decisions, predict outcome and monitor disease. Serum markers of liver fibrosis offer an attractive alternative to liver biopsy, as they are less invasive, may allow dynamic calibration of fibrosis, and can be more cost effective. They can be divided into "indirect" markers that reflect disordered liver function or structural disruption (Alanine amino transferase, Aspartate amino transferase, bilirubin, platelets), and "direct" markers that are molecules associated with the synthesis and degradation of extracellular matrix (e.g. tissue inhibitor of matrix metalloproteinase-1(TIMP-1) and Hyaluronic Acid (HA).

Evidence of the diagnostic performance of serum markers of liver fibrosis in the major causes of CLD (previously introduced in Chapter 1) is needed to assess the clinical utility and effectiveness of such tests in the diagnosis, prognosis and management of liver disease. In order to provide such evidence, three systematic reviews were conducted to locate, collate, appraise and analyse studies that evaluated the performance of non invasive tests in the diagnosis of liver fibrosis in each of CHC, NAFLD and ALD.

3.2. Systematic review of the diagnostic performance of non-invasive markers of liver fibrosis in Chronic Hepatitis C

A narrative systematic review of the performance of single or multiple surrogate markers, compared to histology, in assessing fibrosis in CHC was conducted up to 2002[87]. The diagnostic accuracy appeared greatest and most promising in the studies using a panel of serum markers. We have built on this work and have performed systematic reviews to assess the diagnostic performance of **panels of serum markers** of hepatic fibrosis in CHC, incorporating analyses which place such markers in a clinical context.

3.2.1 Methods (common to all three systematic reviews)

The three reviews were conducted following accepted principles[88]. A systematic literature search was performed to ascertain the diagnostic performance of non invasive markers of liver fibrosis.

Sources searched included:

- Electronic databases
    - CHC 1985 – October 2004: Cochrane Library 2004
    - ALD 1980- October 2005 Cochrane Library 2005
    - NASH/ NAFLD 1996 – October 2005: Cochrane Library 2005

MEDLINE, EMBASE were searched using a search strategy derived from the literature[89;90]. Search terms were added following initial searches as appropriate.

- Relevant websites: American Association for the Study of the Liver, European Association for the Study of the Liver, Digestive Disease Week for conference proceedings or abstracts (2002-2004).

- Reference lists from relevant articles.

- Experts on diagnostic test reviews were consulted (Professor Jonathan Deeks and Ms Jacqueline Dinnes)

Inclusion criteria

Studies were included if they;

- were systematic reviews, meta-analyses or primary studies of diagnostic tests
- were written in English
- used liver biopsy as a reference standard.
- included >30 participants (as smaller studies will be underpowered to produce precise estimates of test performance and would be more likely to produce zero denominator effects in a 2x2 table. Confidence intervals would be very wide and inclusion in SROC where studies are unweighted may result in skewed unreliable results).
- evaluated panels of ≥ 2 serum markers (CHC)
- allowed extraction of data for interferon naïve patients (CHC)
- recorded alcohol consumption of subjects (NAFL/NASH)
- separated data according to the cause of liver disease.

Exclusion criteria

Studies were excluded if;

- data on disease-specific cause were not separately extractable
- the study did not produce a composite serum marker fibrosis score (CHC)
- data on fibrosis stage(s) were not extractable (NAFLD/NASH)
- data were presented only in abstract form.

A serum marker was defined as any measure that could be derived from a blood sample. Studies identified by the search strategy were assessed for inclusion by two reviewers (JP and Dr Neil Guha). Both reviewers read each and all abstracts retrieved by the searches.

Data extraction strategy

Data extraction was undertaken by one reviewer and checked by a second reviewer with any disagreements being resolved through discussion. A third reviewer was consulted to resolve persisting issues. Information collected included patient demographics, test assay details; background prevalence of fibrosis severity, risk factors, histological parameters, statistical methods used, and test performance characteristics.

Quality assessment strategy

The quality of included studies was assessed using the quality assessment of diagnostic accuracy studies (QADAS) tool[91]. (Appendix 3)

Data analysis/synthesis

Data are presented with full tabulation of results of included studies.

Where data were available, 2x2 tables were constructed to derive sensitivity, specificity, predictive values, likelihood ratios (LR) and diagnostic odds ratios (DOR) at each threshold value. (Accepted levels for robust tests are - LR = <0.1, and +LR = >10, >5 and <0.2 give strong diagnostic evidence. For DOR reasonable test performances would be >30) (14). The percentage of patients in each study to which the different thresholds could be applied were derived where possible. We evaluated the performance of tests at thresholds which produced clinically useful predictive values with acceptable disease specific false negative and positive rates, based on local clinical opinion (NPV ≥95% PPV ≥90%). This could then allow liver biopsy to be avoided appropriately, with patients below NPV of <95% assumed to have no significant fibrosis and patients with PPV >90% assumed to have significant fibrosis. The performance of tests was calculated for different fibrosis stages- early versus moderate/severe fibrosis (F0/F1 vs. F2/F3/F4 and also for cirrhosis or no cirrhosis (F0/F1/F2/F3/vs F4). Summative statistics were used to combine results where appropriate, including summative receiver operator characteristic curves (SROC) (Moses Littenberg method) [92; 93] using STATA version 11 package. This method uses a logistic transformation and linear regression to produce a summary receiver operator characteristic curve extending the logic of meta analysis to diagnostic testing. The SROC analysis involves three steps: (1) the pairs of True positive and false positive estimates from each study are transformed onto a suitable scale of log odds; (2) a linear regression equation is fitted using the transformed data; and (3) the coefficients from the linear regression model are used to generate a curve in the original ROC space. The area under the curve (AUC) (in this case, being the area under the SROC curve) presents an overall summary of test performance and displays the trade-off between sensitivity and specificity. An AUC of 1.0 (100%) indicates perfect discriminatory ability of the diagnostic test. In addition, the Q index is another useful global estimate of test accuracy for comparing SROC curves. The Q*index, defined by the point where sensitivity equals specificity on the

SROC curve, is the point on the SROC curve that is intersected by the anti-diagonal. A Q value of 1.0 indicates 100% accuracy (sensitivity and specificity of 100%)

SROC curves were derived for all tests and all thresholds as performed in previous analyses[94], though available software for SROC does not allow for varying numbers of thresholds per study, and also with available data at three thresholds (one per study) low, mid-point, and high (Figures 1a to 1c)[9558] . A summary value of DOR, sensitivity and specificity were calculated for the SROC. Sub group analyses by method of recruitment, quality of biopsy and scoring system were undertaken.

T-test for independent samples was used to compare the effect of characteristics of the studies (such as methods of recruitment and histology classification) on the test performance results.

3.2.2. Results of CHC systematic review

Study characteristics

The electronic search yielded 2,766 abstracts which were read in full. Most were excluded as they did not evaluate serum markers, or the reference standard was not diagnosis of liver fibrosis on biopsy. This was expected as the search strategy was designed to have high sensitivity and low specificity. 25 full papers were retrieved of which 11 were excluded leaving 14 studies in separate populations to be included in the review (see Table 1). Reasons for exclusion were;

- single markers 4 papers
- less than 30 participants 4 papers
- no reference test 1 paper

In addition two reviews were identified; a systematic review of all serum markers and an overview of two markers (Fibrotest and Actitest[96]). All relevant studies fulfilling inclusion criteria have been included in this review. Primary data from several studies presented in the Fibrotest/Actitest review but not reported elsewhere were utilized in the summative analysis.

Table 3.1 Characteristics of studies evaluating the performance of panels of serum markers of liver fibrosis in CHC F0/1 vs F2/3/4

Author Year published (date of study) country (no. centres)	Total no. Patients in study	Serum marker tests in panel	Test used to compare	Validity sample (n)	Patient selection	Age mean (yrs)	% male	% severe fibrosis	% IDU / Alc*	Liver biopsy scoring system	Biopsy Mean Length (L) Portal Tracts (PT) Observers (O)
Imbert-Bismut 2001[97] (08/97-03/00) France (1)	339	AST, ALT, albumin, α_1 globulin, β globulin, γ globulin, bilirubin, GGT, α_2 macroglobulin, haptoglobulin, apolipoprotein A1.	n/r	Training set (n=205) Internal validation set (n=134)	*DOSVIRC cohort Prospective recruitment	47	58	40	n/r	METAVIR	>10 mm (L) n/s (PT) 1 (O)
Rossi 2003[98] (01/98-11/01) Australia (1)	125	bilirubin, GGT, α_2 macroglobulin, haptoglobulin, apolipoprotein A1 corrected for age and sex.(Fibrotest)	n/r	Whole study = external validation of test	Consecutive prospective recruitment	40	66	38	n/r	METAVIR	n/s (L) n/s (PT) 1 (O)
Poynard (2003) (03/98-[99]10/00) Europe Argentina Canada USA (62)	352	FT-AT (Fibrotest + ALT)	Fibrotest	Whole study = external validation of test	From RCT treatment (n=1530) Retrospective recruitment	45	64	17	n/r	METAVIR & Knodell	17 mm (L) >6 mm in 89 % (PT) 1 (O)

Author Year published (date of study) country (no. centres)	Total no. Patients in study	Serum marker tests in panel	Test used to compare	Validity sample (n)	Patient selection	Age mean (yrs)	% male	% severe fibrosis	% IDU / Alc*	Liver biopsy scoring system	Biopsy Mean Length (L) Portal Tracts (PT) Observers (O)
Wai 2003[100] (01/01-01/03) USA (1)	270	AST: platelet ratio (APRI)	n/r	Training set (192) Internal validation (78)	Retrospective recruitment	46	64	64	41	Ishak	n/s (L) n/s (PT) 1 (O)
Le Calvez (2004)[101] 08/97-03/00) France (1)	323	AST: platelet ratio (APRI)	Fibrotest	Whole study = external validation of test	*DOSVIRC cohort Retrospective recruitment	n/r	n/r	41	n/r	METAVIR	> 10 mm (L) n/s (PT) n/s (O)
Forns 2002[102] (07/96-12/99) Spain (1)	476	Age, GGT, cholesterol, platelets (Forns Index)	n/r	Training set (n=351) Internal validation (n=125)	Consecutive prospective recruitment	39	64	25	n/r	Scheuer	n/s (L) >6 (PT) 1 (O)
Thabut (2004)[103] (08/973/00) Fr (1)	249	Forns index score (Age, GGT, cholesterol, platelets)	Fibrotest	Whole study = external validation of test	*DOSVIRC cohort Retrospective recruitment	n/r	n/r	38	n/r	METAVIR	> 10 mm (L) n/s (PT) n/s (O)
Sud 2004[104] (05/99-08/02) Australia (2centres single city)	302	Fibrosis probability index =age, AST, HOMA-IR (fasting glucose* gluc/22.5), total cholesterol, alcohol consumption.	APRI	Training set (n=176) Internal validation set (n=126)	Consecutive Prospective recruitment	41	56	54	61	Scheuer	n/s (L) n/s (PT) n/s (O)

Author Year published (date of study) country (no. centres)	Total no. Patients in study	Serum marker tests in panel	Test used to compare	Validity sample (n)	Patient selection	Age mean (yrs)	% male	% severe fibrosis	% IDU / Alc*	Liver biopsy scoring system	Biopsy Mean Length (L) Portal Tracts (PT) Observers (O)
Leroy 2004[105] (1999-2000) France (1)	388	PIIINP/MMP-1/HA/MMP-2/MMP-9/TIMP-1/TIMP-2	PIIINP HA	Training set (n=194) Age matched controls(n=194)	Consecutive prospective recruitment	43	64	45	40	METAVIR	19 mm (L) 14 (PT) 1 (O)
El Shorbagy (2004)[106] 2000-2003 Egypt (1)	109	Platelets, MMP-9, portal vein diameter, splenic longitudinal axis, ALT, AST, viral load.	n/r	Training set (n=109)	CHC patients from general population screening	47	71	80	n/r	Local scoring system	n/s (L) n/s (PT) n/s (O)
Patel (2004)[107] (1992-2001) USA France (4)	696	HA, TIMP-1, α 2 macroglobulin	HA TIMP1, α2 macroglobulin	Training set (n=294) External validation set (n=492)	Retrospective selection made on equal numbers F0-F4	45	69	51	n/r	METAVIR	13 mm (L) >5 (PT) 1 (O)
Rosenberg 2004[30] (1998-2000) England, (8Italy (2) Sweden (1) Germany(2)	325	3 marker panel – age, HA, PIIINP, TIMP-1.	n/r	Training set (n=164) External validation set (n=261)	Prospective recruitment	44	63	27	n/r	Scheuer	>12 mm (L) >5 (PT) 1 (O)

Author Year published (date of study) country (no. centres)	Total no. Patients in study	Serum marker tests in panel	Test used to compare	Validity sample (n)	Patient selection	Age mean (yrs)	% male	% severe fibrosis	% IDU / Alc*	Liver biopsy scoring system	Biopsy Mean Length (L) Portal Tracts (PT) Observers (O)
F0 F1 F2 F3 vs F4											
Kaul 2002[108] (01/89-10/98) USA (2)	264	Probability model – Platelets, AST, sex, spider naevi.	n/r	Training set (264) External validation set (102)	Retrospective recruitment	n/r	45	61	33	Scheuer	n/s (L) n/s (PT) 1 (O)
Fortunato (2001)[109] Italy (1)	103	Fibronectin, prothrombin, pseudocholinesteras, ALT, manganese superoxide dismutase N-acetyl β - glucosaminidase.	n/r	Training set (63) Internal validation set (40)	Prospective recruitment	n/r	n/r	n/r	n/r	Desmet	n/s (L) n/s (PT) >2 (O)

Abbreviations GGT= γ-glutamyl- transpeptidase, AST=aspartate transaminase ALT alanine amino transferase HA= hyaluronic acid; TIMP-1= Tissue inhibitor of matrix metalloproteinase 1; PIIINP=Amino terminal peptide of procollagen III; MMP= matrix metalloproteinase -9; CAH=Chronic active hepatitis; n/r= not reported * DOSVIRC cohort =single centre cohort liver unit of Pitie-Salpetrie Hospital Paris * co morbidity with respect to % study population IDU/heavy alcohol consumer

The full QUADAS assessments are presented in Appendix 3. In general the studies were of reasonable quality with 12 studies reporting positively to more than ten of the QUADAS criteria, with similar criteria being met or unmet. Most studies met the criteria of all participants having both index and reference test, blinding of test evaluation, and independence of index and reference test. Withdrawals from the study were less well explained and six studies did not include a validation of the score. Overall the studies were of reasonable quality. Ten different panels of serum markers were reported. Ten studies reported sufficient information to derive sensitivity, specificity, predictive values, DOR and LR at specific cut-offs. Patient characteristics varied between studies, the median proportion of male subjects was 64% (range 45-71%) and (where reported) the average age of subjects ranged between 39-47 yrs. Only four studies presented CHC risk factors. The proportion with moderate/severe fibrosis (F2 F3 F4) was 43% (median) with a range of 17-80%.

Histological Staging

The fibrosis staging systems used to classify the histology varied, METAVIR (7), Schemer (4), Ishak (1), Knodell (1), Desmet (1) and local scoring system (1). Sub-group analyses found that there was no significant difference in the AUC results (p=0.6) depending on which of two commonest fibrosis staging systems -METAVIR (6 studies) and Scheuer (5 studies) - was used.

Liver biopsy Size

Quality of liver biopsy as assessed by number of portal tracts and length of biopsy was reported in 7 studies, with 3 having both these criteria. There was no difference in AUC results between those reporting adequate (>15mm length (6) or > 5 portal tracts (4) and inadequate samples (p=0.6). Study design was similar in most studies, with paired histology and serum samples on individual patients with untreated CHC being analysed retrospectively from an existing cohort (n=7), or prospectively recruited (n=8) and then analysed at a single point in an individual's illness.

In 6 studies recruitment was consecutive. Analyses correcting for the differences in recruitment (prospective vs. retrospective) showed no difference in results (p=0.8). 11 studies presented data validated in a different group of patients than the training set including five

studies of patients recruited at the same centre as the training cohort (internal validation) and six studies that recruited subjects at a different centre (external validation).

Results from studies differentiating (F0/F1 vs. F2/F3/F4)

For those studies presenting data on mild versus moderate/severe fibrosis, 10 studies presented data at several thresholds and presented sufficient information to permit the derivation of true positive, true negative, false positive, and false negative rates (Table 2). The number of thresholds presented for each test varied.

AUC for Receiver Operating Characteristic (ROC) curves were presented in 13/14 studies. For most studies the AUC varied between training and validation sets with performance generally being lower in validation than in training (Table 3.2). There was variation in the performance of the same tests using an identical threshold in different populations (e.g. Specificity of Fibrotest at threshold of 0.3 varied in three studies from 45-61%)

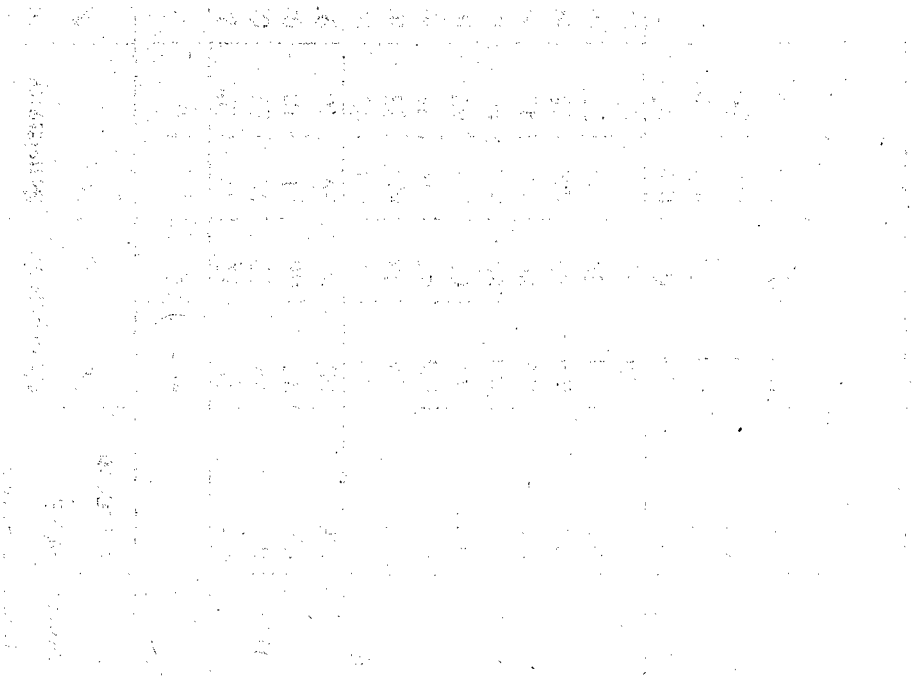


Table 3.2 Results of studies evaluating the performance of panels of serum markers of liver fibrosis in CHC F0/1 vs. F2/3/4

Study and year of publication	Cut off levels reported	Cumulative %		Sensitivity %		Specificity %		PPV %		NPV %		+ VE LR		-VE LR		AUC	
FIBROTEST		T	V	T	V	T	V	T	V	T	V	T	V	T	V	T	V
Imbert-Bismut(2001) ***	0.1	16	12	97	100	24	22	44	51	93	100	1.3	1.3	0.1	0		
	0.3	48	39	79	87	65	59	58	63	84	85	2.3	2.1	0.3	0.2	0.84	0.87
	0.6	77	66	51	70	94	95	84	91	76	80	8.6	12.9	0.5	0.3		
	0.8	86	81	29	38	95	97	78	92	69	66	5.7	14.2	0.7	0.6		
Rossi (2003)	0.1	n/a	21	n/a	92	n/a	29	n/a	45	n/a	85	n/a	1.3	n/a	0.3		
	0.2	n/a	38	n/a	83	n/a	52	n/a	52	n/a	83	n/a	1.7	n/a	0.3		
	0.3	n/a	47	n/a	75	n/a	61	n/a	54	n/a	80	n/a	1.9	n/a	0.4		
	0.4	n/a	61	n/a	67	n/a	78	n/a	65	n/a	79	n/a	3	n/a	0.4		
	0.5	n/a	69	n/a	56	n/a	85	n/a	70	n/a	76	n/a	3.6	n/a	0.5	n/a	0.74
	0.6	n/a	80	n/a	42	n/a	94	n/a	78	n/a	72	n/a	6.4	n/a	0.6		
	0.7	n/a	82	n/a	35	n/a	94	n/a	77	n/a	70	n/a	5.4	n/a	0.7		
	0.8	n/a	89	n/a	22	n/a	96	n/a	79	n/a	66	n/a	5.8	n/a	0.8		
	0.9	n/a	95	n/a	8	n/a	97	n/a	57	n/a	63	n/a	3.2	n/a	0.9		
Poynard (2003) (Before Treatment)	0.1	n/a	6	n/a	97	n/a	80	n/a	41	n/a	81	n/a	1.1	n/a	0.4		
	0.3	n/a	33	n/a	86	n/a	45	n/a	50	n/a	83	n/a	1.6	n/a	0.3	n/a	0.73
	0.6	n/a	67	n/a	50	n/a	79	n/a	61	n/a	71	n/a	2.4	n/a	0.6		
	0.8	n/a	89	n/a	20	n/a	95	n/a	72	n/a	65	n/a	3.9	n/a	0.8		
APRI																	
Wai (2003)	0.5	29	n/a	91	n/a	47	n/a	61	n/a	86	n/a	1.7	n/a	0.2	n/a	0.8	0.9
	1.5	78	n/a	41	n/a	95	n/a	88	n/a	64	n/a	8.2	n/a	0.6	n/a		

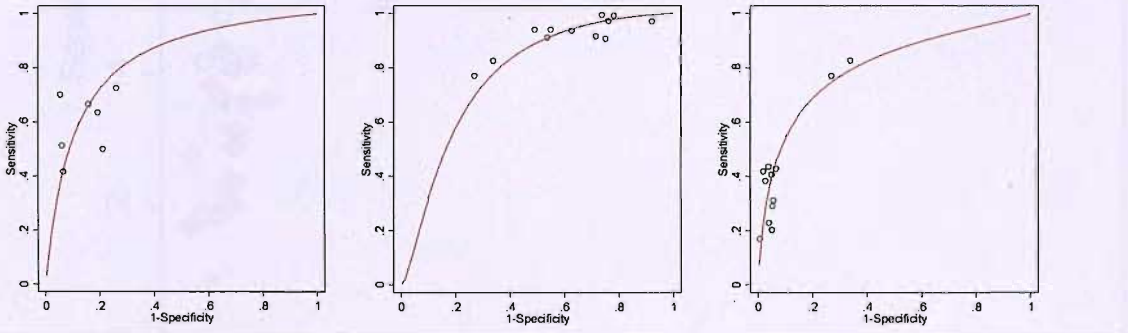
Study and year of publication	Cut off levels reported	Cumulative %		Sensitivity %		Specificity %		PPV %		NPV %		+ VE LR		-VE LR		AUC	
		T*	V**	T	V	T	V	T	V	T	V	T	V	T	V	T	V
Le Calvez (2004) (APRI VAL) ***	0.5 1.0 1.5 2	n/a n/a n/a n/a	41 68 80 87	n/a n/a n/a n/a	81 54 36 24	n/a n/a n/a n/a	56 84 91 95	n/a n/a n/a n/a	56 70 73 76	n/a n/a n/a n/a	81 73 68 65	n/a n/a n/a n/a	1.8 3.3 4.1 4.5	n/a n/a n/a n/a	0.3 0.5 0.7 0.8	n/a	0.74
Sud (2004)	n/r	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.76
		T	V	T	V	T	V	T	V	T	V	T	V	T	V	T	V
FORNS																	
Forns (2002)	4.2 6.9	36 87	39 88	94 44	94 30	45 96	51 95	35 79	40 66	96 84	96 80	1.7 11.6	1.9 6.0	0.1 0.6	0.1 0.7	0.84	0.77
Thabut (2004) (FORNS VAL) ***	1 3 6 8	n/a n/a n/a n/a	2 16 70 91	n/a n/a n/a n/a	1 1 55 19	n/a n/a n/a n/a	4 26 86 97	n/a n/a n/a n/a	39 45 70 78	n/a n/a n/a n/a	1 1 75 66	n/a n/a n/a n/a	1 1.4 3.8 5.8	n/a n/a n/a n/a	0 0 0.5 0.8	n/a	0.78
Sud (2004)	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	0.76
Sud (2004)	0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9	14 24 32 42 52 57 72 76 87	16 29 40 48 55 60 70 75 88	100 96 93 87 73 58 49 43 27	91 85 74 68 64 58 49 42 19	26 43 54 69 74 83 91 94 100	25 48 60 69 81 85 96 98 98	56 61 65 72 72 76 83 87 100	63 70 73 76 83 84 95 97 93	100 93 89 85 75 68 66 64 60	65 69 62 60 61 58 57 54 46	1.4 1.7 2 2.8 2.8 3.4 5.6 6.6 E	1.2 1.6 1.8 2.2 3.3 3.8 12.6 21.8 9.8	0 0.1 0.1 0.2 0.4 0.5 0.6 0.6 0.7	0.4 0.3 0.4 0.5 0.5 0.5 0.5 0.6 0.8	0.84	0.77
Leroy (2004)	0.2 0.3 0.4 0.5	24 61 82 93	n/a n/a n/a n/a	91 65 35 17	n/a n/a n/a n/a	35 85 96 99	n/a n/a n/a n/a	55 76 91 100	n/a n/a n/a n/a	88 75 65 60	n/a n/a n/a n/a	1.5 4.3 8.6 33.1	n/a n/a n/a n/a	0.2 0.4 0.7 0.8	n/a n/a n/a n/a	0.82	n/a

Study and year of publication	Cut off levels reported	Cumulative %		Sensitivity %		Specificity %		PPV %		NPV %		+ VE LR		-VE LR		AUC	
El Shorbagy 2004	0-3 4-6 6-9	32 ^55 ^17 ^	n/a n/a n/a	82 69 80	n/a n/a n/a	80 67 97	n/a n/a n/a	51 77 84	n/a n/a n/a	95 57 96	n/a n/a n/a	4.2 2.1 23.7	n/a n/a n/a	0.2 0.5 0.2	n/a n/a n/a	0.8	n/a
Patel (2004)	0.36	41	47	83	77	66	73	72	76	79	75	2.4	2.9	0.3	0.3		
		T	V	T	V	T	V	T	V	T	V	T	V	T	V	T	V
ELF Rosenberg (2004)	0.063 0.067 0.09 0.126 0.190 0.219 0.268 0.426 0.564			n/a n/a n/a n/a n/a n/a n/a n/a n/a	95 90 85 80 63 52 47 38 30	n/a n/a n/a n/a n/a n/a n/a n/a n/a	29 31 43 58 80 85 90 95 99	n/a n/a n/a n/a n/a n/a n/a n/a n/a	28 28 30 35 48 50 58 70 90	n/a n/a n/a n/a n/a n/a n/a n/a n/a	95 92 91 91 89 86 86 84 83	n/a n/a n/a n/a n/a n/a n/a n/a n/a		n/a n/a n/a n/a n/a n/a n/a n/a n/a			0.77

***** from Comparative Hepatology Poynard T. et al 2004 * T=training population **V=validation population**

Median AUC for training sets was 0.81 (range 0.80 to 0.84), and for the validation sets 0.77 (0.73 to 0.90). Likelihood Ratios (LR) and diagnostic odds ratios (DOR) were derived for 10 studies. Negative LR ranged from 0.1 to 0.9 (– LR), and positive LR 1.2 to 33.1 (+LR). DOR was 9.0 (median) with a range of 5 to 27. The cumulative percentage of patients to whom the panel score was applicable at each cut-off is presented (Table 2). The proportion of people at thresholds where the PPV ≥ 90% and NPV ≥ 95% was 40% (training set), 29% (validation set), overall 35% (median values).Clearly this value will rise if one lowers the predictive value used.

Figure 3.1: Summative Receiver Operator Characteristic Curves



1a: mid range thresholds	1b: lowest reported	1c: highest reported thresholds

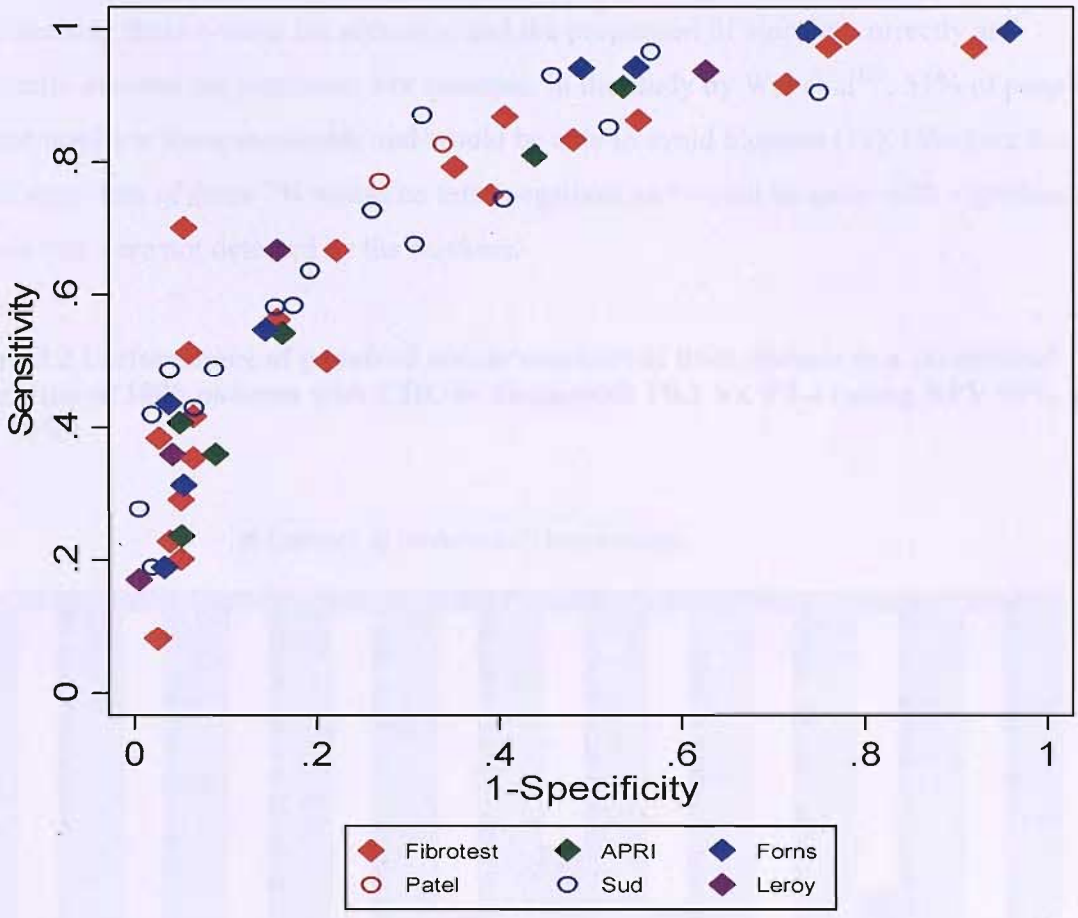
Summative statistics

DOR = 6.52 (1.69-25.23)	DOR = 6.39 (1.89-21.65)	DOR = 9.96 (8.15-12)
Sensitivity = 59.8%	Sensitivity = 94.8%	Sensitivity = 40.1%
Specificity = 87.7%	Specificity = 35.8%	Specificity = 95%

Figures 1a to 1c show SROC curves of panels of serum markers combined at high, mid and low thresholds showing summative diagnostic odds ratios, sensitivity and specificity

At each threshold tests perform with either high sensitivity with low specificity or vice-versa (see Figures 1a to 1d). The summative DORs at low, mid and high thresholds are all <10. None of the panels showed a statistically significant difference in performance from each other.

Figure 3.1d: Summative Receiver Operator characteristic curve for all panel serum markers at all thresholds



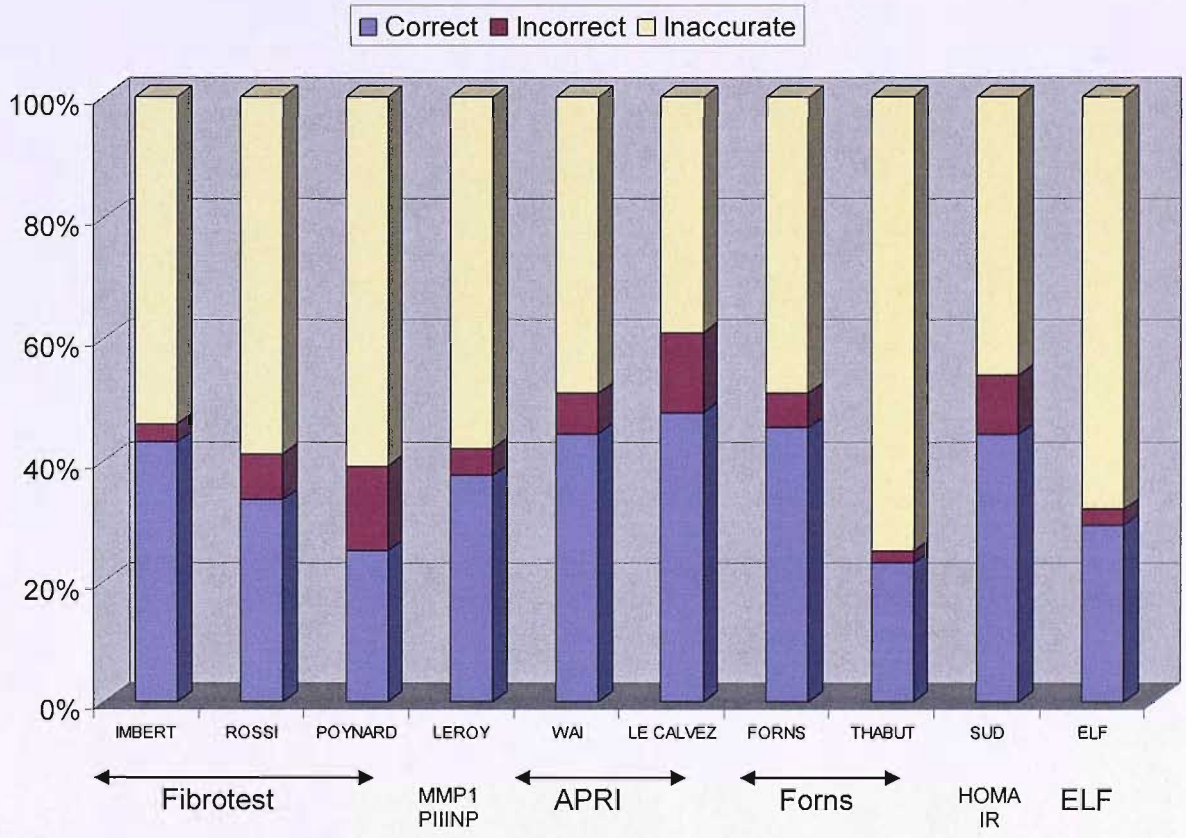
DOR	9.96 (8.15-12)
Specificity	71%
Sensitivity	80%

Test for heterogeneity p=<0.0001

Figure 1d shows SROC curve of all panels of serum markers at all thresholds summative diagnostic odds ratios, sensitivity and specificity

Figure 3.2 shows the results of modelling where thresholds giving PPV of approximately 90% and NPV of approximately 95% are used in an attempt to avoid liver biopsy in a theoretical population of 1,000 patients with CHC (i.e. with a 10% false positive rate and a 5% false negative rate). The percentage of that population in whom the use of markers would generate a result meeting these criteria for accuracy, and the proportion of biopsies correctly and incorrectly avoided are presented. For example, in the study by Wai et al[100], 51% of people had test results at these thresholds and would be able to avoid biopsies (19). However this would mean that of these 7% would be false negatives and would be cases with significant fibrosis that were not detected by the markers.

Figure 3.2 Performance of panels of serum markers of liver fibrosis in a theoretical population of 1000 patients with CHC to distinguish F0,1 Vs. F2-4 (using NPV 95% and PPV 90%)



Utility model showing for each study the percentage of correctly allocated (blue) false test (purple collar) and those tests that cannot be allocated (cream)

Results from studies differentiating (F0/F1/F2/F3 vs. F4) in CHC

In studies that reported results for F0, 1, 2, 3 vs. F4 (i.e. no cirrhosis versus cirrhosis) all of the surrogate markers performed at a higher level, with the AUC and sensitivity and specificity being greater at all thresholds. (Table 3.3)

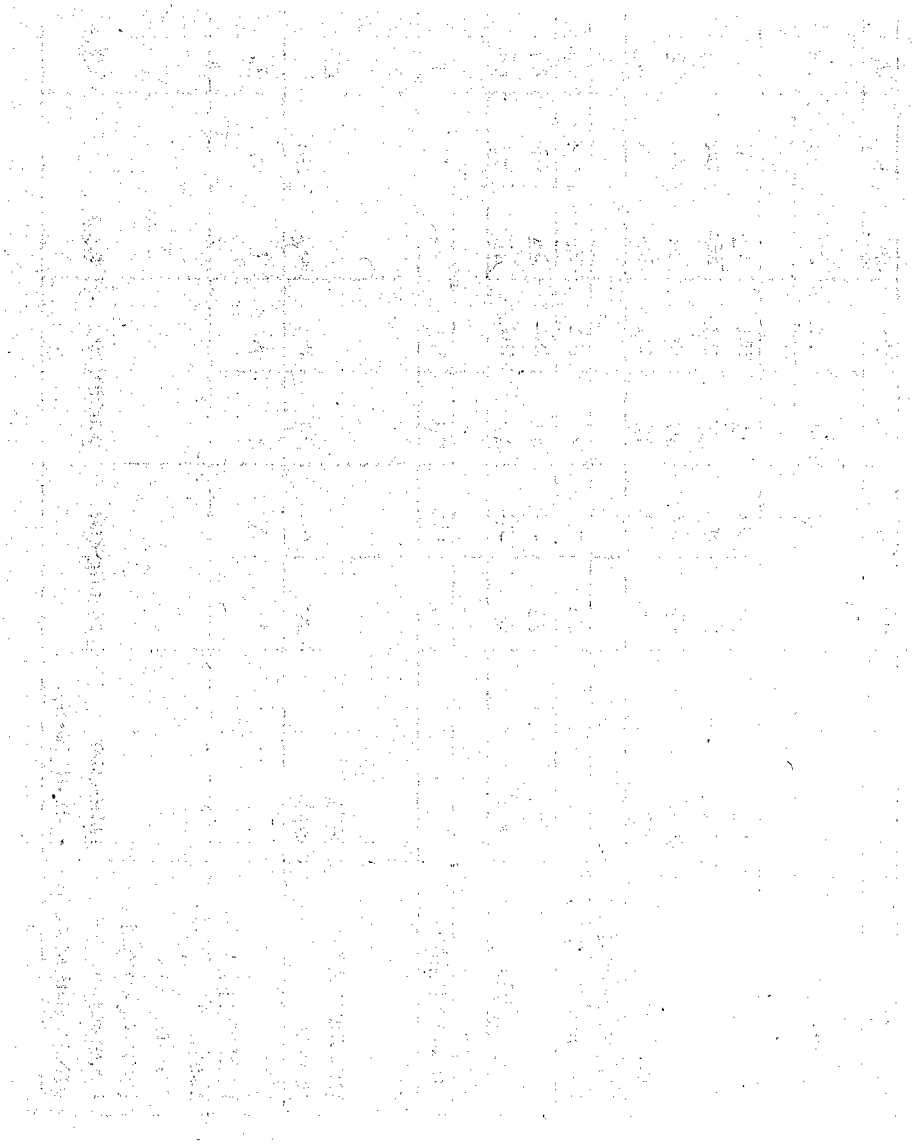


Table 3.3 Results of studies evaluating the performance of panels of serum markers of liver fibrosis in CHC F0 F1 F2 F3 vs F4 (cirrhosis /not cirrhosis)

Study and year of publication (date of study) country (no centres)	Cut off levels reported	Sensitivity		Specificity		PPV		NPV		+ VE LR		-VE LR		AUC	
FIBROTEST		T	V	T	V	T	V	T	V	T	V	T	V	T	V
Imbert-Bismut(2001) ***	<0.8 >0.8	n/r	n/r	n/r	n/r	85	n/r	90	n/r	n/r	n/r	n/r	n/r	0.92	n/r
Poynard (2003)		n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	0.73
APRI															
Wai (2003)	<0.1 <0.2	89 57	n/r n/r	75 93	n/r n/r	38 57	35 65	98 93	100 95	n/r n/r	n/r n/r	n/r n/r	n/r n/r	0.89	0.94
Le Calvez (2004)		n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	0.80
PIIINP MMP1															
Leroy (2004)	0.2 0.3 0.4 0.5	94 85 58 26	n/r n/r n/r n/r	28 74 98 97	n/r n/r n/r n/r	n/r 43 66 77	n/r n/r n/r n/r	95 95 91	n/r n/r n/r n/r	n/r n/r n/r n/r	n/r n/r n/r n/r	n/r n/r n/r n/r	n/r n/r n/r n/r	0,88	
El Shorbagy (2004)		n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	0.8	
Kaul (2002)		n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	0.93	0.93

3.3 Systematic review of the diagnostic performance of non-invasive markers of liver fibrosis in Non-alcoholic fatty liver disease

The methodology differed from that used for CHC in that the search included studies of the diagnostic performance of all non-invasive markers including single serum markers as well as panels of markers, and other methods such as ultrasound. This was because no previous systematic review has been performed to differentiate the performance of single and multiple markers and because other modalities of non-invasive testing have been investigated in NAFLD/NASH.

3.3.1. Results of systematic review of NAFLD

Study characteristics

The electronic search yielded 1,781 abstracts which were read in full. Most were excluded as they did not evaluate serum markers in diagnosis of liver fibrosis on biopsy. This was expected as the search strategy was designed to have high sensitivity and low specificity. 47 full papers were retrieved of which 18 were excluded leaving 29 studies in separate populations to be included in the review.

The demographics of patients included in the final analysis are shown in Table 3.4 .The prevalence of severe fibrosis (grade 3-4) ranged from 9 % to 43 % with a median of 22.5%. The range of mean BMI in the studies was 26 to 60 (median 31); 5 studies recruited from patients undergoing bariatric surgery. The cut-off for alcohol consumption varied amongst studies but the majority excluded patients consuming > 200 g/week. Only 7 studies included details of length of biopsy specimen or number of portal tracts.

Three studies produced a diagnostic algorithm in association with specificities, sensitivities, predictive values and/or area under the receiving operator curve statistics (AUC). The remaining studies investigated the association of individual variables with severe fibrosis vs. moderate fibrosis (17 studies), moderate fibrosis vs. mild fibrosis (4 studies), any fibrosis vs. no fibrosis (7 studies) and no fibrosis vs. moderate fibrosis (1 study).

Table 3.4 Characteristics of studies evaluating non-invasive markers of fibrosis in NAFLD

Study Author Year of publication Date of Study Country	Total No. Patient	Patient selection	Prevalence of Steatosis (S) Inflammation (I) Fibrosis (F)	Age Mean (medn)	% male	BMI Mean (medn)	Alcohol	% diabetes/ HT	Liver biopsy Score Length (L) Portal tract (PT) Observers (O)	Variables measured
Angulo 1999[50] USA	144	NAFLD on biopsy and persistently abnormal LFTS for more than 3 months. Prospective and retrospective recruitment	73 % grade 2-3 (S) 27 % significant fibrosis (F3/4)	(50.5)	33	31.2	<40 g/week	28 % diabetes	Modified Brunt L (n/s) PT (n/s) O (n/s)	Age AST/ ALT >1 ALT Albumin Transferrin saturation Diabetes
Rosenberg 2004[30] Europe	61	NAFLD on biopsy and abnormal LFTS for 6 months. Prospective recruitment	27 % significant fibrosis (F3/4)	44	63	n/s	n/s	n/s	Scheuer L (>12 mm) PT (>5) O (3)	Age HA P IIINP TIMP-1
Sakaguwa 2005[110] Japan	112	NAFLD on biopsy	63 % NASH 43 % significant fibrosis (F3/4)	51	32	29	<30 g/d	30 % diabetes	Modified Brunt L (n/s) PT (n/s) O (2)	Female Platelets Albumin GGT AST/ALT HA, Type IV collagen
Albano 2005[111] UK	167 (NAFLD) 59 (controls)	NAFLD on biopsy. Case controlled: NAFLD vs. controls Prospective consecutive recruitment	44 % NASH 17 % significant fibrosis (F3/4)	55	61	35	<20g/d	29 % diabetes	Modified Brunt L (n/s) PT (n/s) O (1)	Age AST/ALT >1 Diabetes Malondialdehyde (MDA)

Study Author Year of publication Date of Study Country	Total No. Patient	Patient selection	Prevalence of Steatosis (S) Inflammation (I) Fibrosis (F)	Age Mean (medn)	% male	BMI Mean (medn)	Alcohol	% diabetes/ HT	Liver biopsy Score Length (L) Portal tract (PT) Observers (O)	Variables measured
Mofrad 2003[112] USA	51	NAFLD on biopsy with normal ALT	72 % grade 2-3 (S) 36 % severe fibrosis (F3/4)	53	31	29	<20 g/d	57 % diabetes 47 % hypertension	Modified Brunt L (n/s) PT (n/s) O (1)	Diabetes
Shimada 2002[113] Japan	81	NASH on biopsy Prospective recruitment	82 % grade 2/3 (S) 100 % NASH 28 % severe fibrosis (F3/4)	(54)	49	(26)	<20 g/week	31 % diabetes	Brunt L (n/s) PT (n/s) O (1)	Age Platelet count, AST/ALT >1, Albumin, Bilirubin, ferritin, platelets, IgA, PT, type IV collagen, raised lipids
Dixon 2001[114] Australia	105	Patients undergoing laparoscopic banding and liver biopsy with BMI >35. Prospective consecutive recruitment	25 % NASH. 10 % severe fibrosis (F3/4)	41	21	47	<200 g/week	18 % diabetes 39 % hypertension	Brunt L (n/s) PT (>6) O (1)	Male Diabetes Hypertension ALT C peptide
Beymer 2003[115] USA	48	BMI >35 undergoing gastric bypass surgery and liver biopsy Prospective consecutive recruitment	64 % grade 2/3 (S) 33 % NASH 12 % severe fibrosis (F3/4)	42	31	60	<20 g/mth	19 % diabetes	Ishak L (n/s) PT (n/s) O (1)	Diabetes

Study Author Year of publication Date of Study Country	Total No. Patient	Patient selection	Prevalence of Steatosis (S) Inflammation (I) Fibrosis (F)	Age Mean (medn)	% male	BMI Mean (medn)	Alcohol	% diabetes/ HT	Liver biopsy Score Length (L) Portal tract (PT) Observers (O)	Variables measured
Bugianesi 2004[116] Italy	167	Raised transaminases (>6 months) and bright liver on U/S and NAFLD on biopsy. Prospective recruitment	47 % grade 2/3 (S) 21 % severe fibrosis (F3/4)	41	83	28	<20 g/d	8 % diabetes	Modified Brunt L (n/s) PT (n/s) O (n/s)	Age, female, BMI, AST/ALT Ferritin OGIS, 1/QUICKI HOMA-IR
Dixon 2003[117] Australia	105	Patients with BMI >35 undergoing laparoscopic banding and liver biopsy Prospective recruitment	34 % NASH 14 % severe fibrosis (F 3/4)	42	26	>35	<200 g/week	n/s	Brunt L (n/s) PT (>6) O (1)	ALT HOMA IR Polymorphisms in transforming growth (TGF) factor and angio-tensinogen (AT)
Hui 2004[79] Australia	109 (NAFLD) 82 (controls)	Patients referred with abnormal LFTS or hepatic steatosis on U/S and NAFLD on biopsy. Controls matched by age and BMI. CCS /prospective	50 % grade 2/3 (S) 73 % NASH 28 % sever fibrosis (F3/4)	48	63	30	<40 g/week	32 % diabetes in NAFLD group	Brunt L (n/s) PT (n/s) O (1)	Age HOMA-IR
Guidorizzi 2005[118] Brazil	64	Patients with NAFLD on biopsy. Prospective recruitment	84 % NASH 11 % severe fibrosis (F3/4)	45	78	28	<20 g/day	11 % diabetes 27 % hypertens ion	Brunt L (n/s) PT (n/s) O (1)	HOMA-IR

Study Author Year of publication Date of Study Country	**Total No. Patients**	**Patient selection**	**Prevalence of : Steatosis (S) Inflammation (I) Fibrosis (F)**	**Age Mean (median)**	**% male**	**BMI Mean (med)**	**Alcohol**	**% diabetes or hypertension**	**Liver biopsy Score Length (L) Portal tract (PT) Observers (O)**	**Non-invasive variables measured**
Suzuki 2005[119] USA	79	Patients with abnormal LFTs for three months and NAFLD on liver biopsy Prospective and consecutive recruitment	25 % severe fibrosis (F3/4)	46	38	33	<40 g/week	n/s	Brunt L (>15 mm) PT (n/s) O (1)	Age Serum albumin Platelet count Fasting blood glucose Hyaluronic acid Clinical diagnostic score
Angulo 2004[120] USA	88	Patients with abnormal LFTS, NAFLD on biopsy and participants in previous trials. Retrospective recruitment	77 % grade 2-3 (S) 83 % NASH 22 % severe fibrosis (F3/4)	45	35	33	<140 g/week	19 % diabetes	Brunt L (>15 mm) PT (n/s) O (1)	Age Female BMI Diabetes Leptin QUICKI HOMA IR
Marchesini 2003[121] Italy	163	Patients with abnormal LFTS for 3 mth + NAFLD on liver bx Prospective consecutive recruitment	74 % NASH 21 % severe fibrosis (F3/4)	40	88	28	<140 g/week	67 % hypertension	Brunt L (n/s) PT (n/s) O (n/s)	Metabolic syndrome
Hashimoto 2005[122] Japan	247	Patients with NAFLD on liver biopsy Prospective recruitment	36 % severe fibrosis (F3/4)	(53)	53	67 % with BMI>28	<100 g/week	33 % diabetes 46 % hypertension	Local score	AgeSex AST/ALT Albumin, Platelets Diabetes, hyaluronic acid and type IV collagen

Study Author Year of publication Date of Study Country	Total No. Patient	Patient selection	Prevalence of Steatosis (S) Inflammation (I) Fibrosis (F)	Age Mean (medn)	% male	BMI Mean (medn)	Alcohol	% diabetes/ HT	Liver biopsy Score Length (L) Portal tract (PT) Observers (O)	Variables measured
Ong 2005[123] USA	212	Patients undergoing bariatric surgery with BMI >40 and obesity related complications. Prospective recruitment	24 % NASH 8 % advanced fibrosis	42	20	48	<10 g/day	24 % diabetes	Local score L (n/s) PT (n/s) O (1)	Waist to hip ratio (WHR) AST ALT Diabetes HT
Ledinghen 2004[124]	67	Chronically elevated ALT for six months and liver biopsy Retrospective recruitment	40 % NASH 31 % F2/3/4 fibrosis	47	67	26	<40 g/day	n/s	Metavir L (n/s) PT (n/s) O (1)	BMI AST ALT Ferritin
Ratziu 2000 France[125]	93	BMI >25, abnormal LFTS and NASH on liver biopsy. Retrospective consecutive recruitment	30 % F2/3/4 fibrosis	49	34	29	30 g/d	16 % diabetes	METAVIR L (n/s) PT (n/s) O (1)	Age BMI ALT Diabetes Triglycerides
Sorrentino 2004[126] Italy	80	Undergoing liver biopsy for operative procedure(gall stones, large bowel or gastric cancer) + metabolic syndrome + high grade obesity + normal LFTS Prospective recruitment	53 % grade 2/3 (S) 73 % NASH 23 % severe fibrosis (F3/4)	58	38	39	<30 g/day	45 % diabetes 78 % hypertension	Brunt L (>8 mm) PT (n/s) O (2)	Female BMI >45 Duration of obesity Metabolic syndrome

Study Author Year of publication Date of Study Country	Total No. Patients	Patient selection	Prevalence of : Steatosis (S) Inflammation (I) Fibrosis (F)	Age Mean (median)	% male	BMI Mean (med)	Alcohol	% diabetes or hypertension	Liver biopsy Score Length (L) Portal tract (PT) Observers (O)	Non-invasive variables measured
[127] Crespo 2001[127] Spain	181	Patients undergoing bariatric surgery and liver biopsy Prospective recruitment	72 % grade 2/3 (S) 23 % F2/3/4 fibrosis	n/s	16	47	<30 g/d	n/s	Modified METAVIR L (n/s) PT (n/s) O (1)	Age at liver biopsy Elevated blood sugar level
Fierbinteanu - Braticevici 2002[128] Romania	80	Abnormal LFTS and fatty liver on U/S and undergoing liver biopsy Retrospective recruitment	26 % NASH	51	25	32	<200 g/week	n/s	Local score L (n/s) PT (n/s) O (1)	Age BMI >30 ALT >3 N Ferritin Triglycerides (TG) MDA Glutathione (GSH)
Loguerico 2004[129] Italy	305	Abnormal ALT for 12 months and NAFLD on liver biopsy Prospective recruitment	68 % grade2/3 Moderate/ severe pericellular fibrosis	n/a	82	70 % were >25	<20 g/d	n/s	Local score L (n/s) PT (n/s) O (3)	Ferritin HOMA IR
Santos 2005[130] Brazil	30	BMI >25 + U/S diagnosis of steatosis + raised LFTs and undergoing liver biopsy Prospective recruitment	Fibrosis present in 37 %	45	60	31	<20 g/day	23 % diabetes	Modified Brunt L (n/s) PT (n/s) O (n/s)	AST Laminin HA Collagen IV

Study Author Year of publication Date of Study Country	Total No. Patient	Patient selection	Prevalence of Steatosis (S) Inflammation (I) Fibrosis (F)	Age Mean (medn)	% male	BMI Mean (medn)	Alcohol	% diabetes/ HT	Liver biopsy Score Length (L) Portal tract (PT) Observers (O)	Variables measured
Yesilova 2005[131] Turkey	51 (NAFLD) 30 (controls)	Raised LFTS for six months and NAFLD on liver biopsy Prospective recruitment	60 % grade2/3 (S) 88 % NASH 10 % severe fibrosis (F3/4)	36	100	28	<20 g/day	0 % diabetes	Brunt L (n/s) PT (n/s) O (n/s)	HOMA-IR Co enzyme Q10 (CoQ10) Copper zinc oxide dismutase (CuZnSOD)
Koruk 2003[132]	36 (NASH) 32 (controls)	Steatosis on U/S, abnormal LFTs for three months and NASH on liver biopsy	67 % Grade2/3 (S) 100 % NASH 0 % severe fibrosis (F3/4)	44	75	(29)	absent	20 % diabetes	Modified Brunt L (n/s) PT (n/s) O (n/s)	Triglycerides LDL cholesterol Apoprotein A1 (Apo A1)
Hartleb 2005[133]	47	Patients with NAFLD on liver biopsy and ALT > 1.5 ULN Retrospective study	50 % Grade 2/3 (S) 65 % NASH 20 % some fibrosis	45	57	29	<120 g/week	13 % diabetes	Local L (n/s) PT (>5) O (2)	Age BMI Diabetes Hypertension
Chitturi 2002[80] Australia	94	NASH Case-controlled – prospective and retrospective	70 % Grade 2/3 (S) 45 % significant fibrosis (F3/4)	51	57	31	<20 g/d	47 % diabetes	Modified Brunt L (n/s) PT (n/s) O (1)	None
Brunt 2004[134] USA	30	Subjects in NASH treatment trial. Retrospective	43 % grade1-4 fibrosis	45	46	34	<20 g/day	25 % diabetes	Brunt and Metavir L (n/s) PT (n/s) O (1)	AST/ALT ratio Albumin

HT=hypertension medn=median value

Variables associated with fibrosis (F0-2 vs. 3,4)

The variables associated with fibrosis can be subdivided into five groups: socio-demographic and anthropometric, simple liver biochemistry and haematology, features of metabolic syndrome and glucose sensitivity, fibrosis markers and miscellaneous markers as illustrated in Table 3.5. The association of these variables with the different stages of fibrosis is shown in Table 3.6.

Table 3.5. Variables associated with fibrosis

Category	Variable
Socio-demographic and anthropometric	Age, Gender, BMI, waist to hip ratio (WHR)
Simple liver biochemistry and haematology	ALT, AST, AST/ALT ratio, platelets, bilirubin, ferritin, transferrin sat, albumin.
Features of metabolic syndrome or glucose sensitivity	Diabetes, Hypertension, Homeostatic insulin resistance (HOMA-IR), Oral glucose sensitivity index (OGIS), metabolic syndrome, raised triglycerides, Quantitative insulin sensitivity check index (QUICKI), adiponectin, leptin, hyperlipidaemia
Fibrosis markers	Hyaluronic acid (HA), tissue inhibitor of metalloproteases 1 (TIMP 1), laminin, type IV collagen, aminoterminal peptide of procollagen III (PIIINP).
Miscellaneous	Malondialdehyde, C peptide, polymorphisms of transforming growth factor and angiotensinogen, IgA, glutathione, arachidonic acid, oxidised cardiolipin, Co enzyme Q and copper oxide dismutase.

Table 3.6 The association of non-invasive markers with fibrosis stage in NAFLD

F0/1/2 vs. F3/F4	**Age (increased)**	**Diabetes (present)**	**BMI (increased)**	**AST/ALT ratio (increased)**	**HOMA-IR (increased)**	**Platelets (decreased)**	**HA (increased)**	**Miscellaneous (association with fibrosis)**
Angulo[50]	Yes* UVA and MVA	Yes UVA and MVA	Yes UVA	Yes UVA and MVA	Not tested	Not tested	Not tested	Obesity at UVA and MVA. ALT, transferrin sat and Albumin at UVA.
Rosenberg[30]	Yes UVA, MVA and ROC	Not tested	Not tested	Not tested	Not tested	Not tested	Yes UVA, MVA and ROC	PIIINP and TIMP1 also included in discriminant score
Sakugawa[110]	Yes UVA	No**	No	Yes UVA	No	Yes UVA	Yes UVA, MVA and ROC	Female, platelets GGT and albumin on UVA. Type IV collagen at UVA, MVA and ROC.
Albano[111]	Yes UVA	Yes UVA and MVA	No	Yes UVA and MVA	Not tested	Not tested	Not tested	MDA abs UVA and MVA
Mofrad[112]	No	Yes UVA and MVA	No	Not tested	Not tested	Not tested	Not tested	
Shimada[113]	Yes UVA and MVA	Yes UVA	No	Yes UVA and MVA	Not tested	Yes UVA and MVA	Yes UVA and MVA	Albumin, Bilirubin, Ferritin, IgA, Hyperlipidaemia, Type IV collagen and IgA on UVA. Platelet count on UVA/MVA.

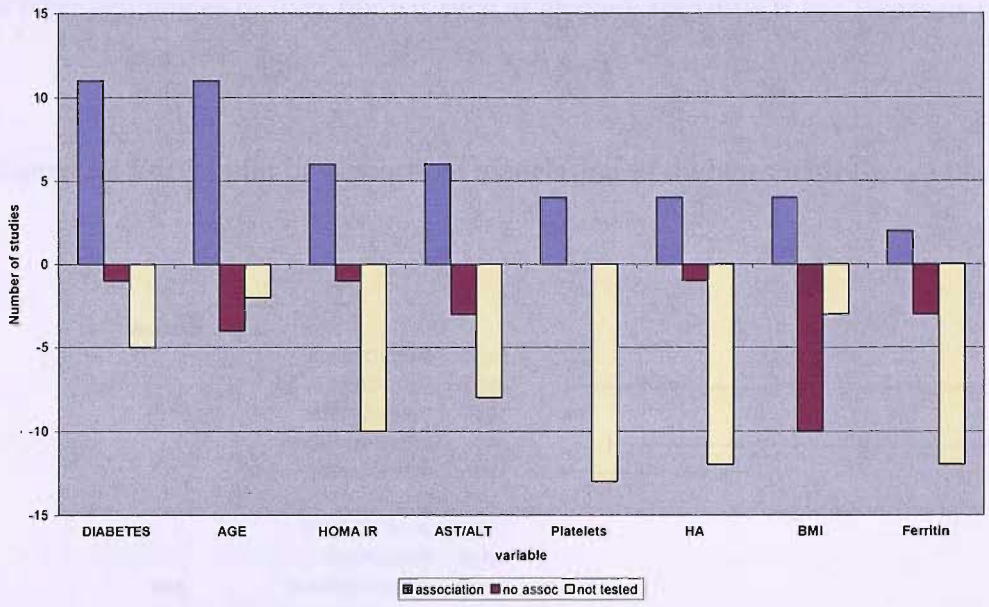
F0/1/2 vs. F3/F4	Age (increased)	Diabetes (present)	BMI (increased)	AST/ALT ratio (increased)	HOMA-IR (increased)	Platelets (decreased)	HA (increased)	Miscellaneous (association with fibrosis)
Dixon[114]	No	Yes UVA	No	No	Yes UVA	Not tested	Not tested	Hypertension , raised C peptide and ALT by MVA
Beymer[115]	No	Yes MVA	No	Not tested	Not tested	Not tested	Not tested	
Bugianesi[116]	Yes UVA	Yes (fasting glucose) UVA	Yes UVA	Yes UVA	Yes UVA	Not tested	Not tested	Female sex, 100/ISI, I/QUICKI, ferritin, OGIS at UVA
Dixon[117]	Yes UVA	Not tested	Yes UVA	No	Yes UVA and MVA	Not tested	Not tested	Raised ALT and combination of high risk phenotypes of polymorphisms (TGF b and AT) on UVA and MVA
Hui[79]	Yes UVA	Not tested	Not tested	Not tested	Yes UVA and MVA	Not tested	Not tested	
Guidorizzi[118]	Not tested	Not tested	Not tested	Not tested	Yes UVA	Not tested	Not tested	
Suzuki[119]	Yes UVA	Yes (fasting glucose) UVA & ROC (clinical diagnostic model)	No	Yes ROC (clinical diagnostic model)	Not tested	Yes UVA	Yes UVA and MVA.	Serum albumin and platelet count at UVA. Ferritin, Clinical diagnostic model (age, diabetes, AST/ALT, obesity) at ROC.
Angulo[120]	Yes UVA and MVA	Yes UVA	Yes UVA	Not tested	Yes UVA	Not tested	Not tested	Leptin and female at UVA QUICKI at UVA & MVA

F0/1/2 vs. F3/F4	**Age (increased)**	**Diabetes (present)**	**BMI (increased)**	**AST/ALT ratio (increased)**	**HOMA-IR (increased)**	**Platelets (decreased)**	**HA (increased)**	**Miscellaneous (association with fibrosis)**
Marchesini[121]	Not tested	Not tested	Not tested	Not tested	Not tested	Not tested	Not tested	Metabolic syndrome by MVA
Hashimoto[122]	Yes UVA	Yes UVA	No	Yes UVA	Not tested	Yes UVA	Yes UVA and MVA	Gender, hypertension, platelet count, albumin, type IV collagen at UVA. Billirubin at MVA.
Ong[123]	No	Yes UVA and MVA	No	No	Not tested	Not tested	Not tested	Raised AST, ALT and WHR on MVA.
F0/1 vs. F2/3/4	Age (increased)	Diabetes (present)	BMI (increased)	AST/ALT ratio (increased)	HOMA-IR (increased)	Platelets (decreased)	HA (increased)	Miscellaneous (association with fibrosis)
Ledinghen[124]	No	Not tested	Yes (BMI >25) UVA	Yes (Raised ALT) UVA	Not tested	No	Not tested	Ferritin at UVA
F0/1 vs. F2/3/4	**Age (increased)**	**Diabetes (present)**	**BMI (increased)**	**AST/ALTratio (increased)**	**HOMA-IR (increased)**	**Platelets (decreased)**	**HA (increased)**	**Miscellaneous (association with fibrosis)**
Ratziu[125]	Yes UVA and MVA	Yes UVA	Yes (BMI >28) UVA MVA	No	Not tested	Not tested	Not tested	BAAT score (BMI, Age, ALT.,TGs) by MVA and ROC
Sorrentino[126]	No	Yes (with metabolic syndrome) MVA	Yes BMI >45 MVA	Not tested	Not tested	Not tested	Not tested	Female sex and duration of obesity MVA
Crespo[127]	Yes UVA and MVA	No	No	Not tested	Not tested	Not tested	Not tested	Raised blood glucose at UVA

F0 vs. F1/2/3/4	**Age (increased)**	**Diabetes (present)**	**BMI (increased)**	**AST/ALT ratio (increased)**	**HOMA-IR (increased)**	**Platelets (decreased)**	**HA (increased)**	**Miscellaneous (association with fibrosis)**
Fierbinteanu[128]	Yes UVA MVA	Not tested	Yes UVA MVA	Not tested	Not tested	Not tested	Not tested	Raised ALT, Ferritin,MDA, GSH and TGs at UVA and MVA. No stats on score (BAMFAGT)
Loguerico[129]	No	Not tested	No	Not tested	Yes UVA	Not tested	Not tested	Ferritin at UVA
Santos[130]	No	Not tested	No	No	Not tested	Not tested	Yes UVA	Laminin, AST and collagen IV UVA
Yesilova[131]	Not tested	No	No	Not tested	Yes Positive correlation	Not tested	Not tested	CoQ10 and CuZnSOD negative correlation
Koruk[132]	Not tested	Not tested	Not tested	Not tested	Not tested	Not tested	Not tested	Raised TGS, LDL shoed positive correlation and Apo A1showed negative correlation
Hartleb[133]	Yes UVA	Yes UVA	Yes UVA	Not tested	Not tested	Not tested	Not tested	HT at UVA
Chitturi[80]	No	No	No	Not tested	Not tested	Not tested	Not tested	
F0 vs. F2/3	**Age (increased)**	**Diabetes (present)**	**BMI (increased)**	**AST/ALT ratio (increased)**	**HOMA-IR (increased)**	**Platelets (decreased)**	**HA (increased)**	**Miscellaneous (association with fibrosis)**
Brunt[134]	No	No	No	Yes UVA	No	Not tested	Not tested	Serum albumin reduced in severe disease

* YES= association at univariate analysis, correlation or multivariate analysis P<0.05 ** NO = no association at univariate or multivariate analysis

Figure 3.3: Variables associated with severe fibrosis

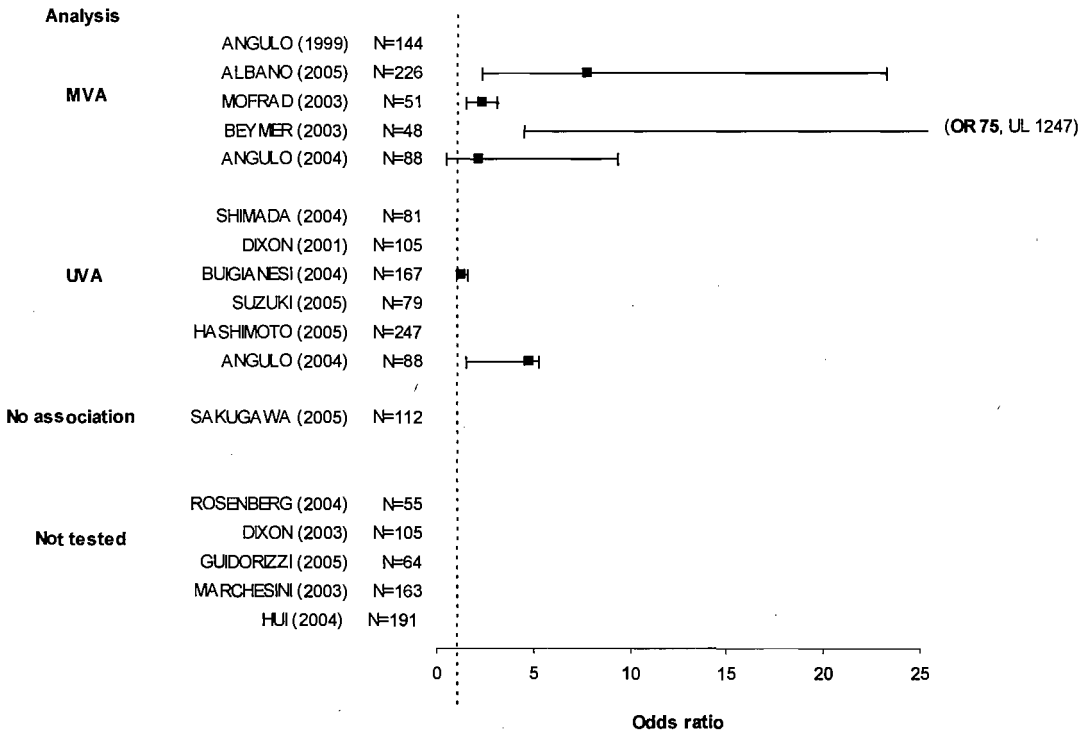


Plot of number of studies which show an association of variables with severe fibrosis -+association (blue column); no association (purple) not tested (cream)

The variables most commonly associated with fibrosis were: presence of diabetes, increasing age, increased HOMA-IR, increased AST/ALT ratio, decreased platelets, hyaluronic acid and BMI (Figure 3.3); each has biological plausibility. In NAFLD, age at biopsy is a reflection of probable duration of exposure to risk (e.g. to obesity or insulin resistance) and there is emerging evidence that the fibrotic response itself may be more exaggerated with increasing age and a similar phenomenon is seen in the context of hepatitis C. The variables diabetes, HOMA-IR, QUICKI and OGIS all reflect insulin resistance which has a fundamental role in the development and progression of fibrosis within NAFLD. The mechanisms by which insulin resistance triggers fibrosis may be through, free fatty acid mobilisation, generation of reactive oxygen species and production of fibrogenic growth factors [135;136]. The AST/ALT ratio has been shown to be elevated in a variety of diseases causing fibrosis and cirrhosis and this may be related to a reduced sinusoidal clearance of AST relative to ALT. Reduction in the peripheral platelet count may be due to splenic sequestration due to splenomegaly resulting from portal hypertension, but also some chronic liver diseases may reduce the hormone thrombopoietin which stimulates platelet production. In NAFLD, unsurprisingly, it has been shown to be a better indicator of severe fibrosis/cirrhosis rather than the earlier stages of fibrosis (see table 3.6). Finally hyaluronic acid may increase in fibrosis due to a mixture of

increased collagen turnover and reduced hepatic clearance and this has been shown to increase in other aetiologies of liver fibrosis such as alcohol, Hepatitis B and Hepatitis C.

Figure 3.4 Forest plot of strength of association of diabetes with severe fibrosis



Panels of markers for the detection of NAFLD

Very few studies were designed as traditional diagnostic studies, comprising comparisons of diagnostic tests with reference standards. The majority have concentrated on finding statistical associations of variables with fibrosis to try and elucidate the mechanisms of NAFLD rather than producing diagnostic algorithms. This is in contrast to hepatitis C where panel marker tests have combined variables found to be significant at multivariate analysis in a mathematical algorithm. As the identification of variables precedes formulation of an algorithm this suggests that non-invasive markers are generally at an earlier stage of development for NAFLD. The three studies producing a panel marker diagnostic test with AUCs and cut-offs with relevant specificities and sensitivities included the BAAT score, HA score and ELF score. Only one of these studies included a validation cohort and the number of patients in these studies was relatively small; two studies compared F3/4 vs. F0/1/2 and the other compared F2/3/4 vs. F0/1. The AUC ranged form 0.84 to 0.92, see Table 3.7.

Table 3.7: Panel marker tests measuring fibrosis in NAFLD

Test	Components of panel	Fibrosis stage	Training or validation	No.	AUC	Cut-off	Sens	Spec	PPV	NPV
HA score [119]	Age >45, obesity, AST/ALT ratio >1, diabetes, Hyaluronic acid	F3/4 vs. F0/1/2	Training	79	0.92	N/S	N/S	N/S	N/S	N/S
ELF score[30]	Age, hyaluronic acid, TIMP-1, PIIINP	F3/4 vs. F0/1/2	Validation	61	0.87	0.37 0.46	89 78	96 98	80 87	98 96
BAAT score [125]	Age, BMI, ALT, serum triglycerides	F2/3/4 vs. F0/1	Training	93	0.84	0 1 2 3 4	100 100 71 14 0	11 47 80 100 100	33 45 61 100 0	100 100 86 73 70

3.4 Systematic review of the diagnostic performance of non-invasive markers of liver fibrosis in Alcoholic Liver Disease

Methods

Methodology differed from the review of CHC in that the performance of **single and panel serum markers** of liver fibrosis were evaluated. Because the ALD literature precedes that for CHC (HCV was first identified in 1989) the searches were conducted from 1980-2005.

3.4.1 Results of systematic review of diagnostic tests in ALD

Study characteristics

The electronic search yielded 436 abstracts which were read in full. Most were excluded as they did not evaluate serum markers in diagnosis of liver fibrosis on biopsy. This was expected as the search strategy was designed to have high sensitivity and low specificity. 34 full papers were retrieved of which 22 were excluded leaving 12 studies in separate populations to be included in the review (see Table 8).

Reasons for exclusion were (may be >1 /study);

- Not primary study (editorial/non systematic review) n=3
- Outcome was not fibrosis (usually alcoholic hepatitis) n=2
- Participants <30 n=1
- No results separable for ALD alone n=6
- No results reported as sensitivity, specificity, ROC curves, diagnostic accuracy n=11 (Most of these studies reported correlation coefficients/differences in means of serum markers between group with fibrosis and those with less fibrosis).
- No results for fibrosis alone separable from data that combined steatosis with fibrosis or fibrosis/cirrhosis with acute alcoholic hepatitis (AH).n=4

Neither systematic reviews nor meta-analyses were identified. Studies were conducted between 1989 and 2006. Study characteristics are shown in Table 3.8.

The median age of participants in included studies was 50 years (range 44-65 years), 77% were male (range 63-100%) and the median number of study participants was 109 (range 44-

1034). The median background prevalence of serious fibrosis/cirrhosis was 42% (14-100%). All of the studies were conducted in secondary/tertiary settings.

There was marked heterogeneity between the studies. Different scoring systems were used: METAVIR (or modified METAVIR) 4; Scheuer 1; Ishak 2; Knodell1; and locally generated 5 (mostly dividing fibrosis into mild, moderate or severe). 10/12 studies presented data that showed the performance of the markers in identifying cirrhosis, 3/12 studies reported information identifying cirrhosis/severe fibrosis (METAVIR stages 4 /3, 4) and 2/12 reported information identifying any fibrosis. All of the studies evaluated performance of markers using cross sectional data for paired samples of histology and serum. All studies recruited prospectively, and half recruited consecutive patients. Although all participants were hospitalized there was heterogeneity of patient selection- both the inclusion criteria and daily alcohol consumption. Inclusion criteria reported were patients with previously diagnosed ALD, and or "alcoholism" or heavy alcohol consumption or patients admitted rehabilitation/detoxification/alcohol withdrawal symptoms. The daily consumption of alcohol (where reported) varied with 1 study recruiting patients drinking >100g of alcohol/day, 3 studies >80g, and 5 studies >50g. Inclusion criteria adopted a varied number of years drinking at these levels (range 5-10 years) reported (See Table 3.8). Two studies used the same patient population, with the earlier study reporting results from 109 patients with compensated ALD recruited in 1994-95 and the later study adding further patients from 1997-98 and reporting from the whole cohort (n=240).

Table 3.8 Characteristics of studies evaluating serum markers in Alcoholic Liver Disease

Study Author: Yr published (date of study when reported) Country *(No. centres)*	**Total no patients**	**Patient selection** **Recruitment details (where reported)**	**Alcohol consumption inclusion criteria**	**% cirrhosis (significant fibrosis*)**	**Age Yr mean (SD)**	**% male**	**Liver biopsy scoring system**	**Serum marker or panel**
Gabrielli 1989[137] Italy *(1)*	44	Patients with ALD on biopsy /clinical Consecutive prospective recruitment	n/r	n/r	52	84	Local	PIIINP
Poynard 1991[138] (1982-1987) France *(1)*	624 (a) 333 training (b) 291 validation	Patients admitted with alcoholism or diagnosed ALD Consecutive prospective recruitment	≥ 50g alcohol daily for last 5 years	29	49	75	Local	PGA PT
Li 1994[139] USA *(1)*	44	Patients undergoing biopsy for clinical reason with h/o heavy alcohol Prospective recruitment	>80g daily at least 5 years	34	45 (range 27-69)	100	Local	PIIINP TIMP 1
Oberti 1997[140] France*(1)*	160 total (a)109 compensated	Admissions for alcoholism/ diagnosed ALD Consecutive prospective recruitment	>50g alcohol daily for 5 years with elevated AST>6m	59	65	n/r	modified METAVIR	HA PT
Tran 2000[141] (1997-1998) France*(1)*	146	Heavy drinkers admitted for detoxification+/-rehabilitation Consecutive prospective recruitment	>80g alcohol daily for >5 year	40 (51)	49	73	Local	HA PGA Tran index: (HA; PT; Apo A1)YKL
Plevris 2000[142]	70	Patients with ALD diagnosed by histology Prospective recruitment	n/r	n/r	n/r	n/r	Local	HA
Croquet 2002[143] France*(1)*	240	Patients admitted for alcoholism or ALD Prospective recruitment	50g daily past 5 years	48 (74)	n/r	n/r	METAVIR	PT HA

Study Author: Yr published (date of study) country *No. centres*	**Total no patients**	**Patient selection** **Recruitment details (where reported)**	**Alcohol consumption inclusion criteria**	**% cirrhosis (significant fibrosis*)**	**Age Yr mean (SD)**	**% male**	**Liver biopsy scoring system**	**Serum marker**
Stickel 2003[144] Germany *(1)*	87	Admissions for alcohol withdrawal symptoms in current drinkers	>100g alcohol daily	14 (44)	n/r	n/r	Ludwig	HA
Rosenberg 2004[30] (1998-2000) *England (8) Germany Italy Sweden*	64	Patients with excess alcohol consumption history and histology Consecutive prospective recruitment	Assessed by each centre	27	44	63	Scheuer Ishak	ELF panel (HA TIMP1 PIIINP age)
Naveau 2005[145] (1996-2000) France(1)	221	Patients with excess alcohol consumption history and with available histology Prospective recruitment	>50g alcohol daily for 1 year	(42)	47	77	METAVIR	*Fibrotest * HA
Cales 2005[146] (1994-2002) France *(1)*	95	Heavy drinkers with ALD on histology Consecutive prospective recruitment	>50g daily >5 years	41 (80)	49.8 (11.2)	71.6	METAVIR	Fibrometer
Lieber 2006[147] USA *(23)*	1034: (a) 507 pre-cirrhotic (b) 527 decompensated cirrhosis	Patients with heavy alcohol consumption +fibrosis/cirrhosis on biopsy/clinical in 2 treatment RCTs Prospective recruitment	80g ethanol daily >5 years HCV negative	(a) 31 (b) 100	(a) 51 (b) 56	(a) 97 (b)98	Ishak	APRI(AST Platelets)

*(significant fibrosis METAVIR F2-4; Ishak 3-6. Cirrhosis METAVIR F4 Ishak Stages 5,6)

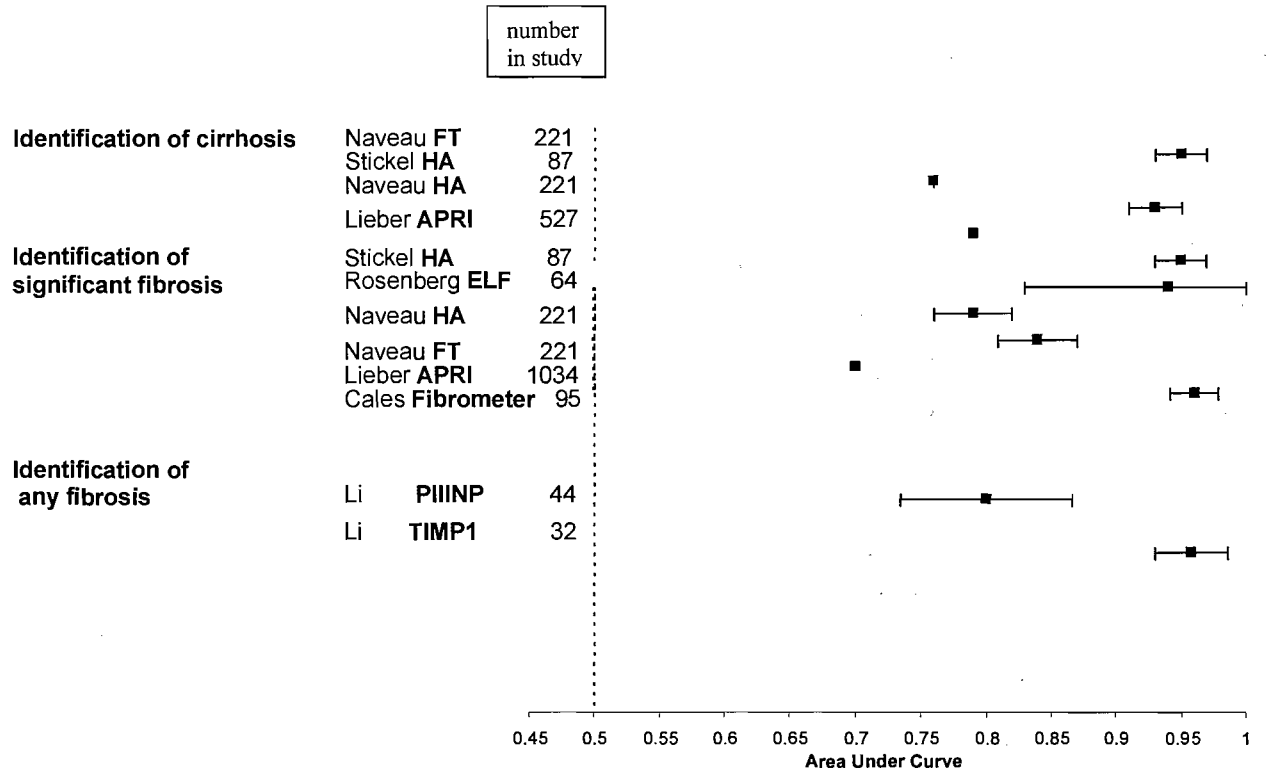
Table 3.9 Diagnostic performance of single markers in Alcoholic Liver Disease

Degree of Fibrosis tested	Study	No.	AUCs (95% CI)	Cut off used	Diagnostic accuracy (TP+TN/Total) %	Sensitivity	Specificity	LR+/ LR-
HA								
Cirrhosis	Oberti (1997)[140]	109*	n/r	60mcg/l	87 (77-93)	100	60	2.5/0.02
	Tran (2000)[141]	146	n/r	60mcg/l	91	100	86.2	7.1/0.01
	Plevris (2000)[142]	70	n/r	100mcg/l	n/r	87	89	7.9/0.15
	Stickel (2003)[144]	87	0.78	250mcg/l	n/r	100	69	3.2/0.014
	Naveau (2005)[145]	221	0.93 (0.91,0.95)	n/r	n/r	n/r	n/r	n/r
F01 vs34	Stickel (2003)[144]	87	0.76	55.5 mcg/l	n/r	82.8	69	2.7/0.25
	Croquet (2002)[143]	240	n/r	n/r	85 (77-89)	n/r	n/r	n/r
F01vs 234	Naveau (2005)[145]	221	0.79 (0.76-0.82)	n/r	n/r	n/r	n/r	n/r

Degree of Fibrosis tested	Study	No.	AUCs (95% CI)	Cut-off used	Diagnostic accuracy (TP+TN/Total) %	Sensitivity	Specificity	LR+/ LR-
P3NP								
F012 vs34	Gabrielli 1989[137]	44	n/r	16ng/ml	n/r	71	50	1.42/0.58
F0 vs. F1-6	Gabrielli 1989[137]	44	n/r	16ng/ml	n/r	90	59	2.2/0.17
	Li 1994[139]	44	0.80 ±0.07 (SD)	1.1 U/ml	n/r	45	100	45/0.55
Prothrombin Time								
Cirrhosis	Oberti 1997[140]	109	n/r	85%	82 (77-87)	n/r	n/r	n/r
	Croquet (2002)[143]	240	n/r	80%	85 (81-91)	81	99	81/0.19
	Tran (2000)[141]	146	n/r	85%	n/r	83	93	11.9/0.18
F01v 2-4	Croquet (2002)[143]	240	n/r	80%	78 (71-83)	n/r	n/r	n/r
TIMP								
Any fibrosis	Li (1994)[139]	44	0.96 ±0.03 (SD)	313ng/ml	n/r	n/r	n/r	n/r
YKL								
Cirrhosis	Tran (2000)[141]	146	n/r	330mcg/l	n/r	51	89	4.6/0.55

*compensated disease

Figure 3.5 Forest plot of the AUC for serum markers in ALD in identification of different fibrosis severity



AUC values (where reported) for all serum markers studies in patients with ALD identifying cirrhosis, significant fibrosis or any fibrosis with 95% CI (where reported). Most studies are small (wide confidence intervals), varying in threshold reported, and where >1 study, per serum marker results are inconsistent.

Table 3.10 Diagnostic performance of panels of markers in ALD

Fibrosis grade	Study	No.	Test	Components of panel	AUCS	Cut off	Sens	Spec	PPV	NPV	LR+ (95% CI)	-LR (95% CI)
Cirrhosis	Poynard 1991[138]	624	PGA	GGT PT	n/r	6	85	85	70	93	5.6 (4.5 7.01)	0.18 (0.12,0.25)
Cirrhosis	Tran 2000[141]	146	Tran	ApoA1 HA, PT	n/r		76	99	98	86	66.8 (9.5,471.2)	0.24 (0.15,0.37)
Cirrhosis	Naveau 2005[145]	221	Fibrotest	α2M apoA1 bilirubin, GGT haptogloblin, corrected for age + sex	**0.95 (0.94, 0.96)**	0.3 0.7	84 60	41 72	39 49	85 80	1.4 (1.2,1.7) 2.1 (1.6,2.9)	0.39 (0.2,0.70) 0.55(0.40,0.75)
Cirrhosis	Lieber 2006[147]	1034	APRI	AST Platelets	**0.79**	>2.0	17	86	56	50	1.2 (0.9,1.6)	1.0 (0.92,1.02)
F012vs 34 Severe	Rosenberg 2004[30]	64	ELF	Age, HA PIIINP TIMP-1	**0.94 (0.84, 1.00)**	0.087 0.431	100 93	17 100	75 100	100 86	1.2 (1.1, 1.4) 68 (37,124)	0.06 (0.01, 0.3) 0.08 (0.05,0.1)
F01 vs 2-4 Mod/severe	Cales 2005[146]	95	Fibrometer	α2M AST, HA PT, Pl, ,	**0.96 (0.94, 0.98)**	n/r	92	93	99	76	18 (2.7,125)	0.08 (0.2)
F01vs 2-4 Mod-severe	Naveau 2005[145]	221	Fibrotest	α2M apoA1 (corrected for age + age) bilirubin, GGT haptogloblin	**0.84 (0.81 0.87)**	0.3 0.7	84 55	66 93	81 93	70 54	2.5 (1.8,3.4) 7.4 (3.3,16.1)	0.25(0.16,0.40) 0.5(0.4,0.6)
F01vs2-4 Mod severe	Lieber 2006[147]	507	APRI	AST Platelets	**0.70**	0.2 0.6 1.0 1.6 2.0	94 47 21 13 9	26 82 90 95 97	71 84 80 83 86	68 44 37 36 35	1.3 (1.2,1.4) 2.6 (2.0,3.3) 2.1 (1.5, 3.0) 2.5 (1.5,4.1) 3.1 (1.6,6.1)	0.24(0.17,0.33) 0.65(0.6,0.71) 0.88(0.83,0.92) 0.92(0.88,0.95) 0.94(0.91,0.96)

Results are presented separately for single markers (Table 3.9) and for marker panels (Table 3.10) in the identification of cirrhosis (F4 METAVIR) cirrhosis, /severe fibrosis (F3/F4 METAVIR) and 'significant' fibrosis (F2-4-METAVIR). There were 12 separate markers evaluated- 6 single markers and the rest as components of 6 panels of markers.

Single markers (Table 3.9)

All single markers were heterogeneous with respect to the grade of fibrosis identified by the test, and the thresholds reported (with the exception of two studies using HA both of which used 60mcg threshold. Two studies reported AUC and sensitivity and specificity at more than one threshold [145,147]

The most commonly measured marker was HA, five studies evaluated its performance in the identification of cirrhosis/severe fibrosis, although only 2 of these reported AUC values. None reported results for the identification of patients with no or mild fibrosis. The AUCs for the 2 studies identifying cirrhosis were discrepant -0.79 and 0.93. The inclusion criteria for each study were similar, but the size of the studies were different (n=221 (Naveau 2005) Vs. n=87 (Stickel 2003), and the prevalence of cirrhosis in one study (Naveau 2005) was twice that in the other (Stickel 2003). There is one direct comparison of a panel and HA and this showed that there was no significant difference between panel (Fibrotest) and HA at both identifying cirrhosis and moderate /severe fibrosis. HA was better at identifying cirrhosis alone than moderate/severe fibrosis. The LRs and predictive values showed that HA is better at excluding cirrhosis/ severe fibrosis than detecting it. The studies were all small (~≤200) and reported different thresholds of HA concentration for positive test results.

There was more limited data on other single markers. Only one small study evaluated the performance of PIIINP in the identification of severe fibrosis /cirrhosis, and three small studies its ability to identify any fibrosis present (two of these presented ROC analyses). Four studies (two independent populations) evaluated prothrombin index and one YKL, and none reported ROC analyses. Prothrombin index was found to have good diagnostic accuracy in identification of cirrhosis (85% (95% CI 81-91). One small study reported performance of TIMP in identifying any fibrosis (AUC 0.96) but this has not been repeated. Overall there were three studies that had positive LR >10 and three negative LR <0.1 (excellent test). There was between study variation with respect to LR derived for individual markers. For example LR+ for P3NP ranged from 2.2 to 45.

Marker panels (Table 3.10)

Five studies assessed the performance in detecting cirrhosis. Two larger studies reported AUCs (0.79 and 0.95) [145,147]. Two panels showed promise in detection of cirrhosis although one was very small (ELF n=64), and one showed no statistically significant difference to HA in direct comparison (Fibrotest). One panel (APRI) had a good but lower AUC compared to ELF and Fibrotest (0.79 vs. >0.94). Two panels had good diagnostic performance in detection of moderate/severe fibrosis, Fibrometer (n=95) and Fibrotest (though not significantly better than HA in direct comparison). In general, panels of markers reported lower diagnostic performance in the detection of lesser stages of fibrosis than in cirrhosis. All AUC data are shown in Figure 3.5. Positive LR for marker panels had a median of 3 (range 1-68), with two studies having values >10 and two with negative LR <0.1 indicating very good performance.

3.5 Discussion

Statement of principle findings by cause of CLD

CHC

14 primary studies were identified using 10 different combinations of serum markers. Median value of AUC- (mild vs. moderate/severe fibrosis) in validation populations being 0.77. All tests showed good performance in differentiating cirrhosis/no cirrhosis with median AUC in validation sets of 0.87.

NAFLD

29 primary studies were included. The variables most commonly associated with fibrosis were: presence of diabetes, increasing age, increased HOMA-IR, increased AST/ALT ratio, decreased platelets, hyaluronic acid and BMI. A subsequent primary study has gone on to show that ELF performs better than simple markers in the detection of fibrosis in NAFLD and that the combination of ELF and simple markers is excellent[148].

ALD

12 primary studies were included. The evaluations used 12 different markers, most commonly HA, and 6 panels. Serum markers appear to be able to identify those people with severe fibrosis/cirrhosis with good diagnostic accuracy at the thresholds presented, performance was less good at less severe levels. Hyaluronic acid as a single marker performs well to identify

cirrhosis but has lower AUC results to panels. Overall the performance of the serum markers was less good at identifying lower grades of fibrosis.
None of the reviews supported the complete replacement of biopsy by serum markers at present, although there seems a place for serum markers to rule in/out moderate fibrosis or cirrhosis at acceptable and reasonable test performance that may be applicable in about 30-40% of the study populations.

The following discussion will initially outline strengths and limitations of the review process, it will go on to focus on those differences between the markers in CHC, NAFLD and ALD, and then broaden out into a general discussion on common issues arising from all three diagnostic reviews.

Strengths and Limitations of the methodology of the systematic reviews

Standard published methods for conducting systematic reviews were used. Recommended search strategies to locate diagnostic tests were used. Two reviewers independently extracted data from selected articles.
Despite the rigorous methods employed, relevant studies may have been missed, in particular smaller studies with negative results or unpublished studies (which were not included in the systematic reviews as abstracts from conferences can be searched for but tend not to present sufficient data). This potential publication bias could be addressed using Funnel plots.
Articles in languages other than English were not searched for nor selected. This may have contributed to any missing data. Heterogeneity of studies precluded meta-analysis of data in NAFLD and ALD.
Systematic reviews of studies in patients with the three main aetiology groups of CLD were evaluated and other important causes such as PBC, HIV-HCV co-infection, Hepatitis B were not performed.
As with all systematic reviews regular updating is needed. This is especially relevant to non-invasive markers of liver fibrosis where the research field has been very active over the past 3 years.
Other test characteristics such as cost effectiveness were not addressed.

Differences between the reviews

Most serum markers in ALD and NAFLD were single markers with few panels of markers derived in these populations. The studies tended to be smaller, had not been validated in external populations and generally used local non- standard histology scoring systems. This is contrast to CHC where most of the liver histology scoring systems were internationally recognised and validated. The literature in ALD was older, included studies were fewer in numbers, had fewer participants, had different inclusion criteria, and had a higher prevalence of cirrhosis/severe fibrosis than in similar studies in other chronic liver diseases. Only in CHC was sufficient data presented to perform any summative analyses. In general the studies included in the ALD and NAFLD reviews were of a poorer quality than those in the CHC review.

Methodology issues

In the CHC review most studies reported AUC in differentiating mild /no fibrosis and significant fibrosis, and 10 had sufficient data to construct a 2x2 table which differed in the other reviews where the minority reported results in terms of AUC or sensitivity and specificity. It was therefore impossible to derive 2x2 tables in any study of NAFLD or ALD. The studies in these two reviews often reported results as the differences in means between those patients with different grades of fibrosis/cirrhosis and no cirrhosis, or correlation between serum markers and grade of fibrosis. Much of the data presented showed the association of variables with fibrosis or cirrhosis in univariate or multivariate analyses, but had seldom reported results as sensitivity and specificity at different thresholds. It was not possible in these reviews to perform meaningful summative analyses. In addition there had been a previous systematic review in CHC which had highlighted the promising performance of panels of markers which were not published in the other two fields. The diagnostic literature in CHC is thus more methodologically advanced than ALD or NAFLD. Study designs lagged behind CHC in that case control studies were found in ALD and NAFLD literature. These are less robust than performing diagnostic performance evaluations in prospectively consecutively recruited cohorts.

Both NAFLD and ALD have complex and overlapping clinical and pathological pictures. ALD is a relapsing remitting condition which leads to a less clear demarcation of severity categories than in CHC, with subsequent difficulty in diagnosis of fibrosis alone. This may have contributed to the general difficulties in untangling results for fibrosis/cirrhosis in the included studies. A similar picture is present in NAFLD with the boundaries between steatosis, inflammation and fibrosis being fluid and difficult to disentangle. These confused and overlapping diagnostic problems may well reflect the real world of patients where, for example, the identification of AH ± fibrosis become important, leading to different management decisions and prognostic information. Clinically in NAFLD and ALD it may be important to identify those patients who;

1. are at risk of developing significant liver disease- those with any fibrosis- both to focus interventions to optimise abstinence/reduce weight, and to monitor disease progress more closely,
2. have severe fibrosis/cirrhosis to focus interventions and to begin to screen for varices and HCC or to prepare for possible liver transplant

The three systematic reviews presented have focused on the ability of diagnostic tests to identify fibrosis, their ability to identify inflammation has not been addressed.

In CHC it is now no longer necessary to distinguish those patients with moderate or severe disease in order to identify people who are eligible for treatment with anti-virals following the most recent NICE guidelines.

Similarities between the reviews

Heterogeneity

All the study populations in each of the three reviews were heterogeneous. Whilst they all recruited patients from specialist clinics in secondary or tertiary settings (there were no studies set in primary care), there was variation between studies in the population characteristics, (such as alcohol consumption which varied even in the ALD review), recruitment methods, prevalence of severe fibrosis, and methods of test validation. Differences in the study population characteristics may lead to spectrum bias. This is where the sensitivity and/or specificity of a test may vary with different populations tested - populations which might vary

in sex ratios, age, or severity of disease as three simple examples. All patients had to have had a biopsy (from inclusion criteria) which could introduce verification bias with those patients not selected for biopsy and not included in the studies potentially having a different disease severity than those who were selected. These methodological issues are discussed further in Chapter 6.

Direct comparison between studies was made more difficult by the use of a range of fibrosis scoring systems, largely locally generated. There was incomplete reporting of co-morbidities and diagnostic test results and it was not possible to conduct any summative evaluation in ALD.

Generalisability

The heterogeneity of study populations may lead to problems of generalisability, and this may be reflected in a reduction or an increase depending on clinical setting. In particular the proportion of significant fibrosis differed between included studies and this has an effect on predictive values. Therefore knowledge of the fibrosis prevalence is necessary to determine appropriateness of a test to individual clinical practice. It is possible some tests might perform better in low or high prevalence populations, for example, those with a high sensitivity across lower test scores, will perform best in low prevalence populations as the NPV will be higher and the test is applicable to a significant part of the study population; the converse would apply in high prevalence populations.

Evaluation of Diagnostic Test used

The Area Under the Receiver Characteristic Curve (AUC) has limitations and may not be the best way to present test performance[149;150]. Whilst the markers and panels in these reviews may have performed with good sensitivity and specificity at highest and lowest thresholds, the AUC does not adequately reflect the test performance at intermediate thresholds where sensitivity and specificity are considerably lower. A misleadingly high value for the AUC may thus disguise the true diagnostic performance across all thresholds. Other diagnostic test evaluations have selected LR and DOR as better ways of evaluating and comparing tests and these measures may be more discriminating. A LR+ describes how many times more likely a

person with the disease will receive a particular test result than a person without the disease. The median LR for panel serum markers was 8.2, with the majority falling outside the conventionally accepted "good" test range (≥10).

Some authors have suggested that the DOR is used[89]. This describes the ratio of the odds of a positive test result in a patient with disease, compared to a patient without the disease (+ LR/- LR). A DOR of 1 suggests a test providing no diagnostic evidence, and a reasonable test may have DOR >30. DORs have been calculated for those studies that presented sufficient data, with the median DOR being 9 (range 5-27). DORs are not easy to apply in clinical practice, but are useful in comparisons such as when combining results in a systematic review and are reasonably constant regardless of the diagnostic threshold. The ability to derive a DOR is dependent upon the presentation of sufficient data to construct a 2x2 table. Further studies should be encouraged to report data in such a way to facilitate these analyses.

The findings of these reviews can be applied in clinical practice to avoid liver biopsy using test threshold levels at illustrative predictive values, identifying the presence or absence of significant fibrosis. This figure can be increased by relaxing the probability of making correct assignment. This method of reporting test performance may be useful and help in the critical assessment of evidence by clinicians before using these markers in their practice.

Diagnostic test reporting and quality of studies

Generally, methods of evaluating the quality of diagnostic tests are not as refined as those for therapeutic trials, with reporting of test evaluations similarly lagging behind. An algorithm for reporting diagnostic tests, similar to the Consort statement for treatment trials, has been published-Standards for Reporting of Diagnostic Accuracy (STARD), and is gaining wider acceptance[151;152]. This will contribute to the improvement of the quality of both the conduct and reporting of diagnostic evaluations. In these reviews QUADAS quality tool was used to assess quality and found most studies used blinded outcome assessment, were explicit about patient selection and exclusions, and used an accepted reference standard, all of which have been cited as the most important criteria that impact on study quality[89] However, sources of potential bias were identified including the incomplete reporting of data (e.g. co-morbidity, alcohol consumption), and sensitivities and specificities at all thresholds. Some studies used

the same cohort both to derive and evaluate the performance of the markers or an internal validation cohort, (where patients were recruited from a group similar to the training cohort) both of which limit generalisability of the findings. A more rigorous methodology would be to derive the panel in one cohort and validate in several external populations or (most preferably) a reference population in which different tests could be tested and directly compared using a standardised reference test.

Liver biopsy as reference standard

A fundamental methodological limitation in assessing non-invasive markers is the use of liver biopsy as a reference standard, and this may underrate the performance of these tests[153]. In addition to the sampling error and observer variability raised in the introduction, there are difficulties in obtaining an adequate sample –some experts have suggested that 20mm in length, others that 11 portal tracts is optimum[60; 61.A] recent systematic review on the quality of liver biopsy specimens concluded that here is poor reporting on the biopsy length, size of needle used and number of portal tracts[154]. Data on the discordant results between histology and one panel of markers have been explored with attribution of discordance to biopsy failure in 18 % cases, failure of markers in 2.4% and non-attributable in 8.2% cases. The authors concluded that in many cases of difference it is the shortcomings of the biopsy that are responsible and this leads to an underestimation of the diagnostic performance of the serum markers[157]. It would be interesting to extend this work to different panels of markers.

Lack of a universal scoring system of fibrosis adds to the difficulties in comparison between studies, and consideration should be given to professional consensus in the standardisation of scoring systems. Whichever score is used, the histological staging of liver fibrosis on biopsy is artificially represented as a quantitative categorical variable with a linear quantum progression in severity from 0 to 4 or 6. This does not accurately reflect the dynamic biological process of fibrosis and constrains the serum marker test performances that are capable of generating continuous variables. Fibrosis progression is likely to be non-linear, and there is not equal temporal progression between sequential stages. It is therefore important to consider how to improve the reference standard in liver disease, or the use of a different reference standard, such as clinical outcomes (mortality or serious morbidity), although these may be limited by

the use of subsequent interventions that may affect test/outcome relationships where there are effective treatments and length of time required to reach clinical endpoints. Further research is needed to evaluate such alternatives.

Clinical implications

For preventing and managing CLD it is important to identify those patients who have clinically silent severe fibrosis/cirrhosis to focus interventions, to begin to screen for varices and HCC, or to prepare for possible liver transplant. Data presented in this review suggest that marker panels could be effectively used in this situation. It would be clinically useful to patient and clinician to identify the proportion of hazardous drinkers who have developed liver disease to monitor disease progress more closely and to offer an opportunity for strategies aimed at reduction/abstention. Repeated serum markers measures showing rise or decline may have an impact on lifestyle choices again allowing scope for reduction in alcohol consumption. These are speculative ideas and require further research. This group of patients often has erratic attendance at outpatient and biopsy appointments and may present in settings where invasive tests are inappropriate/difficult (e g prison). Having non-invasive tests of liver fibrosis would be useful in the management of such patients.

The ideal test for fibrosis would be easy to perform, repeatable and capable of frequent application. It would be highly accurate over the full range and reproducible. The test would provide an accurate assessment of the degree of liver fibrosis throughout the range of matrix deposition from mild scarring, through compensated cirrhosis and then beyond to provide a clear picture of worsening degrees of decompensated cirrhosis. The test would be highly predictive of long term outcomes such as hepatic decompensation, portal hypertension, liver failure, liver cancer, the necessity for transplantation and death from liver disease. It is clear from the systematic reviews presented in this chapter that the current serum markers fall some way short of this ideal, but they are promising, improving and may provide additional diagnostic information in the identification and management of people with CLD. In addition they may fulfill a role in the future in the assessment of outcomes in treatment trials.

Serum marker literature update

Since these systematic reviews were conducted, there have been many more studies adding to the evidence base of the performance of serum markers in identification of liver fibrosis which meet the inclusion criteria of the reviews. Most have been in CHC, and most involve <200 patients. There is a growing literature on other blood borne viruses (Hepatitis B, HCV-HIV co-infection). There have been no further studies on ALD. (See Appendix 4 for summary of studies). There have been two meta-analyses of a single serum marker (Fibrotest), one of the serum panel APRI, and one in HCV-HIV co-infection[156-159].

Fibrotest: In the former non-independent authors pooled 30 studies in 24 articles (6,378 patients) of which half of the included studies were independent validations (of the company licensed and marketing Fibrotest). Most of these were in CHC (20/30) and in 11/20 were in independent studies. Mean AUC for identification of significant fibrosis (stages 2-4) for CHC was 0.77 (95% CI 0.75, 0.79), NAFLD (95% CI 0.81 (95% CI 0.74, 0.86), and ALD 0.85 (95% CI 0.80, 0.89). The authors concluded that Fibrotest could be used as an alternative to biopsy and that biopsy should be used as a second line investigation for clinico-biological discordant cases.

Independent investigators (with no financial interest in the panel) evaluated 9 relevant studies in CHC (1,679 total subjects) and found that in a heterogeneous analysis AUC for identification of significant fibrosis was 0,81, and that at a threshold of 0.60 the sensitivity of Fibrotest was 47% and specificity 90%. For the identification of cirrhosis AUC was 0.90. The diagnostic accuracy was greater the more severe the fibrosis. The authors conclude that Fibrotest has good accuracy for the identification of severe fibrosis but lesser accuracy for earlier stages and that refinement was required before biopsy could be replaced.

APRI: In the systematic review of APRI in patients with CHC, 22 studies were reviewed (n=4,266) and SROC analysis performed. This found summary AUC for significant fibrosis and cirrhosis to be 0.76 (95% CI 0.74, 0.79) and 0.82 (95% CI 0.79, 0.86) respectively with a summary DOR of 11.3 (95% CI 7.9, 16.0). For the identification of significant fibrosis and using a threshold of 0.5, APRI was 81% sensitive and 50% specific. AT prevalence of 40% NPV was 80% and could reduce the need for biopsy by 35%. The performance of APRI was unaffected by study or biopsy quality. The authors concluded the strength of APRI was in

exclusion of significant fibrosis and they recommended that other biomarker should show cost effectiveness and enhanced performance compared to APRI.

HCV-HIV co-infection: 6 studies were evaluated (646 total patients) and AUC values derived form SROC analysis were for significant fibrosis 0.80 (95% CI 0.77, 0.83) and for cirrhosis 0.78 (0.73, 0.82). The authors concluded that more research was needed to elaborate the role of serum markers in co-infection, and before any one panel of markers could be selected for use in this population.

Other main themes in the literature since the reviews are the use of other modalities than serum biomarkers and the use of sequential serum marker algorithm pathways in patient management. There have been many studies in the last 2 years on the performance of other modalities in particular transient elastography and its use in combination with serum markers[27;160;161]. Transient elastography uses an ultrasonic transducer mounted on the axis of a vibrator which generates an elastic shear wave which moves through the tissue. The speed of this wave is measured and this is related to tissue stiffness. The harder the tissue the faster the shear wave moves. Studies have shown that this technique correlates with fibrosis stage. However recent studies have drawn attention to the limitations of this technique in obese subjects and those with inflammation[162;163].

There has been a general trend to publish clinical algorithms to place serum markers in the context of clinical practice. Some of these use a cheaper test (such as APRI) to identify significant fibrosis followed by a further non-invasive test for those that could not be allocated by the first test[164]. In all these suggested pathways the problems of balancing false test rates against keeping the diagnostic performance high, in addition to maximizing the number of patients that can be allocated to a fibrosis category has not been completely solved. There remains no accepted way of using markers in clinical practice, although French health authorities have introduced a national recommendation for the use of Fibrotest ± transient elastography in patients with CHC in order to reduce the number of biopsies that need to be conducted.

Future research

From this overview of the literature of non-invasive markers research recommendations include:

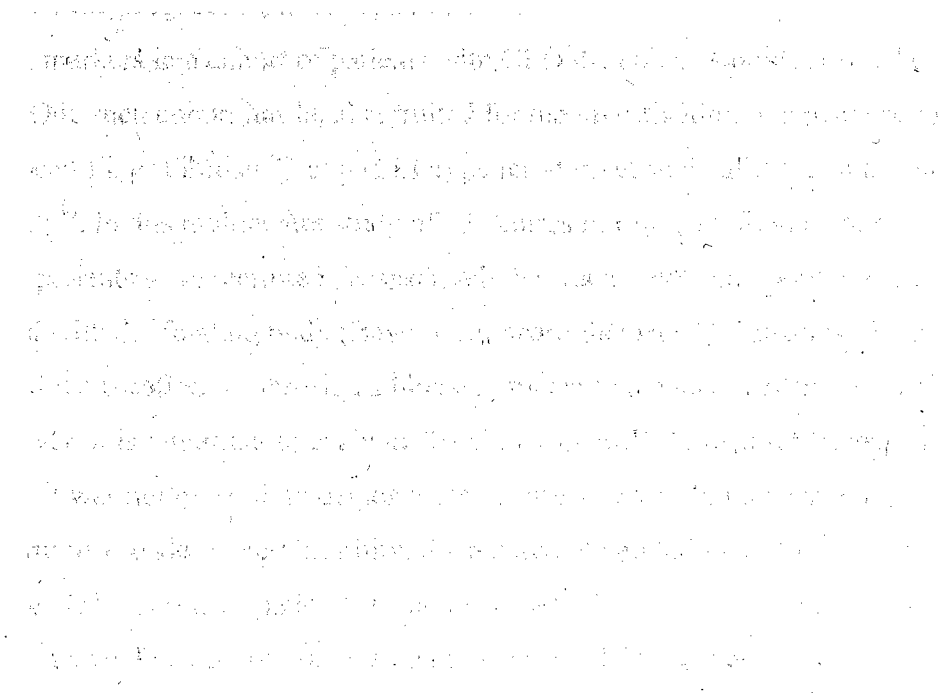
1. the identification of new markers using methods such as proteomics and metabonomics[165] and the use of these technologies with existing or new panels of markers, either in isolation or in combination with emerging quantitative imaging techniques that may improve test performance and the ability to distinguish individual stages of fibrosis.
2. Composite scores from emerging panel markers in CHC, NAFLD and ALD will need validation by independent bodies. This process might be facilitated by establishing an international reference library and quality assurance scheme. The evaluation of diagnostic performance must be accompanied by parallel evaluation of test performance for properties such as reproducibility, stability and linearity.
3. The place of the markers in identifying those people with CLD due to alcohol or obesity who have sustained any fibrotic liver damage in lower prevalence settings such as primary care is of great importance. Should these people be identifiable with good diagnostic accuracy then interventions of lifestyle and pursuance of abstinence could be focused and referral to secondary settings made more appropriate. Further work is needed to ascertain the diagnostic performance of markers in such as setting
4. Use of serum markers in combination with other diagnostic modalities such as serum biomarkers and radiological imaging may increase diagnostic accuracy and allow greater separation of stages of fibrosis. Radiological techniques such as transient elastography, NMR spectroscopy and microbubble ultrasound have been used in the context of chronic hepatitis C, and trials in NAFLD and ALD are awaited.
5. Updating the systematic reviews in CHC NAFLD and ALD is needed with use of meta-analysis and funnel plots. Extension into Hepatitis B, HCV-HIV co-infection and PBC would be timely.

3.6 Conclusion

These reviews provide a systematic evaluation of the published evidence of the diagnostic performance of serum markers of fibrosis in CHC, NAFLD and ALD. They include novel summative analyses, and using methods adaptable by clinicians, have demonstrated the clinical utility of markers to inform use in practice. They have highlighted methodological limitations of commonly used performance measures and suggested alternatives. The current

non-invasive markers potentially allow clinicians to select patients with severe fibrosis or exclude severe fibrosis but in individual patients cannot differentiate the stages of fibrosis reliably, and only in a minority of the population tested will a test result have a high diagnostic performance. In clinical practice, this may allow the reduction in number of biopsies performed. Moreover it is a useful alternative in patients having an absolute contraindication or refusing percutaneous liver biopsy.

The limitations of the liver biopsy may create a glass ceiling for potential non-invasive tests, and continuing improvement of both the index and reference test is needed before the ideal surrogate test for liver biopsy is found, and in this regard clinical outcomes should be evaluated. There is therefore a great need for studies that evaluate the performance of the markers in terms of their ability to predict death or serious clinical outcomes. Such a study is described in Chapter 5.



CHAPTER 4

DIAGNOSTIC PERFORMANCE OF THE EUROPEAN LIVER FIBROSIS (ELF) PANEL MARKER

In this chapter the original study that derived and internally validated the ELFa panel of serum markers will be described in order to provide background for the external validation studies conducted in eight external independent populations of patients with four different aetiologies of CLD. The eight external validation studies are then reported.

4.1 Introduction

The three systematic reviews reported in Chapter 3 have shown that although all serum markers have limitations, there are promising tests that identify severity of liver fibrosis in all three major liver diseases. Currently in the assessment of liver fibrosis none can replace liver biopsy entirely, and although open to limitations of sampling and interpretation, liver histology remains the reference standard for the diagnosis of fibrosis, itself a proxy for clinical outcome. There have been recommendations that clinical outcomes are used directly to evaluate the prognostic performance of serum markers. In order to conduct such a study of serum markers in a cohort of patients with CLD the cohort would need to be followed up over time. One such cohort has been recruited for the investigation of a panel serum markers -the European Liver Fibrosis Group (ELFa) panel -derived and validated in a cohort of 921 patients[30]. In this multicentre study of 13 centres in England (8) and continental Europe (5) 1,021 patients were recruited prospectively between 1998 and 2000. Access to these data was agreed with the funding body (Bayer Healthcare/Siemens Diagnostics). ELFa performed well in the identification of fibrosis on biopsy, especially the more severe spectrum of disease. However, it is important to evaluate the ability of ELFa to identify fibrosis in populations in which it was not derived, to demonstrate its effectiveness in independent populations and to confirm its broader generalisability. Such external validation studies have been conducted in independent cohorts of patients using this panel of markers to evaluate their diagnostic performance. For ease of nomenclature the original ELF panel is denoted ELFa ("European

Liver Fibrosis" panel), and a simplified version of ELF whose derivation is presented in this chapter denoted as ELF (Enhanced Liver Fibrosis Panel).

4.2 Description of original ELFa* cohort study

Patients were recruited to the study if they were undergoing a liver biopsy for the investigation of CLD defined as abnormal biochemical liver function tests persisting for more than three months. Additional inclusion criteria used were the ability to provide informed consent and ages between 18 years and 75 years. Patients were excluded if they fell outside this range, they had any disorder associated with extra hepatic fibrosis including rheumatic renal or pulmonary disease, if they had cardiovascular disease or cancer, had advanced cirrhosis with evidence of decompensation (Child Pugh Class C), consumed regular aspirin, or had hepatocellular carcinoma or drug induced liver disease. Ethical approval was given by UK South and West Multi-centre Ethics committee and all local ethics committees for each site (MREC98/6/08).

Baseline data collected

(i) Nine serum markers of matrix synthesis or degradation were measured: Collagen Type IV (Coll4); Collagen Type V1 (Coll6); Aminoterminal propeptide of procollagen type 111(P3NP); Laminin; Tenascin; Hyaluronic Acid (HA); Matrix metalloproteinases type1(MMP-2); Matrix metalloproteinases type 9 complexed with Tissue inhibitor of Matrix metalloproteinases type1 (MMP-9_TIMP-1); Tissue inhibitor of Matrix metalloproteinases (TIMP-1) (see appendix for role of Serum markers in ECM formation)

(ii) Liver function tests (usual local practice) from: ALT ; AST; GGT; albumin; bilirubin; alkaline phosphatase

(iii) Platelets

(iv) Prothrombin time/INR

(v) Glucose

(vi) Self –reported alcohol consumption

(vii) Body Mass Index

(viii) Self –reported smoking status

Serum samples were obtained at the time of biopsy. All histology was analysed locally and also by one central pathologist (Professor AD Burt University of Newcastle). Individual

fibrosis stage scores (Scheuer 0-4 and Ishak 0-6) were recorded. In addition the scores were categorised into binary outcomes denoting no/mild fibrosis (Scheuer 0-1 and Ishak 0-2) or clinically significant fibrosis (Scheuer 2-4 and Ishak 3-6) or cirrhosis (Scheuer 4 and Ishak 5-6)). The biopsy scores from the central pathologist were used in all subsequent analyses.

The primary analysis of the performance of the serum markers examined the ability of the algorithm to detect significant levels of fibrosis on biopsy.

Statistical analysis

Serum markers and blood parameters (such as liver function tests or platelets) were added sequentially and were included in the algorithm if their addition to the algorithm increased the overall generalized distance between groups in a discriminant analysis as judged by F test of Wilks's lambda. In addition to these variables the subject's age was included in the model. Prior to analyses all variables were log (base e) transformed. This was done to remove problems of concentration differences and to bring the standard error of each marker to within one order of magnitude of each other. An optimal algorithm was selected and the results were then combined to give an overall discriminant score and this was then validated in the remaining 521 cases. The algorithm was evaluated for its ability to identify outcomes:

(a) Identification of significant fibrosis –Ishak (3-6; Scheuer 2-4)

(b) Identification of cirrhosis

Results

Baseline characteristics

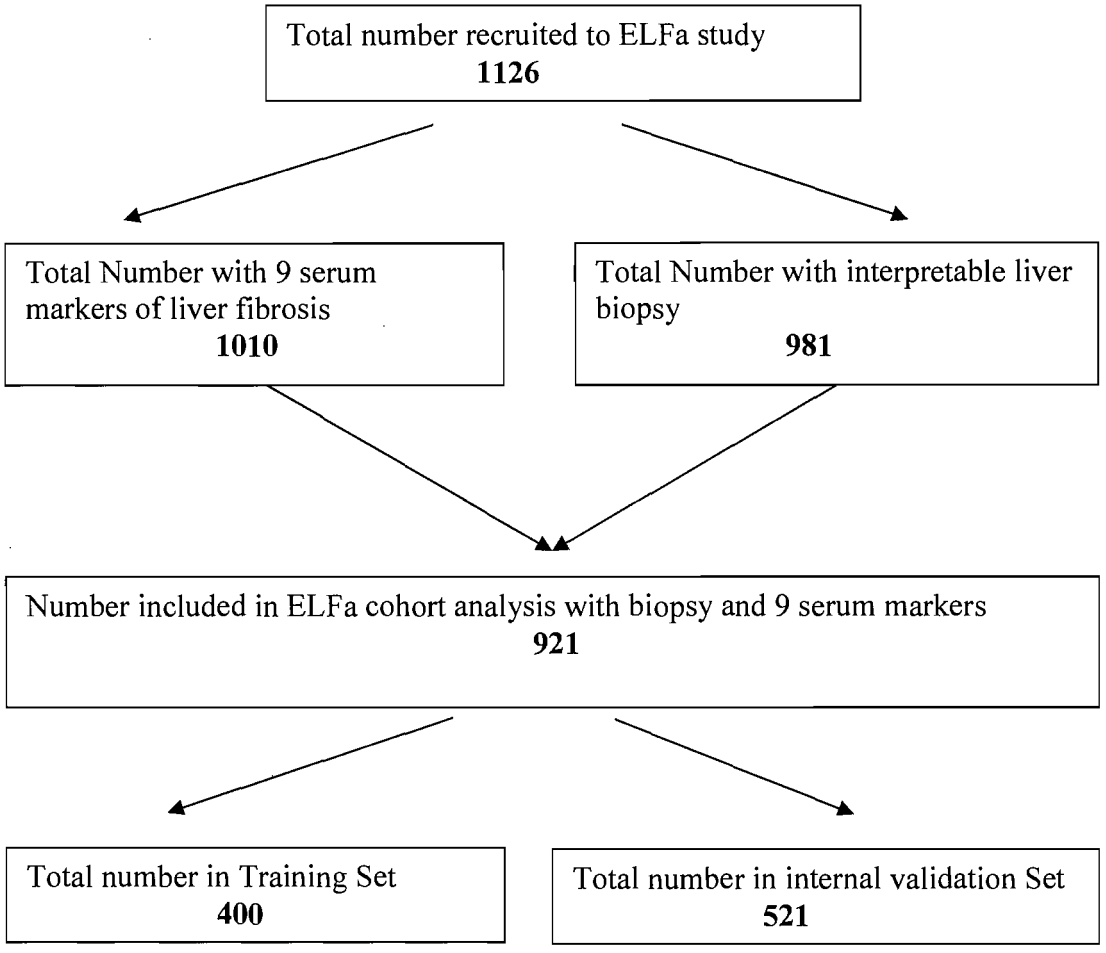
63% of patients were male, with mean age of 44.1(SD 12.8 years range 19-75 years).
26.7% had severe fibrosis/cirrhosis (stages 4-6). Numbers recruited by centre are shown in Table 4.1.

Table 4.1 Numbers recruited by centre and arm of study

Centre	Number recruited to cross sectional study	Number included in cross sectional study
Southampton	105	92
Cambridge	54	47
Liverpool	56	38
Newcastle-Upon-Tyne	148	121
Basingstoke	7	6
Oxford	83	71
Birmingham	90	75
Erlangen	98	77
Hannover	200	181
Florence 1	81	80
Florence 2	92	89
Stockholm	53	45
Total	1126	921

921 (82%) of recruited patients had adequate data on both serum and histology and were included in the final analysis (see figure 4.1). Centres were all secondary/tertiary hepatology centres, three were transplant centres (Cambridge, Newcastle, and Hannover).

Figure 4.1 Flow chart of patients in Original ELFa Study



Flow chart of the patients recruited to original ELF study who had an interpretable biopsy and full complement of serum markers to be evaluated (n=921)

The attrition in participant numbers occurred mainly due to poor quality biopsies which could not be interpreted, and failure of measurement of all nine of the serum markers values for each person.

Table 4.2 Baseline Ishak stage

Ishak Stage	Number	% per stage
0	226	24.5
1	200	21.7
2	133	14.4
3	116	12.6
4	83	9.0
5	56	6.1
6	107	11.6

Most of the participants had no or mild fibrosis on biopsy (46.2%) (Table 4.2). The majority of the patients had Hepatitis C (47%), and NAFLD (10%) and ALD (6%) reflecting the major causes of CLD, and also biopsy practice at the time of recruitment to the study (in order to meet criteria for treatment with anti-viral therapies, patients with CHC had to have a histological diagnosis of moderate or severe liver disease) (Table 4.3).

Table 4.3 Disease aetiology of original ELFa cohort

Disease Aetiology	Recruited (n)	Final dataset (n)	% of final dataset
CHC	496	432	46.9
NAFLD	61	92	10.1
ALD	64	55	6.0
Hepatitis B	61	44	4.8
Post Transplant	48	42	4.6
Autoimmune Hepatitis	45	36	3.9
Primary Biliary Cirrhosis	35	32	3.5
Hereditary Haemochromatosis	32	29	3.1
Primary Sclerosing Cholangitis	18	16	1.7
Normal or very Mild non-specific	12	12	1.3
Cryptogenic or Idiopathic	19	12	1.3
HBV and HCV Co-infection	4	4	.4
Granulomatosis hepatitis	3	3	.3
alpha 1 antitrypsin	2	2	.2
Unassigned	121	109	11.8
Total	**1,021**	**921**	**100**

There were fewer simple liver function biomarkers than serum markers of liver fibrosis, reflecting local laboratory practice as all simple markers were analysed in the recruiting centre (table 4.4).

Table 4.4 Summary of serum markers of liver fibrosis original ELFa cohort (n=921)

	Number tests	Mean (SD)	Median (IQR)	Range
HA	921	107 (395)	36 (18, 76)	1.9-9,325
P3NP	921	6.6 (7.7)	4.6 (3.2,7.2)	0.5-123
TIMP1	921	745 (450)	633 (481, 848)	4- 4,291
Laminin	921	22 (43)	15 (9, 24)	0.4-930
MMP2	921	634 (248)	599 (476, 740)	3.9-2,495
MMP9_timp	921	518 (271)	473 (340, 647)	0.1 - 2081
Tenascin	921	438 (280)	375 (265, 537)	2-2998
Coll IV	921	165 (126)	134 (107, 178)	3.1-1811
Coll VI	921	4.7 (2.4)	4.3 (3.3, 5.7)	0.6-30.8
INR	886	1.0 (0.2)	1.0 (0.9, 1.1)	0.1- 3.7
Creatinine	902	89 (18.9)	87 (77, 98)	49- 250
ALT	885	78 (113)	51 (26, 110)	4-1921
GGT	848	112 (203)	51 (31, 133)	3-1903
AST	706	55 (77)	37 (24, 60)	6-1311
Albumin	668	59 (86)	44 (40, 47)	21-67
Bilirubin	826	18 (35)	12 (9, 17)	3-495
Platelets	910	207 (71)	203 (159, 249)	40-669

Results of original ELF diagnostic study

The optimal panel selected contained age, HA, TIMP1 and P3NP (see below).

Final ELF panel

Scheuer (5 stages)

ELFa =-0.014*Ln (AGE) +0.616*Ln (HA) +0.586*Ln (PIIINP) +0.472*Ln (TIMP-1)-6.38

Ishak (7 stages)

ELFa =-0.08*Ln (AGE) +0.608*Ln (HA) +0.601*Ln (PIIINP) +0.511*Ln (TIMP-1)-6.26

The performance of ELFa panel was evaluated in the internal validation set (n=521), and found to have good performance at identification of significant fibrosis/cirrhosis and cirrhosis (see Table 4.5 for AUC (95% CI) results). At an ELFa threshold of 0.102, the sensitivity was 90%, specificity was 41%, +LR 1.5/-LR 0.24, and PPV was 35% and NPV 92% in this population.

Table 4.5 Diagnostic performance of ELFa panel for different cuts of fibrosis

Histology Ishak	AUC	95% CI	Sensitivity	Specificity	+LR	-LR
(0-2) v (3-6)	0.78	0.74 to 0.82	90	30	1.3	0.3
(0-3) v (4-6)	0.80	0.76 to 0.85	91	41	1.5	0.2
(0-4) v (5-6)	0.89	0.84 to 0.94	91	69	2.9	0.1

ELFa works particularly well in patients with NAFLD and ALD (AUC=0.94 for stages 4-6) although numbers of patients with ALD are small (n=55). (Table 4.6) The performance is less good in terms of specificity, PPV and likelihood ratios.

Table 4. 6 Diagnostic performance of ELFa panel in identifying serious fibrosis (Ishak 4-6) (Validation set n=521)

	AUC	95% CI
Hepatitis C	0.77	0.70 to 0.85
NAFLD	0.87	0.67 to 1.00
ALD	0.94	0.4 to 1.00

Conclusion from original diagnostic ELF study

The ELFa panel excluded significant fibrosis with sensitivity more than 90% and identified significant fibrosis with specificity of 90% using different thresholds of the ELF score. The positive predictive value and LR were less strong. It performed particularly well in a subset of ALD and NAFLD (although these populations were small and so the 95% CI were wide).

4.3 External Validation studies of the diagnostic performance of ELFa panel

The diagnostic performance of the ELFa markers (summarised above) was shown to accurately predict significant fibrosis. Following the publication of the ELFa panel, collaborations with centres in the USA, continental Europe and UK have provided access to cohorts of patients with different aetiologies of CLD. This not only provided an opportunity to externally validate the performance of ELF in independent populations, but also afforded a chance to determine if ELF could be simplified using the whole of the original ELF cohort (n=921) as a training set, whilst maintaining its performance.

The aetiologies and sources of patient cohorts in the 8 external validation studies were:

1. Chronic Hepatitis C
    a. Trent Region HCV dataset
    b. Southampton
    c. Centre for Disease Control and Prevention (CDCP) Atlanta, Georgia
2. NAFLD
    a. Newcastle
    b. Nottingham
    c. Rome Paediatric population
3. PBC
    a. University of Texas Southwestern, Dallas Texas
4. HCV-HIV co-infection
    a. University of California San Francisco

The initial section reports the simplification of the ELFa panel (4.3.1). The following sections report the external validation studies in patients with CHC, NAFLD, PBC and HCV-HIV co-infection. For each external validation cohort the recruitment of the participants of the cohort, the methodology of acquisition of serum and histology, and the baseline characteristics of the cohort will be described. A common method of analysis of these data was used and is described in section 4.3.6. Summary tables of diagnostic accuracy of ELFa and ELF in each

external cohort are reported. Additional analyses (clinical utility model) were conducted for NAFLD and these are presented in section 4.3.7.

4.3.1 Derivation of a simplified three marker panel ELF (Enhanced Liver Fibrosis)

The **entire** original ELF cohort of patients (n=921) with baseline biopsy results, TIMP1 HA, P3NP, and age was used as training population to explore if the ELFa panel could be simplified, and to optimise the diagnostic performance of the ELFa algorithm. The statistical analysis used in this simplification of ELFa panel was backward selection logistic regression using the individual markers in the ELF panel and age for binary grouped biopsy stage (stages0-3 vs. 4-6). This analysis was conducted in the whole of the original cohort (n=921) using SPSS for Windows version 14.0 Results from logistic regression in this training cohort resulted in a simplified version of ELF which contained TIMP1, HA, and PIIINP but did not have the age variable, as age became non-significant in the analysis at the 5% level. It was not possible to simplify the panel further without diminishing diagnostic accuracy. Furthermore it was not possible to improve performance through the addition of any of the other direct markers or simple biochemical tests. The simplified serum marker panel is presented below (ELF). Both published ELFa panel (original published panel) and ELF panels are reported for comparison in all external validation studies. The algorithms for use in both Ishak (7 level staging) and Scheuer (5 level staging) are presented (Figure 4.2).

Figure 4.2 Panel algorithms for published ELFa (original *European* Liver fibrosis panel) and simplified ELF (*Enhanced* Liver Fibrosis panel)

ISHAK classification of biopsy (F0-3 vs. 4-6)

i. ELFa (published) = -6.26 - (ln (age)*0.08) + (ln (HA)*0.608) + (ln (P3NP)*0.601) + (ln(TIMP1)*0.511)

ii. ELF (no age)= -8.468 + (ln (HA) *0.892) + (ln(P3NP)* 0.759) + (ln (TIMP) * 0.410)

SCHEUER (METAVIR) classification of biopsy (F0,1 vs. 2-4)

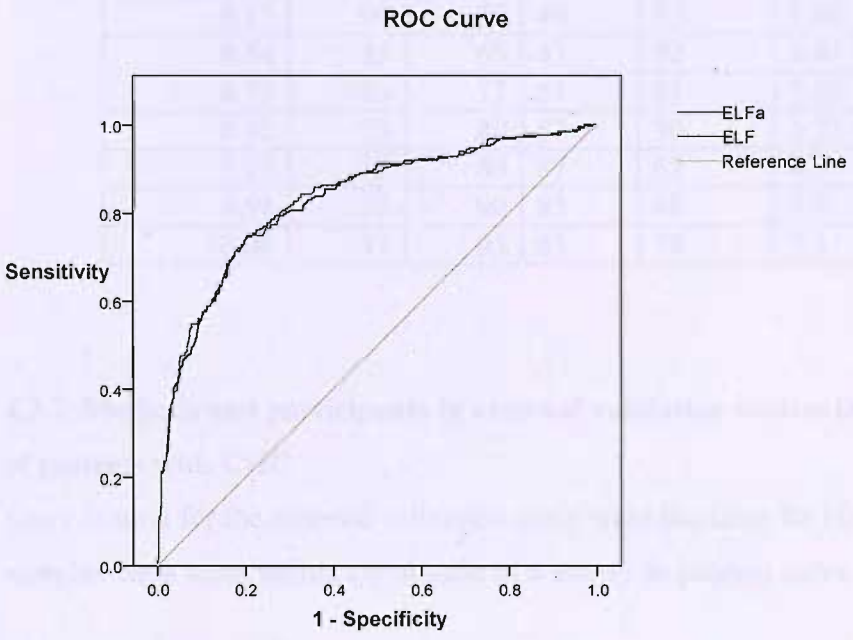
i. ELFa (published) = -6.38 - (ln(age)*0.14) + (ln(HA)*0.616) + (ln(P3NP)*0.586) + (ln(TIMP1)*0.472)

ii. ELF (no age) = -7.412 + (ln(HA)*0.681) + (ln(P3NP)*0.775) + (ln(TIMP1)*0.494)

Equations for ELFa (original published ELF panel) and ELF (simplified ELF panel) for Ishak and Scheuer histology classifications

The diagnostic performances of published ELFa and ELF in the training cohort were similar (Figure 4.3) in the exclusion of any or mild fibrosis and the identification of moderate/severe fibrosis and cirrhosis, producing very similar ROC curves.

Figure 4.3 Performance of the published ELF panel algorithm (ELFa), and a simplified panel derived in whole training set (ELF) (n=921) identifying moderate/severe fibrosis (F0-3 vs. 4-6).



	AUC	95% Confidence Interval	
		Lower	Upper
ELFa	.83	.80	.86
ELF	.83	.80	.86

In Table 4.7 the sensitivities, specificities and predictive values of the Enhanced Liver Fibrosis panel values in identifying severe fibrosis are reported. ELF scores above 8.54 would rule in any fibrosis at a sensitivity of 85% and below 9.27 would rule out fibrosis scores with specificity of 85%.

Table 4.7 Diagnostic performance indices for Enhanced Liver Fibrosis (ELF) panel in original cohort in identification of severe fibrosis (Ishak 4-6).

ELF	Sens	Spec	PPV	NPV	LR+	LR-
7.61	95	29	33	94	1.34	0.17
8.15	90	50	40	93	1.80	0.20
8.54	85	65	47	92	2.43	0.23
8.73	80	72	51	91	2.86	0.28
8.95	75	80	57	90	3.75	0.31
9.27	65	85	87	62	4.3	0.41
9.58	57	90	85	68	5.7	0.48
10.06	47	95	83	78	9.4	0.58

4.3.2. Methods and participants in external validation studies in independent populations of patients with CHC

Entry criteria for the external validation study were the same for all populations. Serum samples were taken within six months of a biopsy in patients naïve to treatment.

Cohort 1 was derived from a population based study of all patients who had had CLD newly diagnosed in all gastroenterology offices in three counties in the USA from 1999-2001. Serum markers were assayed from patients with CHC who had undergone biopsy (362). Of these, 110 had a biopsy performed within 6 months of the sample and 87 of these were naïve to treatment.

Cohort 2 comprised 171 prospectively recruited treatment naïve patients with CHC from the Trent Health Region (UK) HCV cohort study. The Trent cohort study began in 1995 with the aim of evaluating the natural history and epidemiology of Hepatitis C. Patients were recruited from 7 hospital out-patient clinics in one health region.

Cohort 3 was a retrospectively recruited cohort of 87 naïve patients with CHC in a single tertiary hepatology centre with available biopsy and serum samples taken within 6 months of each other (Southampton).

Histology

Biopsies were classified using the METAVIR (cohort1) and Ishak (cohorts 2, 3) scoring systems. In order to facilitate comparisons between cohorts both the actual staging is given and a clinical meaningful grouping presented of no fibrosis, mild, moderate, and severe. The groups were as follows: "no fibrosis" (METAVIR F0 or Ishak 0), mild fibrosis (METAVIR F1 or Ishak 1, 2), "moderate fibrosis" (METAVIR F2 or Ishak Stage 3), and severe fibrosis/cirrhosis (METAVIR F3, 4 or Ishak Stages 4-6). The Scheuer algorithm was used in cohort 1 which presented histology in a 5 stage classification (METAVIR) - both of which are analogous in fibrosis specification.

The baseline characteristics of patients in the three cohorts are presented in Table 4.8.The median ages are similar to the original cohort. Although broadly similar in fibrosis severity there were differences between cohorts. Cohort 3 had a greater prevalence of severe/cirrhosis than cohort 2 and 1 (35% vs. 28% vs. 24%) and overall cohort 2 had more moderate/severe fibrosis (64% vs. 56% cohort 1 and 3). Cohort 3 was the only single centre study.

Table 4.8 Patient characteristics in external validation cohorts

	CHC Cohort 1 (n=87)	**CHC Cohort 2 (n=173)**	**CHC Cohort 3 (n=87)**
Median age (range)	45 (25-72)	44 (23-71)	42 (18-78)
Setting	USA: All Gastro enterology offices in 3 US counties	England: Secondary hepatology outpatient clinics	England: Single tertiary hepatology centre
Recruitment method at entry	Prospective cohort study of patients with newly diagnosed CLD	Prospective recruitment to regional HCV Register	Consecutive retrospective
Histology classification	METAVIR	Ishak	Ishak
Staging levels (%)	n % 0 9 10 1 29 33 2 28 32 3 9 10 4 12 14	n % 0 36 21 1 26 15 2 29 17 3 22 13 4 12 7 5 19 11 6 27 16	n % 0 1 1 1 14 16 2 24 28 3 18 21 4 7 8 5 14 16 6 9 10
Severity of fibrosis (%) Nil Mild Moderate Severe	 10% 33% 32% 24%	 21% 15% 36% 28%	 1% 44% 21% 35%

ISHAK mild= 1, 2: Mod=3: severe/cirrhosis =4, 5, 6 METAVIR mild=1: moderate 2: severe/cirrhosis =3, 4

4.3.3. Methods and participants in external validation studies in NAFLD

Cohorts 1 and 2 (UK centres)

Patients were recruited consecutively from tertiary out-patient liver centres in the UK, in Nottingham and Newcastle-upon-Tyne. All patients had an index liver biopsy and serum taken within three months. Alternative causes of liver disease were excluded by standard clinical history, examination and blood tests. All patients had a weekly ethanol consumption of less than 140 g in women and less than 210 g in men. The following clinical measurements were obtained weight circumference, hip circumference and body mass index (BMI). Serum samples were obtained for routine liver chemistry (including alanine transaminase (ALT), gamma glutamyl transpeptidase, bilirubin, albumin, and alkaline phosphatase), full blood count, measures of insulin resistance (including fasting glucose, insulin and c peptide), ferritin, total cholesterol, HDL, LDL and triglycerides. Serum samples were analysed for levels of TIMP-1, HA and PIINP at an independent reference laboratory (iQur Limited, Southampton, UK) and ELFa and ELF scores were calculated using the Scheuer algorithm for 5 stage classification of histology.

Liver biopsy

Liver biopsies were assessed by two hepato-pathologists, one at each centre. Biopsies were graded for fibrosis using a five stage classification system for fibrosis that has recently been published by National Institute of Diabetes and Digestive and Kidney Disease[42]; stage 0= absence of fibrosis, stage 1 = perisinusoidal or portal, stage 2= perisinusoidal and portal/periportal, stage 3= septal or bridging fibrosis and stage 4= cirrhosis.

Paired serum and histological data were available for 192 subjects. The baseline characteristics of these patients are shown in table 4.9. The demographic data were similar for the two populations; 64 % of subjects were male, the mean age in the study was 49 years and 63 % of subjects had evidence of the metabolic syndrome.

Table 4.9 Baseline patient characteristics in NAFLD individual & combined cohorts

Category	Nottingham centre	Newcastle centre	Entire cohort
Number	88	104	192
Age (years)	50.4 +/- 11.5	47.3 +/- 11.1	48.7 +/- 12.5
Male subjects	65 %	63 %	64 %
BMI (kg/m^2)	30 +/- 4.5	34.4 +/- 5.9	32.4 +/- 5.7
Waist (cm)	104.5 +/- 12.5	111.2 +/- 12.7	107.8 +/- 13
Metabolic syndrome (yes)	66 %	60 %	63 %
Fasting Glucose (mmol/l)	6.0 +/- 1.7	6.5 +/- 3.3	6.3 +/- 2.7
Triglycerides (mmol/l)	2.1 +/- 1.6	2.8 +/- 1.8	2.5 +/- 1.8
HDL (mmol/l)	1.4 +/- 0.42	1.1 +/- 0.28	1.2 +/- 0.4
ALT (U/L)	76.1 +/- 48.9	78.4 +/- 64.6	77.3 +/- 57.8
GGT (U/L)	140 +/- 135	104 +/- 102	121 +/- 119.5
Albumin (g/L)	43.7 +/- 3.4	44.9 +/- 4.9	44.3 +/- 4.3
Fibrosis stage			
0	32 %	49 %	41 %
1	18 %	19 %	19 %
2	27 %	8 %	17 %
3	15 %	12%	13 %
4	8 %	12%	10 %

Values in mean +/- standard deviation unless stated

Cohort 3 Methods and participants in external validation study in a paediatric population with NAFLD

Consecutive patients diagnosed with NAFLD according to accepted criteria (Nobili V et al. Hepatology 2006) seen in the paediatric department in Institute Bambino Gesu Opitale, Roma between June 2004 to June 2006 were included in the study. All patients were referred with elevated aminotransferase levels associated with diffusely echogenic liver in imaging studies suggestive of fatty infiltration of the liver. Ultrasonography is the most commonly used modality for evaluating hepatic steatosis in children. The diagnosis of NAFLD was confirmed with a percutaneous liver biopsy in all cases. All patients were HCV RNA-PCR negative.

Secondary causes of steatosis including alcohol abuse (≥ 140 g/week), total parenteral nutrition, and the use of drugs known to precipitate steatosis (e.g. valproate, amiodarone or prednisolone) were excluded in all cases. Hepatitis A, B, C, D, E and G, cytomegalovirus and Epstein-Barr virus infections were ruled out by appropriate tests. In all cases, autoimmune liver disease, metabolic liver disease, Wilson Disease, and α-1-antitrypsin deficiency were ruled out using standard clinical and laboratory evaluation as well as through histological examination of the liver biopsy.

Paired samples of serum and histology were evaluated to determine the performance of ELF panel in estimating fibrosis severity assessed by liver biopsy.

Liver Histology. Biopsies were performed in all children using an automatic core biopsy device (Biopince, Amedic, Sweden). Liver biopsies were at least 15 mm long and were read by a single blinded liver pathologist. Histology was scored using the Modified Brunt classification: 0 No fibrosis, 1 perisinusoidal/periportal fibrosis, (1a mild zone 3 perisinusoidal, 1b moderate zone 3 perisinusoidal, 1c portal/periportal) 2 perisinusoidal +portal/periportal, 3 bridging fibrosis 4 cirrhosis.

The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was performed according to the recommendations of the Ethics Committee of Children's Hospital and Research Institute Bambino Gesù. Informed consent was obtained from each patient or a responsible guardian.

Baseline characteristics are presented in Table 4.10. The mean age was 11.7 years (range, 3-17.8) with a predominance of male (n = 72, 64%). The mean ALT level was 76 IU/L (range 10-407 IU/L), and the mean AST level was 48 IU/L (range 19-187 IU/L) with all children been evaluated and the liver biopsy performed due to persistently elevated ALT and/or AST. Insulin resistance as indicated by a HOMA-IR > 2, or ISI-comp < 6 was present in 51 (60.7%) and 62 (73.8%) children, respectively, with 66 (79%) of children having either HOMA-IR > 2, or ISI-comp < 6. Obesity was present in 34 (30.5%) subjects including 23 male and 11 female, and 23 (67.6%) of them were insulin resistant as indicated by ISI or HOMA-IR parameters. A significantly higher proportion of children without obesity were insulin resistant as compared to children with obesity [43/50 (86%) vs. 23/34 (67.6%) respectively, p = 0.044].

Table 4.10 Baseline characteristics of the patients in external validation study (cohort 3) (n = 112)

	Mean ± SD or n (%)	Range
Age (years)	11.7 ± 3.3	3 - 178.8
Gender (M/F)	64/48	
BMI (kg/m^2)	26.3 ± 3.7	15.2-38.4
Obesity	34 (30.5%)	
Type II diabetes	2 (2.3%)	
Hypertension	2 (2.3%)	
AST (IU/L)	48 ± 26	19-187
ALT (IU/L)	76 ± 63	10-407
AST/ALT ratio	0.8 ± 0.4	0.4-2.3
AST/ALT ration > 1	16 (19%)	
GGT (IU/L)	26 ± 21	10 – 130
Cholesterol (mg/dL)	156 ± 35	75-243
Hypercholesterolemia	9 (10.7%)	
Triglyceride (mg/dL)	95 ± 55	28-351
Hypertriglyceridemia	6 (7.1%)	
Fasting glucose (mg/dL)	82 ± 12	60-138
Fasting insulin (μU/L)	11.8 ± 6.1	3.5 - 30.7
HOMA-IR	2.69 ± 1.35	0.7 - 6.7
HOMA-IR > 2	51 (60.7%)	
ISI-comp	4.4 ± 2.0	1.2 - 9.1
ISI-comp < 6	62 (73.8%)	
Histology (fibrosis)		
0	37 (33)	
1a	8 (7)	
1b	6 (5)	
1c	44 (39)	
2	9 (8)	
3	6 (5)	
4	2 (2)	

4.3.4. Methods and participants in external validation studies in PBC

161 patients with PBC were prospectively followed between 1993 and 2003 as part of a multi-centre US clinical trial that was designed to investigate whether low dose weekly methotrexate, when added to ursodiol, improved survival or delayed progression of PBC. Methotrexate was not found to affect the course of PBC, so patients from both treatment arms were combined for the purpose of this study[166]. Only patients with established, but not decompensated PBC were enrolled into the parent trial. Participants were required to have both a positive anti-mitochondrial antibody as well as either an abnormal alkaline phosphatase or at least stage 1 disease on liver biopsy. Patients with a history of variceal bleeding, ascites, or encephalopathy were excluded. Of the 161 patients 147 had biopsy and ELF values at baseline.

Baseline values were taken at 0 or 2 years after enrolment for single individuals.

Histology was scored by a single pathologist using the Ishak classification.

The diagnostic performance of ELF panel in the identification of fibrosis stage derived from liver histology was assessed using AUC. Different levels of severity of fibrosis were evaluated, (i) any fibrosis (0 vs. 1-6) ; mild fibrosis (0,1 v 2-6) , moderate/severe fibrosis (0-2 vs. 3-6), severe fibrosis (0-3 vs. 4-6), and cirrhosis (0-4 vs. 5,6. Sensitivities, specificities and predictive values were derived. All analyses were performed using the SPSS software package version 14 (SPSS Inc. Chicago, Il.).

Baseline median age was 53 years (27-70). Prevalence of severity of fibrosis (Ishak) grade 0- 6%, grade 1 -18%, grade 2 - 35%, grade 3- 22%, grade 4 -13%, grade 5 -5%, grade 6- 2%. 110 patients had baseline ELF and biopsy at 0 years and 37 at 2 years after enrolment.

4.3.5. Methods and participants in external validation studies in HCV-HIV co-infection

Patients with HCV-HIV co-infection naïve to treatment for CHC, were recruited into a therapeutic prospective study at 21 sites in US[167]. They all had a biopsy within 48 weeks pre-study. Exclusion criteria included decompensated liver disease. Subjects were randomly assigned to receive 2 regimes of combination therapy with interferon and ribavirin. Baseline serum and biopsy were assessed for ELF markers.

Baseline characteristics

133 subjects were enrolled into the study between December 2000 and June 2001. 85% were male, the mean age was 44.5, and 86% were receiving stable HIV medication. 20% had severe fibrosis (Ishak 4-6) Table 4.11.

Table 4.11 Baseline histology in patients with HCV-HIV co-infection

Ishak staging	Number	%
0	6	4.6
1	35	26.7
2	33	25.2
3	31	23.7
4	13	9.9
5	9	6.9
6	4	3.1

4.3.6 Method of analysis used in all independent cohorts to externally validate ELF panel

TIMP-1, P3NP, and HA (ELF test) were assayed on an automated IMMUNO 1 immunoanalyser (Siemens Medical Solutions Diagnostics, Tarrytown, NY, USA).

The assays are magnetic particle separation immunoassays and were identical to those used for the 2004 European Liver Fibrosis study. The TIMP-1 and P3NP assays each use two monoclonal antibodies (MAbs) that bind to independent binding sites on their respective antigens. The HA assay uses hyaluronic acid binding protein (HABP), which is isolated from cow nasal septum, in the place of MAbs.

ELF markers were analysed in singular and the results continually referred to a set of quality standards to ensure accurate analysis. The ELF assays require a total of 22.2 μL of serum for a single determination: 3.5 μL for HA, 15.0 μL for P3NP and 3.7 μL for TIMP1.

The diagnostic performance of ELF panel in the identification of fibrosis stage derived from liver histology was assessed using AUC. Different levels of severity of fibrosis were evaluated using the Ishak staging, (i) any fibrosis (0 vs. 1-6) ; mild fibrosis (0,1 v 2-6) , moderate/severe fibrosis (0-2 vs. 3-6), severe fibrosis (0-3 vs. 4-6), and cirrhosis (0-4 vs. 5,6).

Statistical analysis

ELF scores were compared to histological staging of liver biopsies from corresponding patients and the sensitivity and specificity of the ELF score for detecting fibrosis was calculated. These results were then used to plot ROC curves and the AUC was calculated. PPV and NPV values and LRs for detecting different degrees of severity of fibrosis were also calculated. Sensitivities, specificities and predictive values were derived. All analyses were performed using the SPSS software package version 14 (SPSS Inc. Chicago, Il.).
In order to compare performances, both the published ELF panel (ELFa) and any simplified panel (ELF) were assessed in the external validation cohorts.

4.3.7 Results

A summary results table of diagnostic performance (AUC) of ELFa and ELF in all external validation cohorts is presented in Table 4.12. The prevalence of severe fibrosis varied between external cohorts, with two of the CHC cohorts being greater than training (34% vs. 27%). Co-infection PBC and NAFLD were lower (19% vs. 27%) and the paediatric cohort lowest of all (7.2%).
In all cohorts ELFa and ELF had similar performance in the identification of all levels of severity of fibrosis. Thus the ELF panel can be simplified without loss of diagnostic performance (see figure 4.3 for ROC curves).
ELF had a better diagnostic performance in the identification of significant fibrosis or severe /cirrhosis, with AUC values consistently higher in all cohorts for the identification of Ishak stages 3-6 and 5-6. In the majority of cohorts ELF has AUC values >0.80 for the identification of significant fibrosis and for cirrhosis.
ELF panel performance as assessed by ROC analysis was as good if not better than that in original training cohort.

NAFLD

ELF had a high diagnostic accuracy in all cohorts but was particularly good in NAFLD. The ELF panel had an excellent performance in distinguishing severe fibrosis in patients with NAFLD with an AUC of 0.91 and a threshold of 10.3576 was associated with a sensitivity of 80 %, specificity of 90 %, positive predictive value of 71 % and a negative predictive value of 94 %. In distinguishing moderate fibrosis the overall AUC was 0.82 and a threshold of -

10.1068 was associated with a sensitivity of 70 %, specificity of 80 %, positive predictive value of 70 % and a negative predictive value of 80 %. In distinguishing no fibrosis the overall AUC was 0.76 and a threshold of 9.793 was associated with a sensitivity of 61 %, specificity of 80 %, positive predictive value of 81 % and a negative predictive value.
Sensitivity and specificity were derived for all cohorts at each cut of fibrosis. In all cohorts ELF panel worked best at a high and low threshold either with a high sensitivity and lower specificity or high specificity and lower sensitivity (PPV and lower NPV or high NPV and lower PPV). This confirms its performance in the original study (and that of other panel markers in the systematic reviews presented in chapter 3. DOR and LR are presented and showed that ELF has a good diagnostic performance (Appendix 5 Table 1).

Table 4.12 Summary of AUC (95% CI) for ELFa (original) and ELF (simplified) in 8 external validation cohorts

Fibrosis stage	Master training cohort n=921		Trent CHC n=171		Southampton CHC n=87		CDC CHC* n=97		NAFLD combined n=192		Paediatric NAFLD n=112		PBC n= 147		HCV-HIV Co-Infection n=131	
	ELFa	ELF	ELFa	ELF	ELFa	ELF	ELFa	ELF	ELFa	ELF	ELF a	ELF	ELFa	ELF	ELFa	ELF
Prev 456 (%)	26.9		34		34.4		23.3		19.5		7.2		19		19.9	
Prev any fibrosis (%)	75.4		79.2		99		89.5		57.2		67		94		94.3	
0 v 1-6	0.72	0.72	0.78 (0.70, 0.86)	0.79 (0.71, 0.86)	n/a	n/a	0.71 (0.53 0.89)	0.71 (0.53, 0.89)	0.78 (0.72, 0.85)	0.78 (0.72, 0.85)	n/a	0.92 (0.86, 0.97)	0.74 (0.57, 0.91)	0.72 (0.54, 0.89)	0.86 (0.75, 0.97)	0.86 (0.74. 0.97)
0,1 v 2-6	0.77	0.77	0.79 (072, 0.86)	0.80 (0.73, 0.87)	0.69	0.69	0.77 (0.67, 0.86)	0.77 (0.67, 0.87)	0.84 (0.78, 0.90)	0.85 (0.79, 0.91)	n/a	0.99 (0.97, 1.0)	0.72 (0.63, 0.81)	0.71 (0.61, 0.80)	0.84 (0.76. 0.91)	0.83 (0.76, 0.91)
0-2 v 3-6	0.79	0.79	0.83 (0.76, 0.89)	0.83 (0.76, 0.89)	0.87 (0.78, 0.95)	0.87 (0.80, 0.95)			0.90 (0.84, 0.96)	0.90 (0.85, 0.96)	n/a		0.76 (0.68, 0.84)	0.76 (0.67, 0.84)	0.81 (0.73, 0.88)	0.78 (0.70, 0.86)
0-3 v 4-6	0.83	0.83	0.86 (0.81, 0.92)	0.87 (0.81, 0.92)	0.90 (0.83, 0.96)	0.89 (0.83, 0.96)	0.85 (0.76, 0.94)	0.85 (0.76, 0.94)	0.93 (0.90, 0.97)	0.93 (0.89, 0.97)	n/a	0.99 (0.97, 1.0)	0.84 (0.76, 0.92)	0.83 (0.75, 0.91)	0.84 (0.77, 0.92)	0.84 (0.75, 0.92)
0-4 v 5,6	0.86	0.86	0.88 (0.82, 0.93)	0.88 (0.82, 0.93)	0.89 (0.82, 0.96)	0.89 (0.82, 0.96)	0.90 (0.82, 0.98)	0.90 (0.82, 0.98)	0.91 (0.86, 0.95)	0.91 (0.86, 0.95)	n/a	n/a	0.86 (0.77, 0.94)	0.85 (0.76, 0.94)	0.77 (0.66, 0.88)	0.77 (0.65, 0.88)

Figure 4.4: ROC curves of ELF panel identifying serious fibrosis (Ishak 4-6) with and without age (a) NAFLD (b) PBC

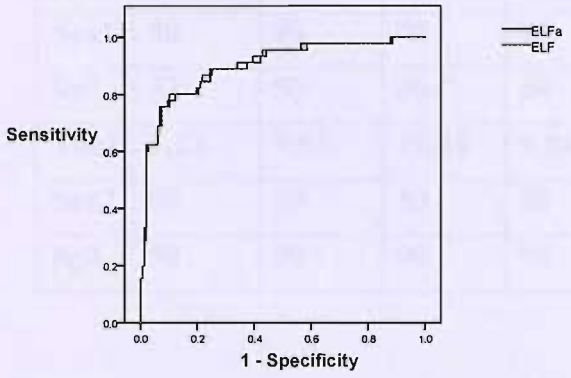


Figure 4(a)

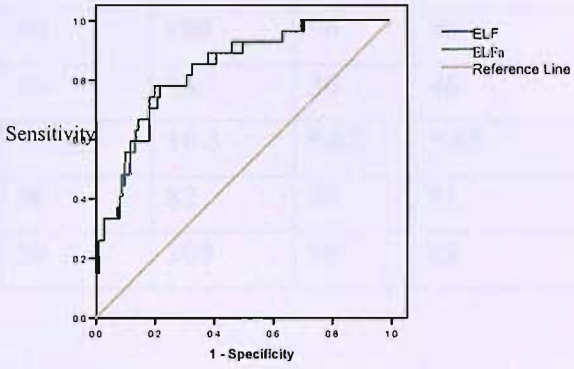


Figure 4(b)

ROC curves of the performance of ELFa (original published ELF) and ELF (simplified ELF) in cohorts of NAFLD(a) and PBC (b) showing that their performance is very similar

High and low thresholds at which the sensitivity was ~90% and another where specificity was ~90% in each cohort were compared to the thresholds from the original ELF cohort (Table 4.13). Thresholds differed between the original training cohort and those in the external validation cohorts, where all of the thresholds were higher.

Conversely those thresholds derived in the original ELF cohort for certain values of sensitivity and specificity do not yield the same results for diagnostic accuracy with the equivalent thresholds in the external cohorts (data not presented).

Table 4.13 Comparison of thresholds to identify significant fibrosis (Ishak 3-6/Sheuer 2-4)

	Bayer	CHC 1	CHC 2	CHC 3	NAFLD (1&2)	NAFLD (3)	PBC	HCV-HIV co-infection
Thr1	7.53	8.92	9.03	8.37	9.33	10.09	8.3	8.51
Sen1	90	90	90	95	90	100	90	90
Sp1	32	53	35	54	50	88	30	46
Thr2	9.12	9.93	10.18	9.84	10.31	10.3	9.82	9.85
Sen2	50	60	53	50	56	82	53	51
Sp2	90	90	90	90	90	100	90	89

Additional analyses

Using NAFLD as an example, additional analyses were conducted to show:

(i) the development of a clinical utility models which showed

The use of high and low ELF thresholds and the balance of false negatives/false positive results with those correctly allocated and those patients who could not be given an ELF value at a particular predictive value.

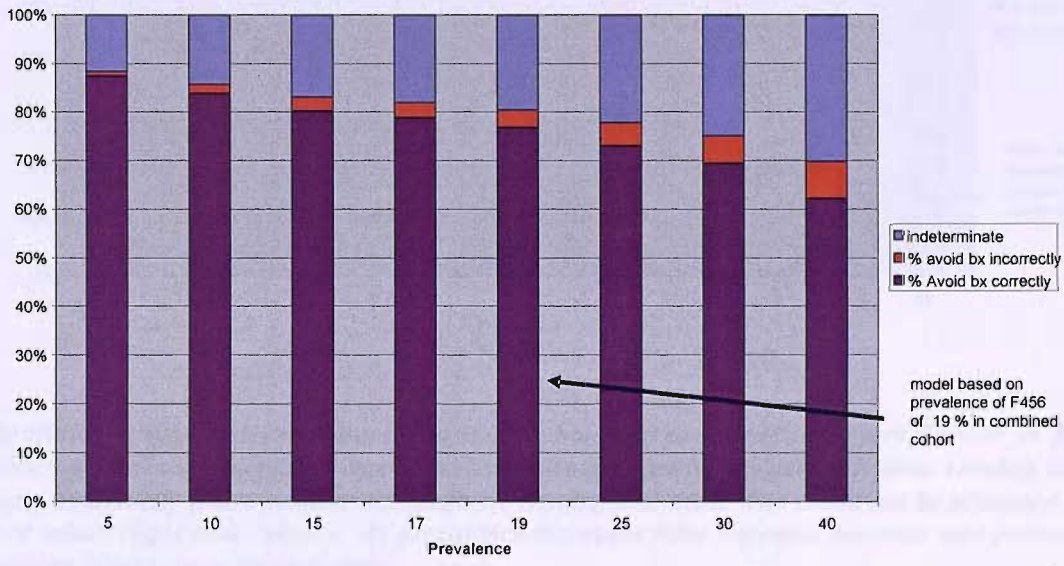
(ii) how ELF performance may vary with different prevalence of fibrosis

Development of a clinical utility model

The clinical utility model in figures 4.4 and 4.5 shows the number of patients who could avoid a liver biopsy if only extreme thresholds were used to determine who should have a biopsy. For the diagnosis of severe fibrosis using a high threshold and a low threshold, 77 % patients would avoid a liver biopsy (true positives and true negatives), 4 % of patients would incorrectly avoid a biopsy (false positives and false negatives) and 19 % of patients would have indeterminate scores requiring consideration of a biopsy. Predictive values are strongly influenced by the prevalence of the underlying disease therefore we have modelled what would happen if the prevalence of severe fibrosis increases or decreases (see figure 4). With increasing prevalence, both the number of patients incorrectly assigned and the number of indeterminate biopsies would increase.

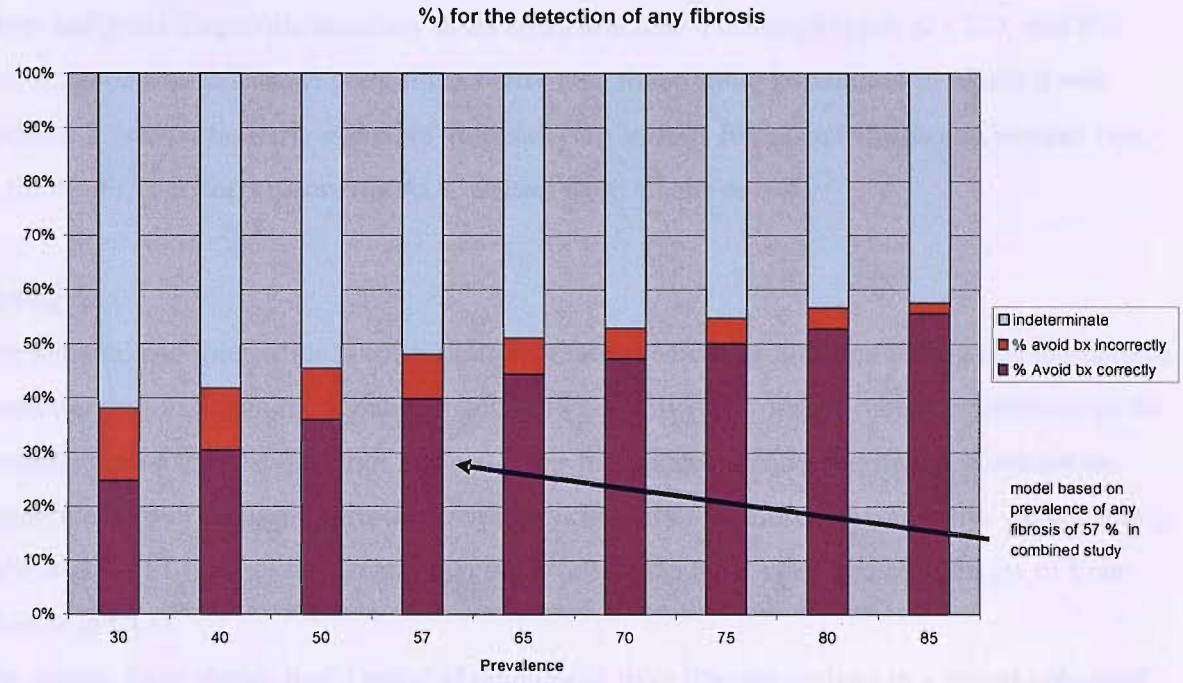
Using this clinical utility model for the diagnosis of any fibrosis (Stage1-6), 40 % of patients can avoid a biopsy correctly, 8 % will incorrectly avoid a biopsy and 52 % will have indeterminate scores requiring consideration of liver biopsy. Figures 5 and 6 illustrate the influence of prevalence on these figures. With increasing prevalence the both number of incorrectly assigned biopsies and indeterminate biopsies would reduce in number.

Figure 4.5 Clinical Utility models showing effect of varying prevalence for the detection of F4-6 fibrosis in the combined NAFLD cohort (at threshold of 95 % for NPV and PPV)



The clinical utility model showing the effect of change in prevalence of severe fibrosis at fixed predictive values on the percentage of patients who avoided biopsy correctly (true results), avoided biopsy incorrectly (false positive and negative results) and those who could not be allocated at these fixed values (blue colour). As prevalence increases the proportion that cannot be allocated increases.

Figure 4.6 Clinical Utility models showing effect of varying prevalence for the detection of F1-6 (any) fibrosis in the combined NAFLD cohort (at threshold of 95% for NPV and 80% PPV)



The clinical utility model showing the effect of change in prevalence of severe fibrosis at fixed predictive values on the percentage of patients who avoided biopsy correctly (true results), avoided biopsy incorrectly (false positive and negative results) and those who could not be allocated at these fixed values (light blue colour). As prevalence increases false test rates decrease and proportion avoiding biopsy correctly increases

4.4 Discussion

The use of the whole of the original ELF cohort as a training set allowed the development of a simplified version of ELFa. This was shown to maintain its performance compared to the original panel in 8 independent cohorts of patients with CLD. The Enhanced Liver Fibrosis panel had good diagnostic accuracy in all cohorts across 4 different types of CLD, and this performance was at least as good if not better than the training population in which it was derived. It was particularly effective in identifying serious fibrosis/cirrhosis and worked best in the NAFLD cohorts where the AUC values were all above 0.80.

Strengths

The national and international collaborations that have been established during this thesis have created access to 8 cohorts of patients with CLD in which to conduct robust evaluations of the performance of the ELF panel of markers. This has added considerably to the literature on serum markers in predicting fibrosis on biopsy. In particular these 8 studies have substantively validated the ELF panel of markers making it one of the most validated biomarkers of liver fibrosis in CLD.

The studies have shown that a panel of markers of liver fibrosis derived in a mixed cohort of patients with CLD (predominantly CHC) maintained its performance in single aetiology cohorts. This broadens its generalisability.

The evaluation of ELF in a large paediatic population will add to the scant literature on non-invasive markers in children. With the rise of childhood obesity it is becoming important to identify those children who are at risk of developing significant and severe fibrosis/cirrhosis due to obesity. It is alarming that such a large cohort could be assembled fairly rapidly over 18 months, and that here was such a high prevalence of significant fibrosis in children whose median age was 13 years.

Limitations

The limitations of liver biopsy as a reference standard may lead to the misclassification of ELF. This is likely to be a blunting of association and therefore any diagnostic performance shown is probably an underestimate. This is discussed in more detail in Chapter 6.

The studies are relatively small (all under 200) leading to reduction in precision of estimates of effect. However both co-infection and PBC are not the most common of CLD, and recruitment takes longer to achieve and may require multi-centre participation in order to attain sufficient numbers to power studies, which are more costly and require more complex administration. Combining several smaller studies in a meta-analysis may be one solution. However problems arise when the inclusion criteria of the studies are dissimilar, different histology classifications were used and patients may be on treatment which may affect staging if biopsy and serum are not taken at the same time point. Two of the NAFLD cohorts could be combined as the methods of recruitment were similar and the methods of classification of histology compatible.

In the PBC cohort the paired samples of serum and biopsy were taken at baseline and 2 years after the start of the study. This was done to increase the power of the study. A sensitivity analysis was performed with only those patients with baseline values at zero time (n=110) considered in the analysis compared to the analysis where all patients are included (n=147). ELF was shown to have greater diagnostic accuracy in the identification of mild/significant fibrosis (stages 2-6) or any fibrosis (1-6) when only patients at time zero were included. AUC values in moderate/severe fibrosis were the same when patients at time zero and at zero and 2 years were considered. (See Appendix 5 for analysis). For *any* fibrosis –baseline and 2 years AUC=0.72 (95% CI 0.55, 0.86) and for baseline AUC =0.80 (0.57, 1.0), and for mild/moderate/severe fibrosis AUC baseline and 2 years =0.71 (0.60, 0.78) and baseline alone AUC=0.75 (0.65, 0.86). This difference may reflect the change of fibrosis over time leading to a bias in the results.

Such progression/regression of fibrosis over time may have affected the results in the HCV-HIV co-infection as the inclusion criteria permitted biopsy results up to 48 weeks prior to study entry and the first serum sample. However, most biopsies were taken within 6 months or at study entry reducing this potential bias.

Both PBC and HIV-HCV co-infection cohorts consisted of patients recruited to treatment trials whose pre-treatment baseline tests were used in the analyses. In the case of PBC the treatment was not effective and both treatment arms were used in the study. There may have been selection bias in this participant selection reducing its generalisability to hepatology

clinical settings, where patients with PBC and co-infection are not subjected to strict exclusion criteria.

In these external studies there were limited opportunities for direct comparison of other markers with ELF in predicting fibrosis. It was possible to compare ELF and the Mayo clinical scale in 141 patients with PBC. ELF had consistently higher AUC values- for severe fibrosis AUC for Mayo scale was 0.71 vs. 0.85 ELF; for cirrhosis 0.73 vs. 0.85. There have been no other direct comparison of ELF in the published literature, although indirect comparison was conducted using different populations where ELF was shown to have comparable AUC value (0.834±0.037) for the identification of significant fibrosis (METAVIR stages 2-4) with other pane such as APRI (0.82±0.037, Fibrotest 0.87±0.034) and Fibrometer 0.89±0.029)[148] . This comparison has limitations in that it was not conducted in the same population under the same conditions and using uniform standards of testing. Such a drawback can be addressed by using a reference population derived from many centres, all of which would have the same inclusion/exclusion criteria and study methodology in which to directly compare all of the biomarkers which are currently in the public domain as non-invasive markers of liver fibrosis. Such an approach is an urgent research imperative for researchers, clinicians and patients interested in CLD.

As outlined in Chapter 3 there have been external validations in other biomarkers the most extensive being with APRI and Fibrotest. The comparison of APRI performance with other biomarkers including those using markers of extra cellular matrix formation and breakdown (Hepascore; Fibrometer) found comparable performance for significant fibrosis for all markers[168]. All of these comparisons have been in CHC and there are few comparisons in other aetiologies. Most of these individual studies have been conducted in populations <200. (See Appendix 4 for components of the marker panels).

The performance of ELF in CHC in the external studies was better than the summative values derived from the APRI SROC analysis, with median AUC for significant fibrosis being 0.85 (range 0.83-0.87) and for cirrhosis 0.89 (range 0.88-0.90). In the CHC external validations of ELF, the impact of the sensitivity, specificity and predictive values can be modelled in a theoretical cohort of 1000 patients with hepatitis C in a hospital outpatient setting. Permitting 10% false negative and positive test rates only 23% of patients in a theoretical cohort of 1000

could not be reliably assigned to severe fibrosis/cirrhosis, 67% correctly avoid biopsy with a 9% false test rate (mostly false positive).

There have been several serum markers suggested in the identification of fibrosis stage in co-infection[169;170]. A recent comparison of such markers was conducted in a population of 272 patients with HCV-HIC co-infection comparing Fibrotest, SHASTA, Hepascore, Fibrometer, APRI, Forns index, FIB-4. The panels performed best at identifying cirrhosis, and those that containing ECM components did better than those with direct liver function markers, although differences were not significant. Results showed that for significant fibrosis AUC values were for Fibrometer 0.70 (95% CI 0.64, 0.76); Hepascore (95% CI 0.69 (0.63 0.74); Fibrotest 0.64 (95% CI 0.58, 0.70) and APRI 0.65 (95% CI 0.59, 0.71). Although not in the same population the diagnostic accuracy of the ELF panel appears to be better in the cohort of co-infection patients in which it was tested (AUC =0.81 (95% CI 0.73, 0.88)[171].

There are few small studies in PBC using non-invasive markers. AST/ALT ratio as been used in a retrospective study in 121 patients and the AUC for identifying cirrhosis was 0.87, with $\geq$1.1 threshold sensitivity was 82% specificity 79%[172]. This study was limited by the haphazard nature of the participants, who were derived from those with availability of paired biopsy and AST Alt values on patients retrieved from records over a 24 year period and also of the difficulties in external validity when using local laboratory assays. An HA-Bilirubin model was constructed in 153 patients undergoing treatment for PBC which had at one threshold a sensitivity of 64%, a specificity of 74% for the identification of cirrhosis, with 71% of patients being able to be allocated[173]. Three smaller studies evaluated serum markers of ECM production and found that HA and P3NP were associated with identification of severe fibrosis/cirrhosis[174-176] . Such early studies on the effectiveness of ELF component markers are consistent with the performance of ELF in PBC. More recent studies have been conducted on the performance if transient elastography in patients with PBC. One study in 55 people showed AUC values of 0.86 (95% CI 0.72-0.94) (for severe fibrosis –Scheuer stages 3-4 and 0.96 (95% CI 0.87, 0.99) for cirrhosis (Scheuer stage 4)) At a threshold of 14.7 kPa the sensitivity for detecting severe fibrosis was 56% and specificity 100%[177]. Problems associated with this technology are explored in Chapter 6 but include failure rate of the initial procedure

often due to obesity. In a study of patients with PBC (n=73) and Primary Sclerosing Cholangitis (n=28) this failure rate was reported as 6 %[178].

Performance in external validation cohorts is usually worse than performance in the population in which the panel/marker was derived. This is because panels are derived and their performance assessed using analyses that make the model fit the dataset in an optimal way. Independent datasets will differ in many respects to the original cohort (for example gender, age, distribution of disease severity aetiology of CLD), and consequently the panel does not fit this external population as well as that in which it was derived. This usually results in a reduction of diagnostic accuracy. This was not the case with ELF whose performance was as good if not better in all external validations. This may be because the original ELF cohort, being a large multi-centre study was truly representative of general hepatology practice that was similarly reflected in the external populations. It may reflect the universal biological applicability of the serum marker components in the fibrotic process or it may be the consistency of measurement of the components that were assayed using automated processes in two laboratories (central one for the original cohort and a different laboratory but using the same sort of analyzer, for all external cohorts). Or finally it may be that the chance differences that would often arise were favourably distributed –by chance.

ELF performed optimally in NAFLD. This confirms the result in the internal validation set in the original ELF cohort which found the highest performance in patients with NAFLD and ALD (although numbers were small). This could be due to the different pathophysiological process that results in fibrosis in these aetiologies, compared to CHC which has a significant inflammatory element. The markers in ELF could play a more important role in NAFLD and ALD whereas more inflammatory markers could add to the performance of ELF in CHC and other viral hepatitides. This needs further research elaboration.

There are few recent, good quality studies of the performance of serum markers in ALD. This may reflect clinical practice in the use of biopsy in patients with ALD, or the difficulties of recruiting this patient group into studies. Future research recommendations from the work presented in this thesis include expansion of the literature by conducting larger studies of

patients with ALD which can directly compare and validate in external populations, performance of existing markers, the identification of new markers or enhancement of existing tests to identify any, mild or moderate fibrosis.

There is some debate as to the threshold that should be used in evaluation studies. Each external study can generate a dataset driven set of thresholds at which sensitivity, specificity and predictive values within the prevalence of the included populations can be calculated. The thresholds for optimal discrimination may differ by CLD aetiology and by population. Use of a single set of thresholds would be desirable in order to be more convenient in the clinical setting, and to facilitate research using common diagnostic criteria. It may be that thresholds derived from a large population such as the original ELF cohort give sufficient precision to allocate patients with acceptable false test rates to broad categories of fibrosis such as any fibrosis, significant fibrosis or cirrhosis. It may be that different aetiologies such as CHC, NAFLD and ALD may need to have separately defined thresholds. This has been identified as a dilemma with other biomarker panels and further research is needed to clarify these issues.

Future research

There are exciting opportunities to carry on this validation research. This includes conducting a meta-analysis using summative receiver operator characteristic curves for all of the external validation cohorts using new advanced methods of analysis being developed by the Cochrane Diagnostic Methodology Committee. Research is needed to evaluate if ELF performance can be enhanced by the addition of simple markers such as those that measure inflammation (in CHC-CHB co-infection) or metabolic pathways (NAFLD). Initial work has been conducted and published using the NAFLD cohorts 1 and 2 producing a model that combined ELF and simple markers which had a better performance than each alone. This panel needs validation in an external population. Use of different modalities with ELF such as transient elastography and microbubble contrast ultrasound may add to their performance, although recent studies have begun to elaborate the limitations of transient elastography such as difficulty in getting reliable measurements in the obese and problems of interpretation in those with active inflammation.

The cost effectiveness of the use of serum markers in the clinical setting also needs to be addressed. Economic modelling studies using different panels of markers should be conducted.

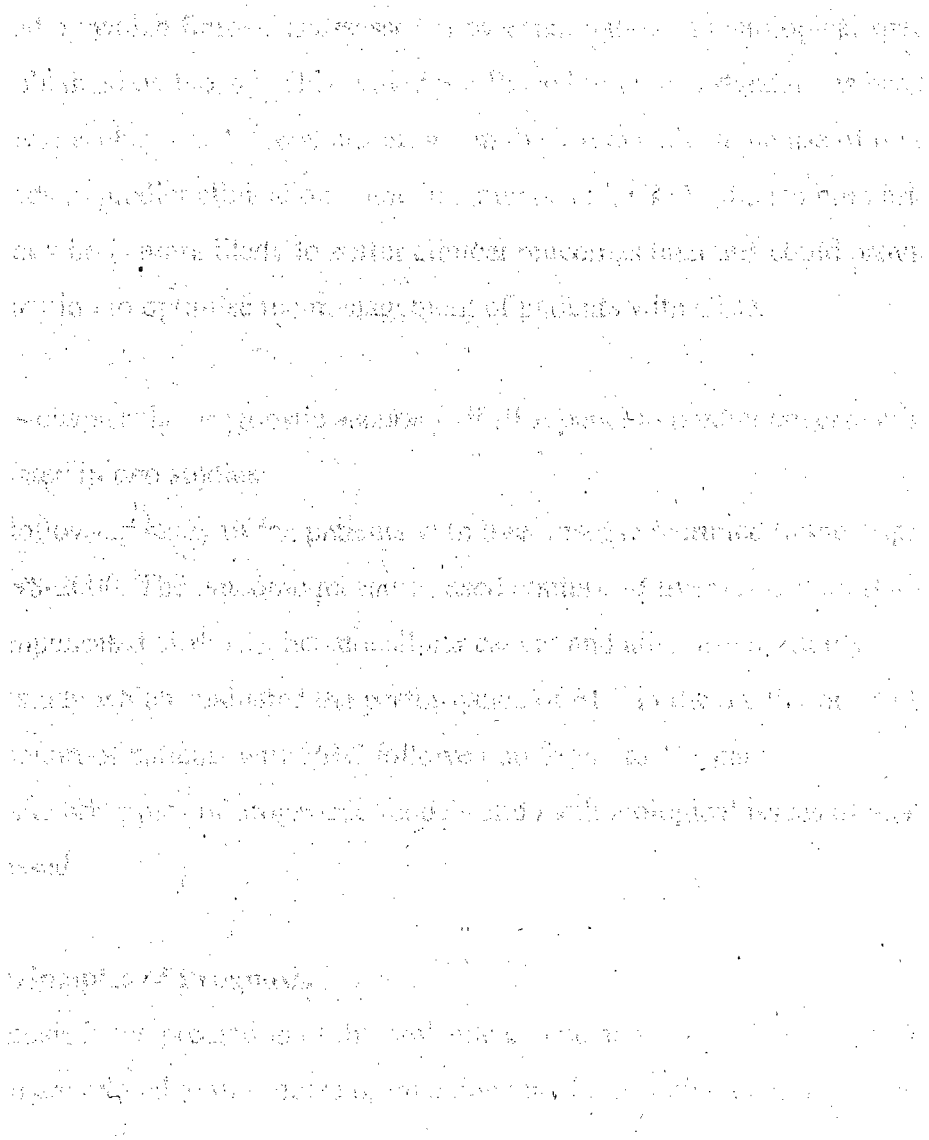
The use of serum markers in clinical practice is not yet endorsed by the National Institute of Clinical Excellence in England and Wales. There is considerable interest in their utility in clinical practice by the hepatology/gastroenterology community, and their use is seeping into routine care of patients with CLD in a haphazard way. This was the pattern observed in France where a survey of the use in practice by clinicians was conducted prior to national guidance. Although not a very high response rate (65%), the results of the French survey found that many physicians had incorporated serum markers into their practice but not in a systematic or uniform way[179].

Opportunities for longitudinal follow up studies using serum markers exist. In the co-infection cohort serum were taken at regular intervals during a RCT of Hepatitis C treatment. Evaluation of treatment outcome and changes in biomarkers including ELF would be a valuable addition to the evidence on the use of serum markers. In addition a cohort of 267 patients were recruited by the original ELF study, who had quarterly serum markers measured for 2 years and then had a repeat biopsy at the end of this time. Drop out rates were high (only 84 paired liver biopsies were obtained). However, an analysis using serial ELF values to evaluate the performance of changes in ELF score over time in the prediction of clinical outcomes would be of interest. Such longitudinal follow up studies could be extended to evaluate the role of ELF in prediction of longer term clinical outcomes rather than detection of fibrosis severity on biopsy -which is itself a surrogate outcome for clinical outcomes. Two such studies are presented in Chapter 5.

4.5 Conclusion

The ELF panel can be simplified to HA TIMP1 and P3NP without loss of diagnostic performance. It has been shown to have a good diagnostic profile in 8 external validation studies in CHC NAFLD PBC and HCV-co-infection. It appears to perform as well as most published serum markers of liver disease although there are no direct comparison studies. In CHC patients modeling suggests that ELF may avoid biopsy correctly in 67% of patients with

9% false test rates and with 77% of the population successfully being given a test result in a setting where the prevalence of serious fibrosis is ~ 30%). Future research should include extension of this validation work to ALD cohorts, and to the development of the ELF panel by addition of simple markers. Research attention should be directed at the establishment of a large reference population for direct comparison of available non-invasive markers.



CHAPTER 5

PROGNOSTIC PERFORMANCE OF THE EUROPEAN LIVER FIBROSIS (ELF) PANEL MARKER

5.1 Introduction

Liver fibrosis/cirrhosis is asymptomatic in most people until the disease has progressed to the end stages of disease, when hepatocellular failure and /or portal hypertension develop. This may take decades. Consequently ascertainment of liver disease before clinical symptoms occur is difficult, and prediction of future morbidity and mortality problematic. The more severe the fibrosis, the more likely the probability of occurrence of a clinical outcome. Liver fibrosis has been used as a surrogate for clinical outcomes in CLD. The method by which fibrosis is assessed is by examination of histological specimens of the liver obtained by biopsy. This provides a flawed reference standard as has been described in previous chapters. Interest has grown in the last decade in the use of non-invasive methods to predict clinical outcome in patients with CLD. If such biomarkers could predict who is more likely to suffer clinical outcomes then this could provide valuable information to optimise the management of patients with CLD.

In this chapter the prognostic accuracy of ELF panel to predict adverse clinical outcomes is evaluated in two studies:

(i) a follow-up study of the patients with liver disease recruited to the original ELF cohort in 1998-2000. The outcome measures used comprised liver related mortality and decompensated cirrhosis, hepatocellular cancer and all-cause mortality.

(ii) a study which evaluated the performance of ELF in the prediction of clinical outcomes in a cohort of patients with PBC followed up for up to 10 years.

First the principles of prognostic models and methodological issues of analysis are discussed.

5.2 Principles of Prognosis

Prognosis is the prediction of the probable course and outcome of a disease. Variables which are related to the course or outcome may be combined into a prognostic model to improve prediction.

Assessment of the prognosis is needed for rational decision making about treatment and management by patient and doctor in addition to providing information to the patient and

their family regarding prospects for the future. Prognostic models commonly use regression analyses of baseline characteristics of patients and/or diagnostic tests in defined populations with well defined outcomes. The ideal prognostic model would be widely generalisable using accepted diagnostic criteria. Derivation should be in representative, well described populations, consecutively recruited with adequate sample size allowing for the number of events and the number of possible variables being considered in the model. Follow up should be as complete as possible. Variables should be cheap to measure by easily available means, with a high degree of precision, accuracy, reliability and reproducibility (see glossary for definitions). Prognostic models are best derived from studies that fulfill criteria for high quality[180] . These are:

- use of an inception cohort (patients assembled at a similar stage of disease),
- description of referral patterns (ways in which patients were recruited into the study),
- complete follow up achieved,
- use of objective outcome criteria,
- outcome assessment blinded, to choice of model variables,
- adjustment for extraneous prognostic factors carried out.

In clinical medicine a true inception cohort is difficult to assemble, especially in chronic disease. Patients tend to present to healthcare services in a haphazard manner at varying stages of illness and comparability of staging is only really possible for acute illness such as myocardial infarction. In chronic diseases it is possible to gather an inception cohort all of whom are at the same stage of an incident event such as treatment.

Liver fibrosis due to CLD has an asymptomatic phase making time of onset/stage of disease difficult to ascertain and assembly of an inception cohort very difficult. Presentation may be early or late leading to heterogeneity, which may impact on precision of any model as prognosis in later stages may differ from earlier ones. Many factors may need to be considered in the prediction of outcome in CLD- for example sociodemographic (age, gender, ethnicity socio-economic status), clinical (cause of CLD, biopsy stage at presentation), disease activity (for example inflammation, viral load in CHC or CHB), presence of other factors that would affect outcome (for example alcohol, obesity), and the presence of decompensated liver disease. Adjustment for the effect of any effective therapeutic interventions during the course of CLD is also important. All of these factors

may interact in a complex way. Outcomes used in published models in CLD have been death or the first signs of portal hypertension or liver decompensation or HCC or transplantation.

In most studies Cox's regression model for censored survival has been the model of choice. This assumes proportional hazards throughout follow-up time and log-linear additive effects of the variables. Logistic regression modeling has also been used when the risk of the event within a defined time interval is of interest. Accelerated failure time models and neural network systems (requiring large numbers of patients and estimated variables) have been used. The latter is a "black box" analysis and does not provide information on how the variables contribute to the prediction. The number of variables fitted in the model should not exceed 10% of the number of patients reaching end-points, i.e. >=10 events per variable. All follow up information can be used in Cox regression model for time dependent variables. Use of time dependent analysis provides stronger prognostic information than corresponding time fixed models, is more useful to update prognosis during follow-up and allows effect of treatment to be taken into account[181].

5.3 Prognostic models in CLD

Prognostic models in liver disease have most often been used to predict outcome in cirrhosis, the most widely used ones at present are Child-Turcotte Pugh classification (CPC)[182], the Model for End-stage Liver Disease (MELD)[183], and the Mayo risk score (developed for PBC)[184] (See Appendix 1). In a systematic review of prognostic indicators of survival in cirrhosis[53] 118 prognostic studies of cirrhosis were included (of which 31 were "good quality" studies but with only one study meeting all quality criteria, and only 17 having validated results). The major methodology problems were reported to be the inclusion of non-consecutive patients, incomplete reporting of inclusion criteria, incomplete follow up, and inclusion of patients at different stage of disease. Statistical methodology limitations were that in the majority of studies the ratio of deaths to number of variables was <10. Such "over-fitting" of the model may lead to false positive results. Variability of survival times across studies may be explained by heterogeneity of included patients who were at different stages of cirrhosis. Notwithstanding these limitations this review concluded that the most robust predictor of death in cirrhosis is the Child Pugh score (see Appendix 1 for formula)[53].

Prognostic variables differ depending of the stage of cirrhosis e.g. compensated versus decompensated. As decompensated cirrhosis becomes worse, parameters assessing circulatory deterioration become important and other clinical scores which include these (such as MELD) have an important place at this more severe end of disease. There are time dependent models for PBC and primary sclerosing cholangitis (PSC).

The variables used in these published prognostic models generally have selected parameters that are not central to the disease processes –e.g. ascites, and bilirubin which are indirect indicators of hepatic function. There are very few studies of prognosis in pre-cirrhotic patients. The use of components of extracellular matrix formation and breakdown may be more direct measures of the fibrotic process and therefore may provide more accurate information on prognosis across all stages of disease.
The ELF serum markers (which are markers of extracellular matrix formation and breakdown) are thought to provide more direct information on disease processes, and by using clinical outcomes rather than results of a liver biopsy, a more accurate prediction of prognosis may be provided.

5.4 Existing literature of serum markers to predict clinical outcome

A literature review was conducted to evaluate the evidence on serum markers predicting clinical outcomes. A simple search strategy was used in Medline database searching years between 1995-2006. The last decade was chosen as previous reviews in chapter 3 suggest that higher quality larger studies are more likely to be found in this timeframe. All reference lists of retrieved articles were searched so earlier studies could be identified and assessed. 3 studies were found[185-187] .

Critical Appraisal of studies

Korner (1996)[185] investigated the prognostic value of laminin and HA in 38 patients with fibrosis (n=4) and established cirrhosis (34) followed up for one year. 16 events were reported during this time and of these 6 patients died. Laminin (≥2.6u/ml) had a sensitivity of 71% and a specificity of 86% in the prediction of liver related outcomes, HA (≥200ng/ml) had a sensitivity 90% and specificity 67%; and both had a Relative Risk of 2.7 for the development of clinical outcomes. This was a small study, the selection of included patients was not described and AUC values were not reported.

Guechot (2000)[186] conducted a study to evaluate the prognostic value of HA in 91 patients from one centre with CHC who had biopsy proven cirrhosis with no decompensation

events. The outcomes used were decompensated cirrhosis and liver related death. Median follow up time was 38 months with 24 events, 14 of which were fatal. Hazard ratios (HR) were derived from Cox proportional hazards analysis. Multivariate analysis showed 4 biomarkers were independently predictive of liver related events – (i) HA >350mcg (HR 3.5(1.4-8.4)); (ii) bilirubin >18 mmol/l (HR 2.6 (1.03-6.6)); (iii) albumin <36g/l (HR 2.6 (1.04, 6.3)) ;(iv) PT <63% (HR 2.5(1.05, 6.1)). **Ngo (2000)**[187] This was a study from a single centre evaluating the prognostic performance of Fibrotest in 537 patients with CHC recruited over 5 years. 243 were followed up for 5 years. Outcomes were liver related mortality/decompensated liver disease, and all-cause mortality. There were 49 events (9 liver related deaths, 11 non liver-related deaths, 20 non fatal liver events). Loss to follow up was 48% at 5 years even for ACM. The AUC at 5 years for Fibrotest predicting liver related outcomes was 0.96 (0.93, 0.97), and the AUC at 5 years for Fibrotest predicting all-cause mortality (ACM) was 0.76 (0.63.0.84). Results for biopsy in prediction of liver-related outcomes was AUC of 0.91 (0.85 0.94) and all-cause mortality 0.66 (0.52, 0.78). Crude Kaplan Meier survival analysis showed a difference between severe fibrosis and all other fibrosis. Using Cox Proportional Hazard model adjusted for histology, anti-viral treatment, alcohol consumption and HIV infection, the hazard ratio for Fibrotest predicting survival with no liver related outcome was 7.31(2.71,25.44) and for biopsy the hazard ratio was 1.84 (1.16,2.91). In the prediction of all cause mortality the adjusted HR for Fibrotest was 2.27 (1.40, 3.66) and for biopsy 1.10 (0.69 1.74). There was no flow chart to show the follow up status of the patients and how successful the strategies outlined in the methods section had been in determining outcomes in those lost to follow up.
No novel predictive models were reported in these studies.

Conclusion of literature search

There is scant literature on the performance of serum markers in predicting clinical outcomes, and only one in the last 5 years. There were 2 small studies in patients with cirrhosis, largely conducted in patients with CHC. Nevertheless Fibrotest did show promising performance in the prediction of clinical outcomes though missing data is a problem in this study. The studies on ELF will add to the body of evidence using (i) a large cohort of patients with mixed aetiology CLD (Study 1) (ii) a cohort of patients with PBC (Study 2).

5.5 The performance of ELF serum marker panel in predicting clinical outcomes in a cohort of patients with mixed aetiology CLD (STUDY 1)

5.5.1 Aim of the study

1. To evaluate the prognostic performance of the ELF panel and to compare it to biopsy in predicting:

   (i) Liver-related outcomes (Primary outcome)

   (ii) All-cause mortality (Secondary outcome)

2. To determine if there are alternative biomarkers or combinations of biomarkers that may predict clinical outcomes in a mixed aetiology CLD cohort.

5.5.2 Methods

As described in Chapter 4, the simplified panel of ELF is used in the prognostic study. This is particular important in that it permits age to be adjusted for in statistical analyses.

Data collection

13 centres participated in the original ELF study recruiting 1,126 patients, of which 922 were included in the cross sectional diagnostic study (see chapter 4). The original principal investigators for each centre were contacted to get their agreement to the follow-up study, and an individual staff member was identified in each centre to help with medical note retrieval and administration of the study. Ethical permission to follow up the patients for mortality and morbidity via medical case note analysis was granted by MREC South West (MREC/98/6/08). In each centre local procedures for R&D were established and followed. There was no centrally held patient list. Anonymised lists were held by each centre and following a protocol resulting from a pilot study in Southampton, the archive of original ELF study was located and the patient list re-assembled and medical notes of participants retrieved. Data collection forms were constructed (see Appendix 7) and modified following the Southampton pilot. Personal face-to-face contact was made with each centre principal investigator to discuss the study protocol and formulate a study process acceptable to each centre. Key personnel were identified with the principal investigator, and visits to each centre in England and two in continental Europe were arranged by direct liaison with this key person. For the larger centres data extraction was performed with support from a senior research nurse with hepatology training (CG) and a medical student doing a BSc in the department (MW). Standardisation of data collection was achieved using the same data

collection forms which incorporated definitions of outcomes. For three centres in Continental Europe a named clinical person was delegated to the study and extracted data using study data collection forms. Close liaison was maintained with these persons via telephone and email communications.

Clinical data were collected on the date of first clinical outcome/transplant/ and (if occurred) death. Clinical outcome were defined as death (liver related and non-liver related death), any episodes of decompensated cirrhosis post-recruitment (ascites-defined by paracentesis, ultrasound, or clinically), encephalopathy (defined clinically) oesophageal variceal haemorrhage (defined clinically and with endoscopy), acute alcoholic hepatitis on background of biopsy proven cirrhosis (defined clinically), liver transplantation, and hepatocellular carcinoma (HCC)). There was standardisation in terms of collection of data on required test procedures, with the case definitions being those used in standard clinical practice.

Any further biopsy evidence of cirrhosis since the original ELF study ended was noted. Date last known alive and loss to follow up from hepatology centre were collected. Dates and details of any treatment and the outcome of that treatment that may have affected prognosis (e.g. response to treatment with anti-viral therapy for Hepatitis B and C, patients with auto immune hepatitis on steroids /immunosuppression, loss of weight for NAFLD, abstinence for ALD) were collected. In England data were collected independently by three observers in the larger centres. Data collection and analysis were conducted by individuals blinded to test results.

Follow up in England and Continental Europe

It was important to minimise the lost to follow up in the cohort and to obtain as complete as possible ascertainment of mortality and liver related outcomes, in order to minimise the potential for selection bias (differences in associations of those lost to follow up vs. those studied), and to increase the power with information from as many patients as possible to be included in the analysis. Methods used to do this differed in England and Continental Europe.

(i) England

Mortality in the patients in England was ascertained using the national systems of patient tracing and follow up (National Strategic Tracing Service). All death certificates were obtained from the Family Records Office or local Register Offices. For those patients who were lost to follow up or discharged from care from the original recruiting centre in

England, MREC approval was gained to contact the last recorded General Practitioner in order to ascertain when the patient was last known to be alive and to determine if there had been any clinical hepatic decompensation. A short questionnaire was constructed to collect these data (Appendix 8). In order to maximize response to the questionnaire a second letter requesting information was sent out to Primary Care to those GPs who had not responded to the first letter.

(ii) Continental Europe

In 3 out of 4 centres in Continental Europe it was not possible to ascertain mortality of those patients who had been lost to follow up as there were no processes which allowed national flagging of deaths. Such a process is possible in Sweden but these data are pending and have not been included in this thesis. Mortality and cause of death of those who remained under the care of the recruiting centre, and the date last known alive were ascertained. Clinical decompensation events were ascertained in those patients who remained under the care of the recruiting centre. One centre contacted the General Practitioner recorded in the clinical records of those discharged to ascertain any clinical events. The decompensation events in those lost to follow up in centres in Continental Europe could not be ascertained.

Statistical methods

Medians with inter-quartile ranges (IQR) were used to describe continuous variables, and mean values are presented with standard deviation. Comparison between medians was made using Mann Whitney or Kruswal Wallis statistics. Survival analysis to derive cumulative probability of survival was conducted using Kaplan Meier curves with ELF score divided into tertiles, and biopsy into three groups (nil/mild; moderate; and severe/cirrhosis). The ELF score was divided into tertiles. The method of deriving tertiles can be (i) by having equal numbers of patients per tertile or (ii) having equal ranges of score per tertile. This latter method was chosen as the former is dataset specific and it is likely that in the wider clinical context the latter method would be more generalisable and appropriate. The number of events and persons in each tertile were assessed. The greatest number of people and greatest number of events were found in the middle tertile. This middle tertile was then divided into two groups by score, to evaluate whether this further division of ELF score could result in more precise prediction by score.

Adjusted hazard ratios (HR) of the association of ELF with risks of liver-related outcomes were derived for ELF and biopsy using Cox proportional hazards model. Variables chosen for adjustment in this analysis were chosen a priori from knowledge of factors associated with possible liver fibrosis development and progression. They included age, sex, alcohol consumption, treatment response, and smoking. Other terms were added from results of data analysis. The proportional hazards assumption was checked using log minus log graphics (SPSS version 14, SPSS Inc Chicago Illinois) and Schoenfeld residuals in STATA (version 9). Linearity of ELF was assessed by testing if a quadratic expression was required, and by fitting ELF as categorical and continuous predictors with the categorical having no significant improvement over the simple linear term. Logistic regression at a time-specific point was used to provide adjusted odds ratio of unit increase in ELF predicting liver-related outcomes and all-cause mortality adjusting for factors associated with liver outcomes such as age, gender, aetiology of CLD, smoking and alcohol consumption. Those persons who did not reach the stated time point and who did not have an event were excluded from the analysis.

The development of a model that may best predict liver related outcomes using available variables of serum markers and simple blood tests was conducted using methods described by Collett [188]. This is a model building method that uses all available information and requires three stages. The first step was a **sequential** approach with each of the nine serum markers and eight simple markers (having sufficient values to make numbers in the analysis meaningful) being added **on their own** into a logistic regression predicting clinical outcome. Those with a **significance level <0.1** were noted. Subsequent analyses included those variables whose p value was <0.1. The second step was to derive a **parsimonious set of markers** having entered all of the markers which reached <0.1 signfdicance on the first step, using the **backward elimination** (retention threshold p=<0.05). The final step is to **add in separately** to the final model those markers that did not reach p=<0.1 at the first step, to evaluate their impact on the model. ROC curves were plotted for ELF, biopsy and the new model. As standard methods do not exist for deriving ROC curves for time to event data, occurrence of event as compared to non-occurrence of events within 6 years was used as the outcome for these analyses. The Hosmer Lemeshow statistic was used to show the goodness of fit of models to the data. ROC curves were compared using the Hanley MacNeil method (Stata 9 package).

The reliability of ELF score was evaluated by calculating an intra-class correlation (ICC). This is a measure of the correlation, consistency or conformity for a data set when it has multiple groups. It is the ratio of between-groups variance to total variance. Serum marker scores were available in a subset of 84 patients from the original cohort who had agreed to have repeat serum marker measures over 2 years, and who had baseline and 3 month values. ICC values were derived for each of the serum markers in ELF, and for the ELF score itself. A value of 1 being indicative of complete reliability which is when there is no measurement error and the proportion of the total variability is explained by between-subjects and not within-subject variability. If the serum markers variability produced an ICC of 0 then most of the variability would be explained by differences within subject. ICC was estimated using a two way random effects model.

5.5.3 Results

In order to compare centres in Continental Europe and centres in England, data are presented separately.

Baseline Data characteristics

998 of the 1,126 patients had data on all nine serum markers and 904 patients from 11 centres had available follow up data and were included in this prognostic study. Recruited patients excluded from the published diagnostic study largely had uninterpretable or missing biopsy, or missing serum marker values. Data were obtained in 7 centres in England and 4 centres in Europe. Baseline characteristics of the cohort are presented in Table 5.1. The characteristics are similar between English and European centres. Most of the patients in all centres had CHC but there was little ALD in the patients recruited in the continental European centres. More patients in England had self-reported moderate / heavy alcohol consumption (21% vs. 7%) and more were current smokers (49% vs. 37%). The majority of the patients in all centres were white (96%). The proportion of patients with serious fibrosis/cirrhosis on biopsy was similar in England and Europe (26% and 27%).

In order to check how close the cohort is to an inception cohort, information on how long the participants had been under hepatology care before recruitment was collected in one centre (n=102). The mean time was 0.3 years implying that in this centre most of the patients could be regarded as being part of an inception cohort.

Table 5.1 Baseline characteristics of cohort

	England (498)	**Europe (500)**	**Complete cohort (998)**
Median age (IQR)	43 (17)	44 (21)	43 (19)
% male	66	61	63
Main Aetiology*			
HCV	210 (42)	236 (47)	443 (44)
ALD	88 (18)	6 (1)	94 (9)
NAFLD	44 (9)	26 (6)	73 (7)
PBC	31 (6)	15 (11)	46 (5)
AIH	28 (6)	30 (6)	60 (7)
PSC	12 (2)	10 (2)	22 (2)
HBV	30 (6)	27 (5)	57 (6)
HHC	14 (3)	12 (2)	26 (3)
Other	42 (8)		74 (7)
Alcohol (self reported)			
Abstinent	168 (34)	277 (56)	445 (45)
Light	223 (45)	187 (38)	410 (41)
Moderate	60 (12)	30 (6)	90 (9)
Heavy	46 (9)	3 (1)	49 (5)
Missing			
Smoking status n			
Never	178 (36)	208 (42)	386 (39)
Past	75 (15)	84 (17)	159 (16)
Present	245 (49)	183 (37)	427 (43)
Missing			
Ethnicity			
White	475 (95)	488 (97)	962 (96)
Asian	11 (2)	4 (1)	15 (2)
Black	5 (1)	4 (1)	9 (1)
Other	8 (2)	4 (1)	12 (1)
Liver fibrosis (Ishak score)	457 (92)	471 (94)	928 (93)
0	119 (26)	110 (23)	229 (25)
1	108 (24)	89 (19)	197 (21)
2	59 (13)	77 (16)	136 (15)
3	52 (11)	65 (14)	117 (13)
4	45 (10)	39 (8)	84 (9)
5	19 (4)	39 (8)	58 (6)
6	55 (12)	52 (11)	107 (12)

Transplant pre study 41 (8%) -allocated aetiology of pre transplant pathology; 4% remained unknown aetiology
*derived from information collected from clinical records where possible.

As the outcome measures were clinical outcomes and not histology, additional patients were included in the prognostic study above those who had been originally recruited to diagnostic study. These patients had been excluded due to absence/inadequacy of liver biopsy. Table 5.2 shows the baseline characteristics of these additional patients. They were similar except that the additional patients have a higher proportion of ALD than in the diagnostic study.

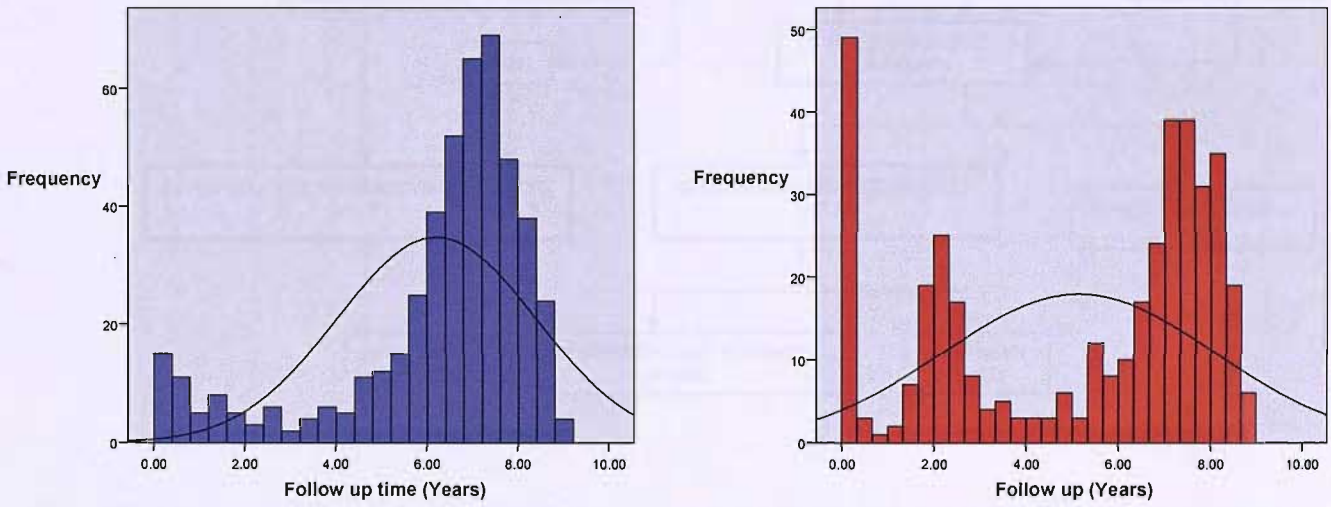
Table 5.2 Baseline characteristics of patients in prognostic but not in original diagnostic study

Characteristic	ELF diagnostic n=922	Prognostic n=998	Patients in prognostic study not in diagnostic study n=76
Age yr (range)	43 (19-75)	43 (19-75)	47 (21-70)
% male	63	63	67
HCV (%)	47	50	36
ALD (%)	6	10	20
NAFLD (%)	10	8	5
PBC (%)	6	5	7
Other (%)	31	27	32
Mean Baseline ELF (SD)	-1.29 (1.54)	-1.26 (1.54)	-0.89 (1.64)

Follow-up of patients

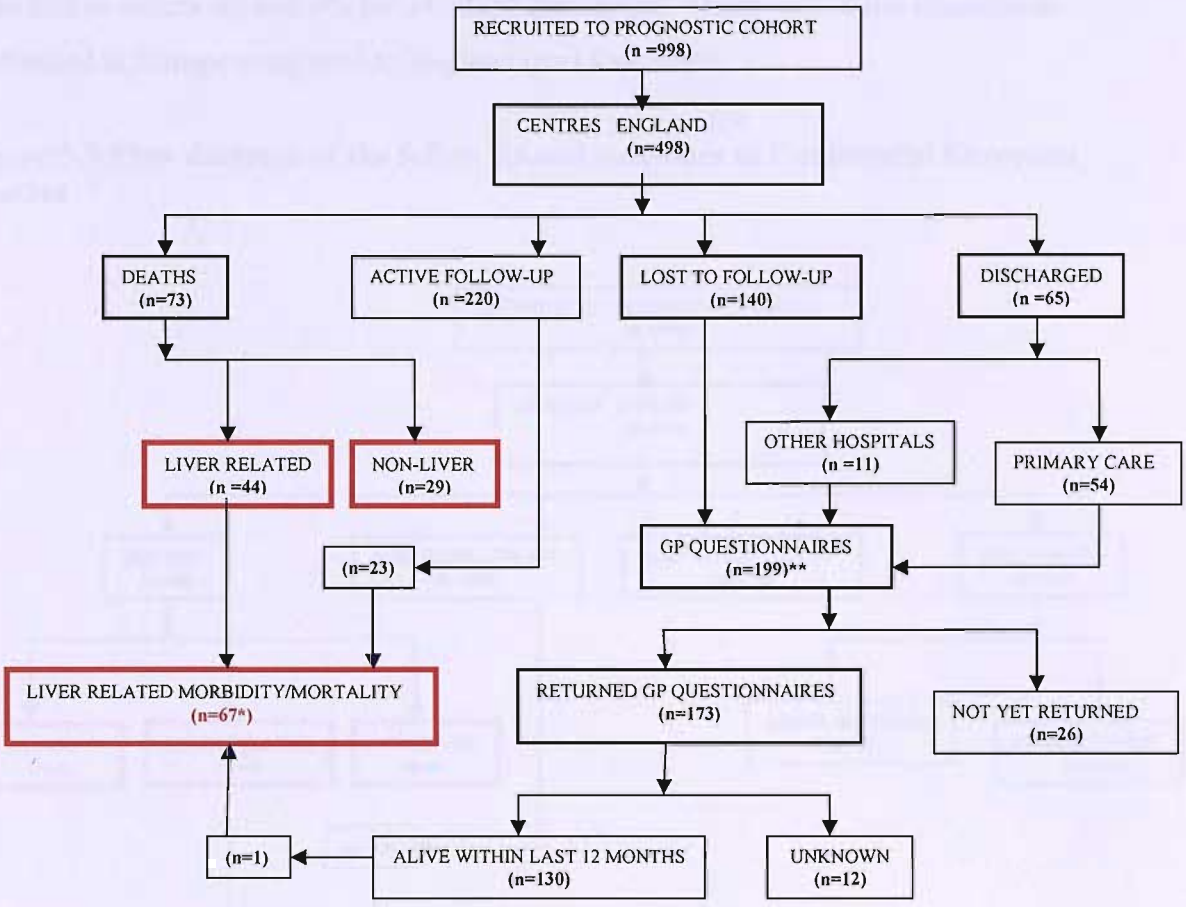
Figures 5.1 to 5.3 show the follow up status of recruited patients of those who had not had an event in England and Europe centres. The median follow-up time (time between last known alive and recruitment in those patients with no events) in England was 6.87 years (IQR 1.85) and Europe 6.6 (5.4) with distribution of loss to follow up in Europe being skewed with large loss to follow from the recruiting centre at 4 years.

Figure 5.1 Follow up times in England and Europe



The follow up times (difference in years between last known alive and date recruited to study in those patients who did not have clinical events) in England centres and European centres. The follow up time was longer in England (mean 6.2 years (SD 2.17) vs. 5.1 years (SD 3.0)) with many of those lost to follow up in Europe centres occurring soon after recruitment)

Figure 5.2 Flow diagram of the follow up and outcomes in England Centres



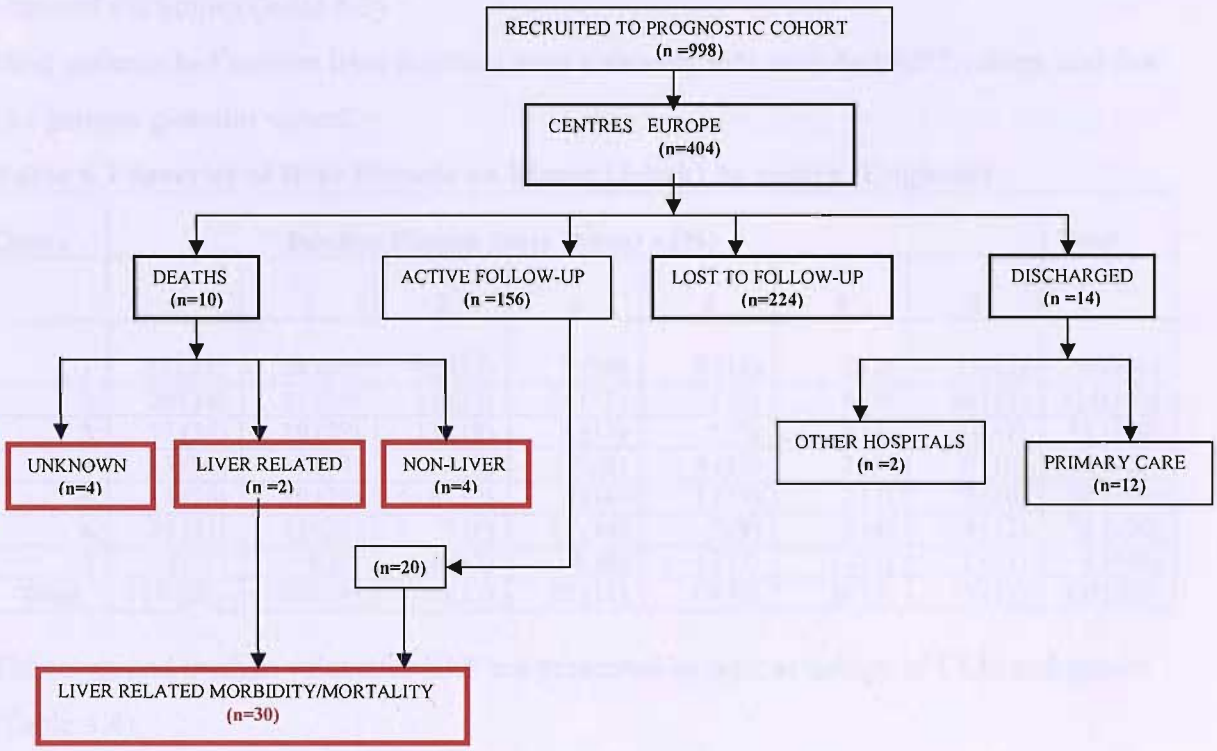
Flow diagram showing the follow up and outcomes for England centres (n=498) *5 people had a liver related decompensation before death by non-liver cause **emigration (3), no/erroneous GP details available (4) or discharged with LKA within last 12 months so no GP questionnaire sent)

The lost to follow up rate from original recruiting centre was 42% (29% did not attend appointments and were therefore lost to recruiting centre, and 13% being discharged from the recruiting centre). Of the questionnaires sent to GPs for 199 patients to determine outcomes in those lost to follow-up or discharged by the recruiting centres the response rate was 86% (n=173), with only one reporting a clinical decompensation event. Ascertainment of clinical status and outcome was therefore possible in 92% of patients and all outcome data are presented as a proportion of the inception cohort.

In the centres in England, the overall mortality rate in those patient recruited to this study was 15% (n=73), of which 44 (60%) were liver related, defined as any mention of liver disease on Part 1 on the death certificate. There were 67 liver–related outcomes (liver-related deaths or hepatic decompensation). In the centres in Continental Europe the overall

mortality rate was 2.5% (n=10) with 30 liver-related outcomes. 56% (n=224) of patients were lost to follow up and 4% (n=14) were discharged. There were more transplants performed in Europe compared to England (n=11 vs. n=6).

Figure 5.3 Flow diagram of the follow up and outcomes in Continental European Centres



Flow diagram showing the follow up and outcomes for European centres (n=404). Non attendance at recruiting centres was high (56%)

There was a large loss to follow up in the European centres (60%). This compromised the validity of survival analyses as there was no information on clinical outcomes and mortality in these patients. Whilst the prevalence of severe fibrosis (Ishak stages 4-6) was similar in Europe and England, the mortality rate was different (between the English and European centres (15% vs. 2.5%) and there were 30 (7%) liver related outcomes compared to 67 (13%) in England centres. It is therefore likely that there are many deaths/liver related outcomes that have not been ascertained. Consequently further analyses are presented using only the England centres (n=498), where data were robust and more complete. The data collected from continental Europe, and a sensitivity analysis in which the whole cohort (including the European centres n=902) are reported in Appendix 6.

Baseline investigations in 7 centres in England

All 498 patients had all values for nine serum markers involved in extracellular matrix formation and breakdown. 457/498 patients had a biopsy that could be interpreted (92%). The proportion of serious fibrosis/cirrhosis in each centre (Ishak 4-6) ranged from 20-34% (median 26%). The majority of patients in each centre had nil/mild fibrosis on histology obtained via biopsy (table 5.3).

Most patients had routine liver function tests although 53% only had AST values, and few had gamma globulin values.

Table 5.3 Severity of liver fibrosis on biopsy (Ishak) by centre (England)

Centre	Baseline Fibrosis Score (Ishak) n (%)							Total
	0	1	2	3	4	5	6	
1	22 (23)	24 (26)	16 (17)	9 (10)	9 (10)	2 (2)	11(12)	93(100)
2	29 (24)	31 (26)	14 (12)	14 (12)	11 (9)	6 (5)	16 (13)	121(100)
3	17 (24)	19 (27)	13 (18)	8 (11)	5 (7)	3 (4)	6 (9)	71 (100)
4	13 (34)	7 (18)	2 (5)	3 (8)	5 (13)	2 (5)	6 (16)	38 (100)
5	13 (26)	10 (20)	6 (12)	7 (14)	7 (14)	2 (4)	5 (10)	50 (100)
6	24 (31)	17 (22)	7 (9)	11 (14)	7 (9)	3 (4)	9 (12)	78 (100)
7	1(17)	0 (0)	1 (17)	0 (0)	1 (17)	1 (17)	2 (33)	6 (100)
Total	119 (26)	108(24)	59(13)	52 (11)	45(10)	19 (4)	55 (12)	457(100)

The mean and median values for ELF are presented by age, aetiology of CLD and gender (Table 5.4).

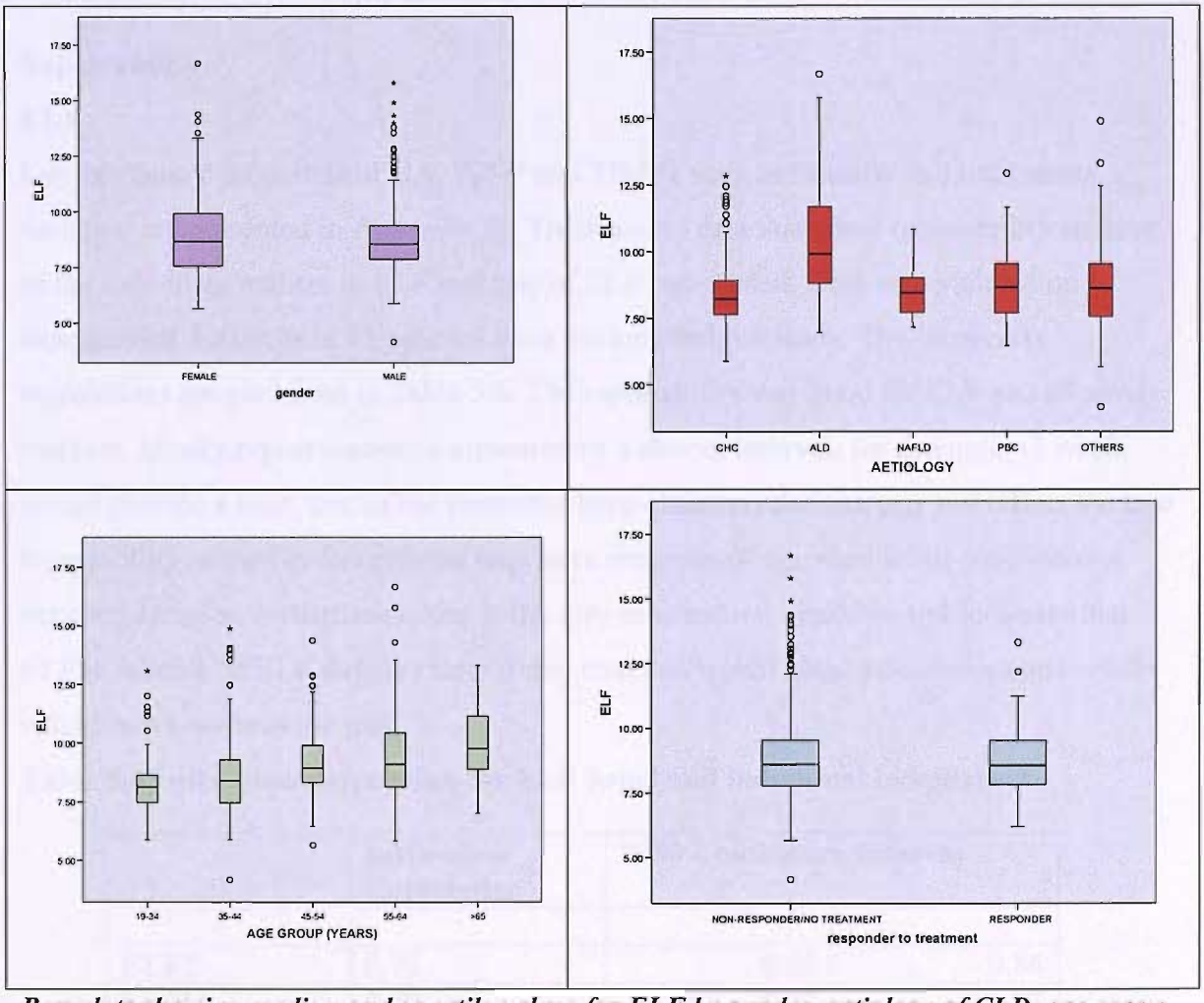
Table 5.4 Available routine blood tests England (total n=498)

	Number tests	Mean (SD)	Median (IQR)
HA	498	150 (25.6)	38 (19, 83)
P3NP	498	7.43 (0.43)	4.8 (4, 8)
TIMP1	498	772 (20.3)	663 (520, 848)
Laminin	498	24 (2.5)	15 (9, 23)
MMP2	498	652 (12.7)	597 (466, 750)
MMP9_timp	498	545 (12.3)	491 (360, 658)
Tenascin	498	458 (14.3)	384 (258, 567)
Coll IV	498	184 (7.1)	141 (112, 192)
Coll VI	498	5.1 (0.12)	4.5 (3.5, 6.0)
INR	489	1.0 (0.23)	1.0 (0.9, 1.1)
Creat µm/l	483	89 (18.9)	87 (77, 98)
ALT iµ/l	460	80 (123.3)	47 (29, 97)
GGT iµ/l	417	156 (276.8)	63 (31, 133)
AST iµ/l	263	66 (98.6)	45 (30, 71)
γ Globulin g/dl	38	32 (6.4)	32 (29, 36)
Albumin g/l	489	43 (6.6)	44 (40, 46)
Bilirubin µm/l	487	30 (71.9)	10 (7, 19)
Alkaline Phosphatase iµ/l	488	141 (126.6)	99 (75,178)
Platelets 10^9/l	488	229 (69.7)	242 (175, 274)

There was variation in availability of the simple blood tests (e.g. ALT 92%; AST -53%; γ Globulin 8%). The distributions of the serum markers were positively skewed. The routine liver function tests (AST, ALT, GGT) all have median values that were above the upper limit of normal, whereas the median values of platelets, bilirubin, alkaline phosphatase and albumin were within normal ranges.

There was a significant difference between the median value of ELF in patients with ALD as compared to all other disease categories; (p=<0.0001) (Table 5.5; Figure 5.4). No other differences between other disease categories reached statistical significance. There was no significant difference between medians of ELF between genders. There was a significant difference between the age categories with means of ELF increasing with age category (Jonckheere Terpstra test).

Figure 5.4 Box plots of ELF by age, aetiology of CLD, sex and response to treatment (England)



Box plots showing median and quartile values for ELF by gender, aetiology of CLD, age group categories and response to treatment. *Responders to treatment defined as those with SVR in CHC, sero-conversion in CHB and use of steroid/azathiaprine >3 months in AIH compared to non responders & not treated in the rest of population.

Table 5.5 Median values of ELF for gender, CLD aetiology, age group, responder to treatment for centres in England

		Median	Inter quartile range(range)
Gender	**Male**	8.56	1.52 (4.16-15.78)
	Female	8.68	2.37 (5.66-16.67)
Aetiology CLD	**CHC**	8.24	1.26 (5.88-12.83)
	ALD	9.91	2.78 (6.95-16.67)
	NAFLD	8.48	1.34 (7.16-10.84)
	PBC,PSC,AIH	8.81	2.49 (5.84-13.31)
	others	8.48	1.73 (4.16-14.89)
Age group (years)	**19-34**	8.14	1.24 (5.88-13.06)
	35-44	8.41	1.84 (4.16-14.89)
	45-54	8.93	1.60 (5.66-14.37)
	55-64	9.12	2.38 (5.84-16.67)
	>65	9.77	2.82 (7.01-13.31)
Responder to treatment	**No/no treatment**	8.61	1.79 (4.16-16.67)
	responder	8.56	1.91 (6.20-13.31)

Repeatability

ELF

Key laboratory properties of HA, P3NP and TIMP1 such as linearity and inter-assay variation are presented in Appendix 6. There are no data that allow repeatability analysis of the individual makers in ELF and also of ELF panel itself. Data are available on baseline and 3 months in 75 patients from the longitudinal study. The intra-class correlations are presented in Table 5.6. The repeatability was good for ELF and all single markers. Ideally repeat measures separated by a shorter interval, for example <1 week, would provide a truer test, as the presented intra-class correlations may not reflect the true repeatability as the fibrotic process may have progressed/regressed in the time interval between samples. Nevertheless this is the only information available and indicates that ELF is reliable. If ELF did vary day-to-day then this would blunt associations and results would move towards the null.

Table 5.6 Intra-class correlation for ELF panel and individual biomarkers

	Intra-class Correlation	95% Confidence Interval	
ELF2	0.78	0.65	0.86
Ln HA	0.79	0.67	0.87
Ln TIMP1	0.76	0.63	0.85
Ln P3NP	0.67	0.49	0.79

Histology

Repeatability of histology was evaluated in the original diagnostic study. Three pathologists blinded to results staged the histology on 620 of the biopsies from the cohort. Inter-observer reliability between pathologists varied (kappa scores of 0.97 and 0.5). Intra-observer reliability for one pathologist was very good (kappa = 0.93 95% CI 0.91, 0.96).

Clinical outcomes for England

The all-cause mortality did not vary substantially by centre (Table 5.7).

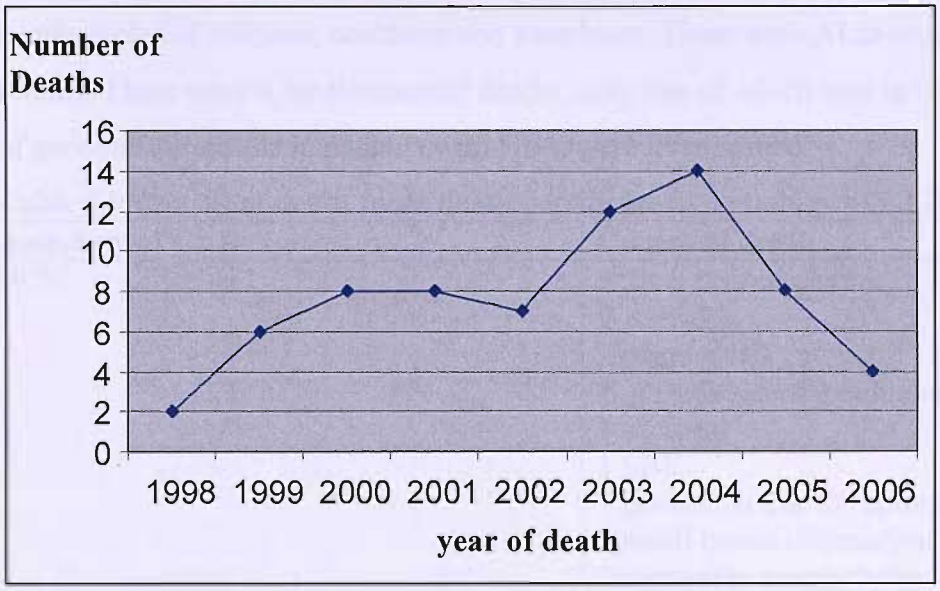
Centre 7 joined the study later than other centres and recruited many fewer patients which may account for the lower crude mortality. The median age of death was 47 yrs (30-73) and 73% were in males (n=50). The median age of liver-related outcomes was 47 yrs (28-72) with 66% in males (n=44).

Table 5.7 Liver related outcomes and All cause mortality by centre (England) n (%)

Centre	Number recruited (n)	Liver related outcomes (n) (%)	All cause mortality n (%)
1	102	13 (19)	13 (13)
2	125	22 (33)	16 (13)
3	79	9 (13)	13 (17)
4	45	6 (9)	6 (13)
5	52	5 (7)	8 (15)
6	89	11 (16)	16 (18)
7	6	1 (2)	1 (17)
Total	**498**	**67 (100)**	**73 (100)**

Deaths occurred throughout the period of follow up with a peak in 2004 (figure 5.5).

Figure 5.5 Number of deaths by calendar year (England)



Number of deaths in patients recruited in centres in England by year of death, showing that deaths occurred throughout the period of follow-up.

Table 5.8 shows the clinical outcomes by liver disease aetiology. Almost half of the deaths were in patients who had ALD (48%) and 27% were in patients with CHC. Similarly most (54%) of the liver-related outcomes were in those with ALD. 61% of liver related deaths were in patients with ALD. Prognosis varies by aetiology of CLD, with those patients with ALD having higher than expected incidence of clinical events (observed/expected for ALD for liver related outcomes = 5.4 compared to CHC observed /expected = 0.4 and NAFLD observed /expected = 0.2 See Appendix 6 for detailed analysis Tables 1a-2). Aetiology of CLD should therefore be taken into account in appraising the prognostic performance of markers.

Table 5.8 Deaths and liver related clinical outcomes by baseline liver disease aetiology (ENGLAND) n (row % within mortality group)

	HCV	ALD	NAFLD	HBV	PBC	PSC	AIH	Other	Total
Liver related outcomes	13 (19)	36 (54)	1 (2)	4 (6)	3 (5)	3(5)	4(6)	3 (5)	67 (100)
ACM	20 (27)	35 (48)	2 (3)	3 (4)	3 (4)	5 (7)	3 (4)	2 (3)	73 (100)
Liver related mortality	9 (21)	27 (61)	0 (0)	3 (7)	1(5)	2 (5)	1 (2)	1 (2)	44 (100)

37 % of HCV deaths were liver related; 77% ALD death were liver related; none of NAFLD 40% PBC all of HBV The cause of death (derived from the death certificate) by CLD aetiology is reported in Table 5.9. In those patients with CHC the non-liver related deaths included suicides, accidents and overdoses. Those with ALD were mostly cancer related. There were 4 cardiovascular deaths, only one of which was in the NAFLD group of patients. All deaths in patients with HBV were liver related.

Table 5.9 Causes of death from death certificate by aetiology of CLD

Aetiology of CLD	Causes of death
HCV	9 liver related deaths 2 suicides 2 overdoses 1 misadventure (pneumonia) 1 accident 1 SBE 1 peritonitis due to ruptured appendix 1 small bowel obstruction 2 metastatic cancer ?primary 1 myocardial infarction **Total 20**
ALD	27 Liver Related death 3 Primary Cancer 2 Metastatic cancer 1 Renal failure 1 CVA 1 Open verdict **Total 35**
NAFLD	1 CHD 1 COAD **Total 2**
PBC	1 liver related death 1 cancer 1 cardiac failure **Total 3**
PSC	2 Liver Related death 1 Cancer 1 pancreatitis 1 MI **Total 5**
HBV	3 Liver Related death **Total 3**
AIH	1 Liver Related death Lymphoma AML **Total 3**
Other	1 Liver Related death 1 Pulmonary hypertension **Total 2**

Survival Analysis

The number of clinical events per ELF tertile and for each biopsy stage is reported in table 10. Most of the events occur in the middle ELF tertile, but the greatest proportion of events per person occurs in the highest tertile (15/18-83%). A similar pattern was observed in the biopsy stage 6 (30/55-55%).

Table 5.10a Number of clinical events per ELF tertile

ELF tertile	1 (lowest)		2		3 (highest)	
	No people	No events	No people	No events	No people	No events
LRO	203	4	277	48	18	15
ACM	203	4	277	52	18	14

Table 5.10b Number of clinical events per Biopsy stage (Ishak)

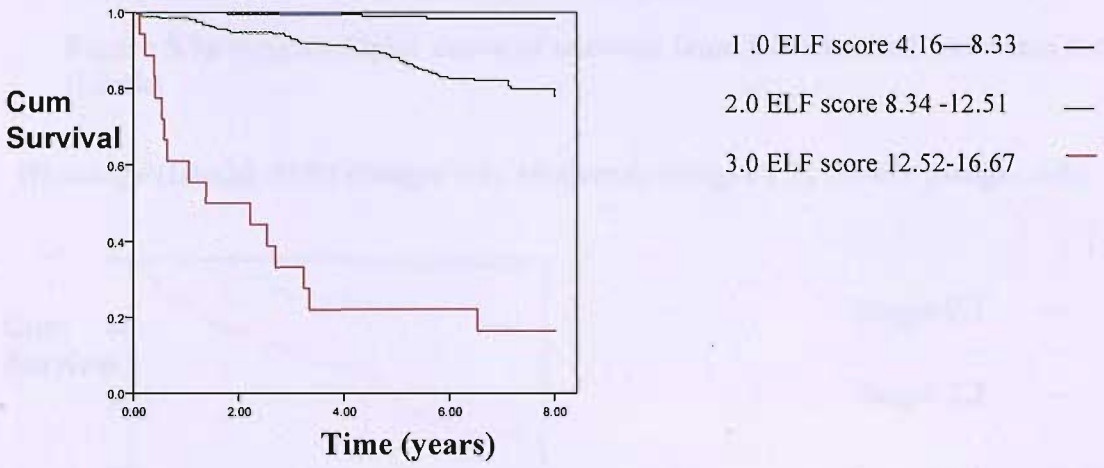
Biopsy stage (Ishak)	0		1		2		3		4		5		6	
	n	evnt	n	evnt	n	evnt	n	evnt	n	evnt	n	evnt	n	evnt
LRO	119	3	108	4	59	1	52	8	45	6	19	6	55	30
ACM	119	10	108	7	59	1	52	8	45	7	19	7	55	24

LRO Liver Related Outcome ACM All Cause Mortality evnt=clinical event

Figures 5.6a and 5.6b are Kaplan Meier curves showing the cumulative probability of survival from liver related death or hepatic decompensation (primary outcome) and all-cause mortality (secondary outcome) with tertiles of ELF representing low, medium, and high scores of ELF.

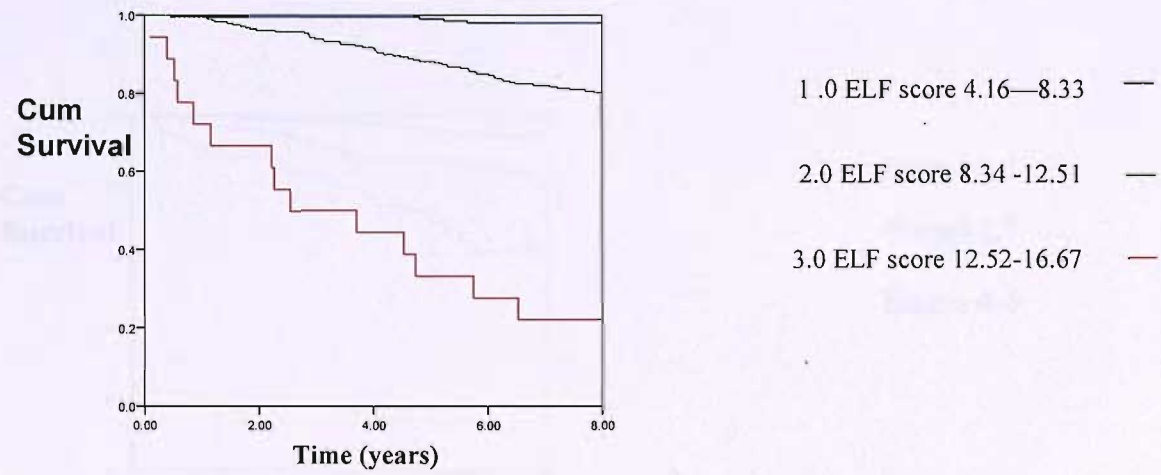
Crude unadjusted analyses by Kaplan Meier plots showed that tertiles of baseline ELF score can predict liver outcomes and all-cause mortality, with those people having ELF scores in the highest tertile at baseline being significantly more likely to have clinical outcomes than those in the middle tertile, who in turn were more likely to have outcomes than those with the lowest tertile score (log rank test (Mantel-Cox) p= <0.001)

Figure 5.6a Kaplan Meier curve of survival from liver related outcomes for ELF



Patients at risk	0	2	4	6	8
Lowest	203	189	179	153	31
Middle	277	249	231	177	34
Highest	18	9	5	4	1

Figure 5.6b Kaplan Meier curve of survival from all cause mortality for ELF



Patients at risk	0	2	4	6	8
Lowest	203	202	202	198	161
Middle	277	266	254	235	154
Highest	18	12	8	5	3

Kaplan Meir survival curves of liver related outcomes and all-cause mortality by tertile of ELF showing that survival is better in lowest ELF tertile and worst in highest.

Histology was also predictive of outcome when classed as mild, moderate or severe fibrosis (figures 5.7a and 5.7b). Differences between tertiles are significant p=>0.0001 Log Rank (Mantel-Cox) bottom tertile compared to 3rd and bottom compared to middle tertile.

Figure 5.7a Kaplan Meier curve of survival from liver-related outcomes for histology (Ishak)

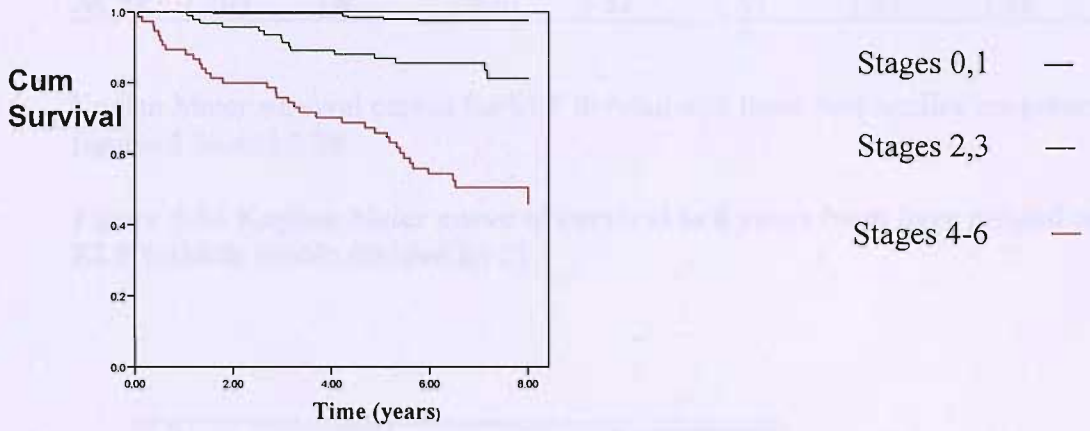
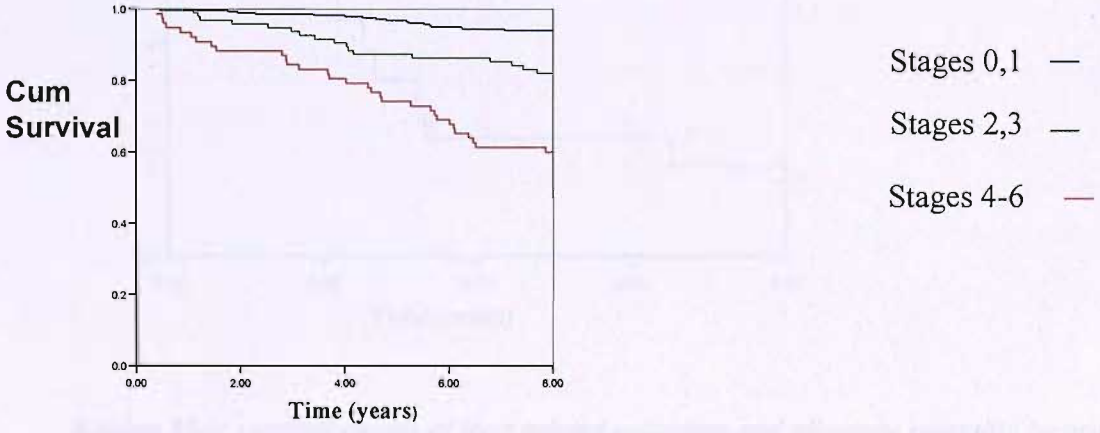


Figure 5.7b Kaplan Meier curve of survival from all-cause mortality for histology (Ishak)



Kaplan Meir survival curves of liver related outcomes and all-cause mortality by category of histology showing that survival is better in mildest disease group and worst in most severe disease.

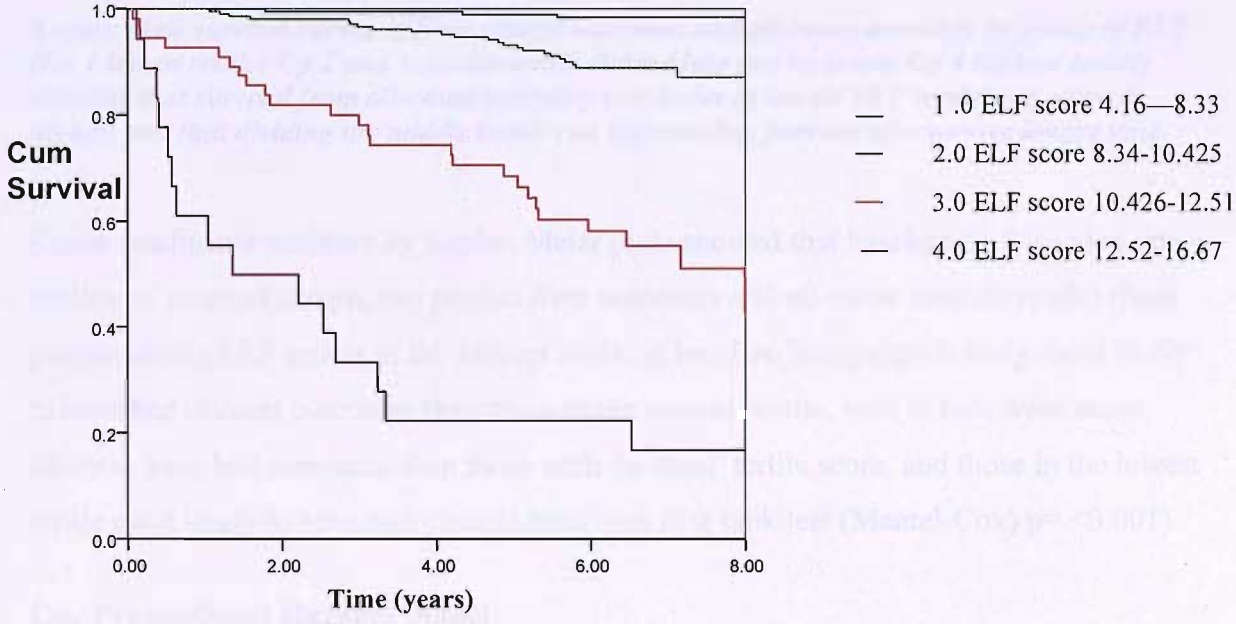
When the middle tertile of ELF is divided by 2 by score, the number of people per tertile and clinical events are more evenly distributed as in table 5.11.

Table 5.11 ELF panel divided into 4 groups (middle tertile divided into two by score)

	1 4.16-8.33		2 8.33-10.425		3 10.426-12.51		4 12.51-16.67	
	No people	No events	No people	No events	No people	No events	No people	No events
LRO	203	3	220	24	57	25	18	15
ACM	203	4	220	32	57	23	18	14

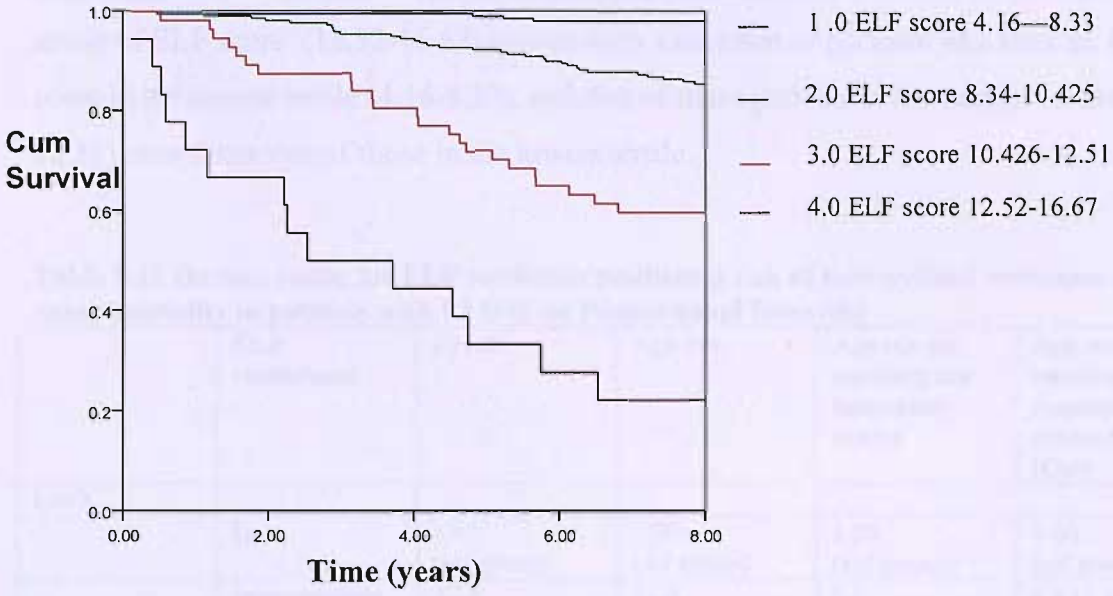
Kaplan Meier survival curves for ELF divided into these four tertiles are presented in figures 5.8a and 5.8b.

Figure 5.8a Kaplan Meier curve of survival to 8 years from liver related outcomes for ELF (middle tertile divided by 2)



Kaplan Meir survival curves of liver related outcomes and all-cause mortality by group of ELF (Gp 1 lowest tertile; Gp 2 and 3 middle tertile divided into two by score; Gp 4 highest tertile) showing that survival from liver –related outcomes was better in lowest ELF tertile and worst in highest and that dividing the middle tertile can differentiate persons who survive longer time.

Figure 5.8b Kaplan Meier curve of survival to 8 years from all-cause mortality for ELF (middle tertile divided by 2)



Kaplan Meir survival curves of liver related outcomes and all-cause mortality by group of ELF (Gp 1 lowest tertile; Gp 2 and 3 middle tertile divided into two by score; Gp 4 highest tertile) showing that survival from all-cause mortality was better in lowest ELF tertile and worst in highest and that dividing the middle tertile can differentiate persons who survive longer time.

Crude unadjusted analyses by Kaplan Meier plots showed that baseline ELF divided into 4 tertiles by score as shown, can predict liver outcomes and all-cause mortality, with those people having ELF scores in the highest tertile at baseline being significantly more likely to have had clinical outcomes than those in the second tertile, who in turn were more likely to have had outcomes than those with the third tertile score, and those in the lowest tertile most likely to have had clinical outcomes (log rank test (Mantel-Cox) p= <0.001).

Cox Proportional Hazards model

Proportional hazards assumption was checked and found to apply for tertiles of ELF. Multivariable hazard ratios for liver related outcomes adjusted for age, gender, baseline self-reported alcohol consumption and smoking status (derived from a priori based on prior knowledge of risk factors for progression) in patients with low intermediate and high ELF scores are presented in Table 5.12. Aetiology of CLD and centre were added as prognosis was found to vary with aetiology (patients with ALD having worse prognosis than other aetiologies) and centre of recruitment (see Appendix 6 Tables 1-3). The fully adjusted

hazard ratios for the highest ELF scores for liver related and all-cause mortality were 68.0 and 76.2 relative to the lowest tertile, and 9 and 7 for the middle tertile relative to lowest tertile.

This suggests that the hazard/risk of developing a liver-related outcome in the highest tertile of ELF score (12.52-16.67) is over sixty times that of patients who have an ELF score in the lowest tertile (4.16-8.33), and that of those patients in the middle tertile (8.34-12.51) nine times that of those in the lowest tertile.

Table 5.12 Hazard ratios for ELF tertiles in predicting risk of liver-related outcomes and all-cause mortality in patients with CLD (Cox Proportional Hazards)

	ELF tertile/cont	Crude	Age Sex	Age sex alc smoking aet responder centre	Age sex smoking alc responder centre biopsy (Cat)
LRO					
	Low	1.00 (ref group)	1.00 (ref group)	**1.00 (ref group)**	1.00 (ref group)
	Intermediate	12.9 (4.0 41.3)	11.4 (3.5,36.7)	**9.1 (2.8, 30.0)**	4.9 (1.1,22.0)
	High	133 (38.3,461.1)	136.4 (38.2,487.0)	**68 (17.0, 273.0)**	30.7 (5.5,173.0)
	Continuous	2.01 (1.8,2.2)	2.1 (1.86,2.37)	**1.87 (1.62,2.16)**	1.67 (1.38,2.03)
ACM	**Low**	1.00 (ref group)	1.00 (ref group)	**1.00 (ref group)**	1.00 (ref group)
	Intermediate	11.06 (4.0,30.5)	8.5 (3.1,23.7)	**7.2 (2.5,20.3)**	4.74 (1.6,14.0)
	High	86.03 (28.2,262.4)	75.7 (24.2,237.1)	**76.2 (20.7,280.3)**	28.0 6.9, 113.8)
	Continuous	1.65 (1.51,1.80)	1.64 (1.5,1.80)	**1.7 (1.45,1.90)**	1.5 (1.27,1.79)

Using the ELF score as a continuous variable in a Cox proportional hazards model results in a fully adjusted Hazard Ratio (HR) of 1.87 Adjusting for biopsy the HR became 1.7, indicating that ELF has something additional to biopsy to offer in the prediction of liver related outcomes.

As the test for trend across age categories was significant (Jonckheere Terpstra), an interactive term for age and ELF was added to the analysis and found to be non-significant. When those patients with ALD were excluded from the survival analysis the ELF tertiles are still able to detect those people who were at risk of developing liver related outcomes and differences between tertiles remained significant. In a Cox Proportional Hazards model

the HR for ELF was increased to 3.5, suggesting that for those patients who had an aetiology other than ALD the risk of outcome tripled with a unit change in ELF.

Prognostic accuracy of ELF in prediction of clinical outcomes

ROC curves were plotted to evaluate the diagnostic performance of ELF and biopsy in predicting clinical outcomes. As standard models do not exist for deriving ROC curves for time to event data, event versus no event within a specified time was used. The time selected was 6 years as the number of events and persons remaining in the analysis were maximized at this time (see table 5.13). ELF panel and biopsy are compared and ROC curve presented in Figure 5.9.

Table 5.13 Clinical events and censoring by year of follow up (England).

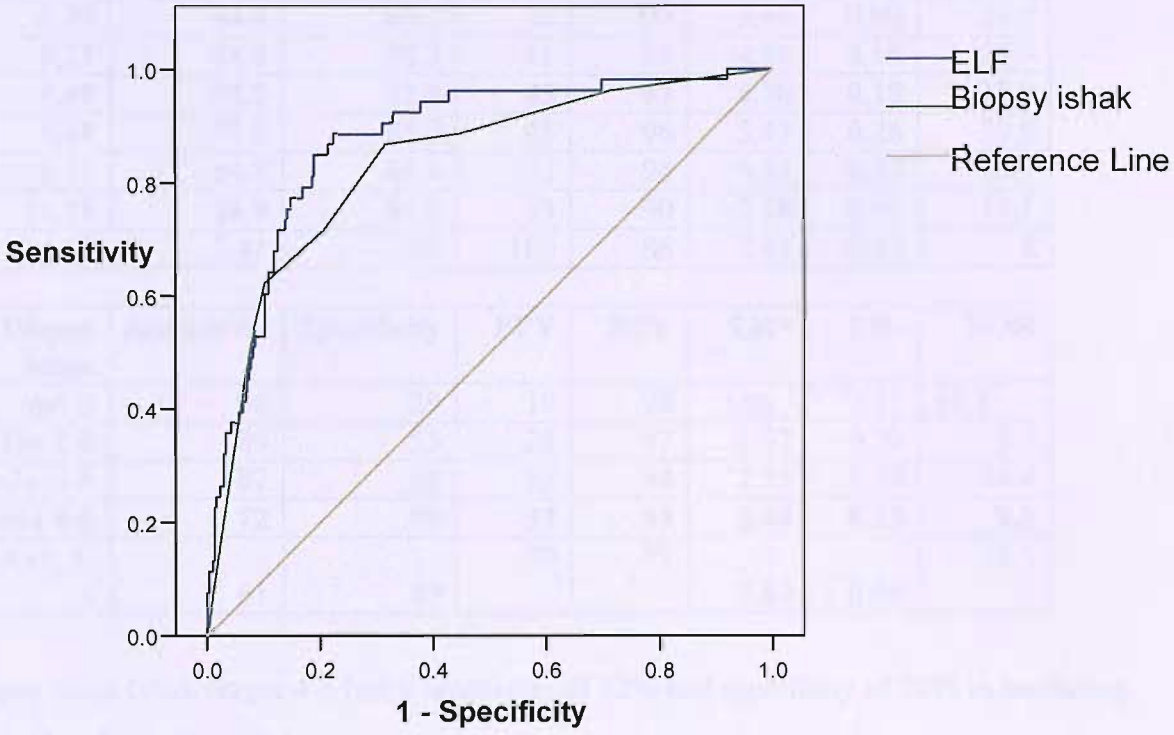
(i) ELF

Years of survival	No events	No people lost
Liver related outcomes		
5	48	117
6	61	103
7	64	208

(ii) BIOPSY

	No events	No people lost
Liver related outcomes		
5	42	107
6	54	154
7	56	271

Figure 5.9 ROC curves of ELF and Biopsy (Ishak) at 6 years Liver related outcomes



ROC curve of histology and ELF panel in prediction of liver–related outcomes at 6 years

The unadjusted AUC for ELF was high (0.88 (95% CI 0.83, 0.93)) and showed very good performance in predicting liver-related outcomes at 6 years. It was greater than the AUC for biopsy (0.88 vs. 0.83 (95% CI 0.77, 0.89)) but this difference did not reach statistical significance (p=0.1).

Diagnostic odds ratios (reported in Table 14) showed very good test performance in predicting clinical outcome, and using a threshold of 9.49 the sensitivity was 85% and specificity was 82% with a very high NPV of 97% and PPV 45 %. Therefore ELF showed very good performance in the identification of those people who are unlikely to have a clinical event within 6 years, and can predict events in patients with 85% sensitivity.

Table 5.14 Diagnostic performance of ELF and biopsy predicting liver related clinical outcomes at 6 years

ELF threshold	Sensitivity	Specificity	PPV	NPV	LR+	LR-	DOR
7.83	98.2	29.9	19	99	1.22	0.49	2.5
8.79	94.4	63.1	31	99	2.56	0.09	29.0
9.27	88.9	78.3	41	99	4.10	0.14	28.9
9.49	**85.2**	**82.0**	**45**	**97**	**4.70**	**0.18**	**25.9**
9.88	77.8	85.7	48	96	5.43	0.26	20.9
10.25	64.8	89.5	52	94	6.16	0.39	15.7
11.18	38.9	94.6	55	90	7.18	0.65	11.1
14.18	7.41	100	100	86	7.41	0.93	8

Biopsy stage	Sensitivity	Specificity	PPV	NPV	LR+	LR-	DOR
0v1-6	96	29	19	98	1.35	0.13	10.5
0,1vs 2-6	89	55	25	97	1.97	0.20	9.7
0-2vs3-6	87	68	32	94	2.73	0.19	14.4
0-3vs 4-6	**72**	**79**	**37**	**93**	**3.44**	**0.35**	**9.8**
0-4 vs. 5-6	61	89	50	91	5.82	0.44	13.4

Biopsy stage Ishak stages 4-6 had a sensitivity of 72% and specificity of 79% in predicting liver related outcomes at 6 years.

The ability of ELF to predict all-cause mortality at 6 years was good, with an AUC of 0.82 (95% CI 0.76, 0.88) compared to biopsy which had an AUC of 0.70 (95% CI 0.62, 0.79) (p=0.0004) (figure 5.10).

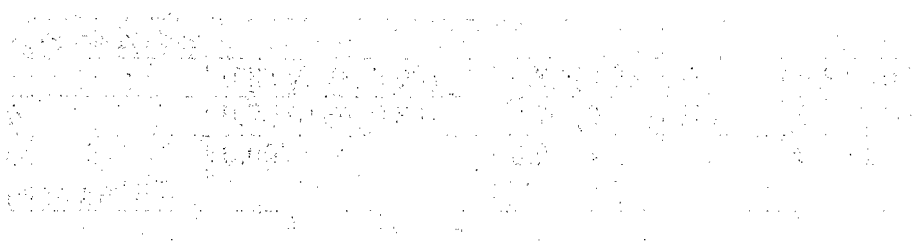
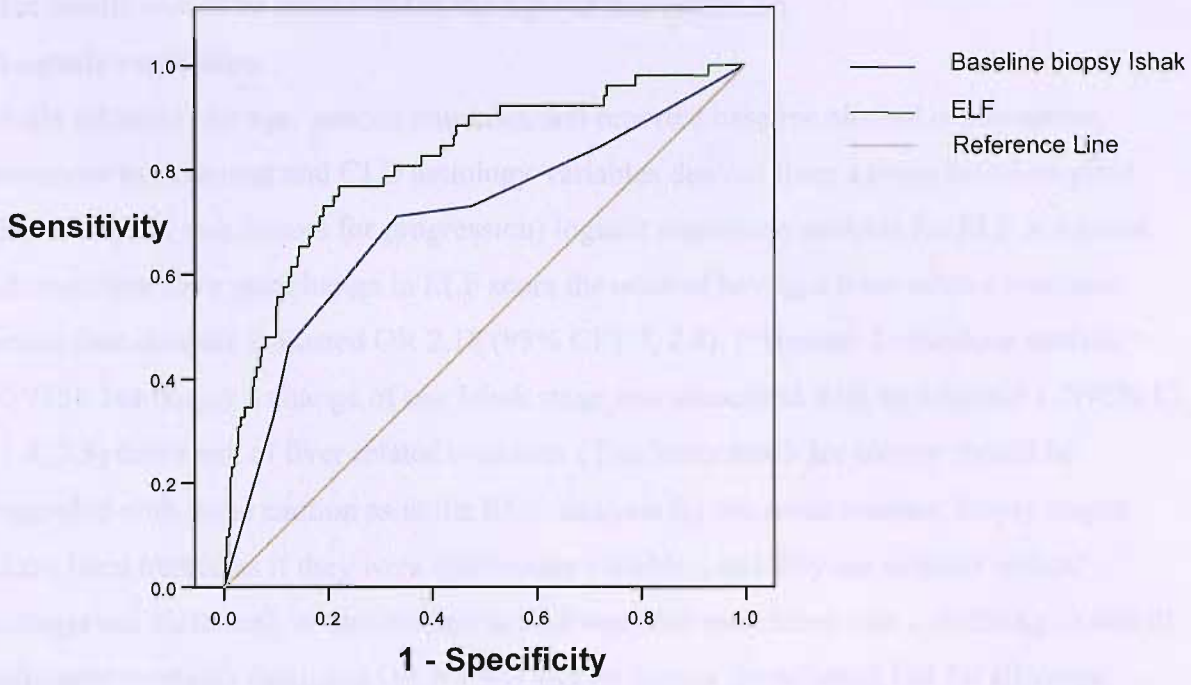


Figure 5.10 ROC curves of ELF and Biopsy (Ishak) at 6 years All-cause mortality



Performance of ELF panel compared to histology (treated as continuous variable) in the prediction of all-cause morality at 6 years. *Difference in AUC significant (p=<0.01)*

Table 5.15 shows complete results for AUC values for ELF and biopsy at 5, 6 and 7 years survival. The predictive performance of ELF for liver-related and all-cause mortality clinical outcomes at 5, 6 and 7 years are good.

Table 5.15 Area under the Curve for ELF and Biopsy predicting clinical events at different survival times

	AUC at 5 y survival (95% CI)	**AUC at 6 y survival (95% CI)**	**AUC at 7 y survival (95% CI)**
Liver related outcome			
ELF	0.87 (0.81,0.93)	0.86 (0.81,0.92)	O.88 (0.82,0.93)
Biopsy	0.83 (0.77, 0.90)	0.82 (0.76,0.89)	0.84 (0.77, 0.90)
p value (comparing AUC)	0.3	0.1	0.3
All cause mortality			
ELF	0.83 (0.78, 0.90)	0.82 (0.75,0.88)	0.84 (0.78,0.89)
Biopsy	0.74 (0.66, 0.83)	0.70 (0.62,0.79)	0.73 (0.65, 0.80)
p value (comparing AUC)	0.007	0.0004	0.0002

ROC analysis for biopsy should be regarded with some caution as biopsy stages have been treated as if they were continuous variables with 7 thresholds, and they are ordinal

categorical variables. This is an acceptable method of directly comparing ELF and biopsy but results should be considered in the light of this limitation.

Logistic regression

Fully adjusted (for age, gender, smoking, self reported baseline alcohol consumption, response to treatment and CLD aetiology-variables derived from a priori based on prior knowledge of risk factors for progression) logistic regression analysis for ELF at 6 years showed that for a unit change in ELF score the odds of having a liver-related outcome more than doubles (adjusted OR 2.18 (95% CI 1.7, 2.8). (*Hosmer Lemeshow statistic = 0.776). For biopsy a change of one Ishak stage was associated with an adjusted 1.7(95% CI 1.8, 2.5) times risk of liver related outcome. (This latter result for biopsy should be regarded with some caution as in the ROC analysis for the same reasons; biopsy stages have been treated as if they were continuous variables, and they are actually ordinal categorical variables). A unit change in ELF was also associated with a doubling of risk of all-cause mortality (adjusted OR = 1.96) and for biopsy the adjusted OR for all-cause mortality was 1.30 (1.12, 1.51).

Adjusted odds ratios using logistic regression for ELF (tertiles) and biopsy (mild moderate and severe stages) as categorical variables were; ELF highest tertile compared to lowest tertile OR=90 (95% CI 16,512) and middle compared to lowest OR= 9 (3, 30). For biopsy severe compared to nil/mild OR=22 (95% CI 8, 62) and moderate/severe compared to nil/mild OR=8 (2.7, 33).

Derivation of new predictive model

To evaluate whether there was a model of biomarkers that could improve the prediction of clinical outcome, all nine of the serum markers and simple markers were entered into a multivariable regression analysis using the Collett method of selection. MMP_TIMP1, AST, ALT, alkaline phosphatase and creatinine variables were not significant at the 0.1% level and did not add to the model when introduced at step 4.

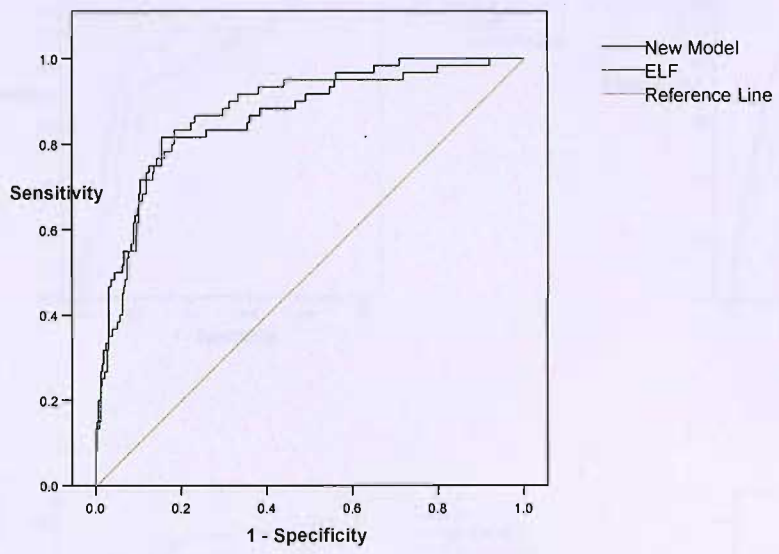
The final model was fitted on 333 patients with 50 liver related events and included HA, platelets and laminin.

Algorithm: (LnHA*0.640) + (Ln laminin*0.819) - (platelets*0.008) -5.402

(Hosmer Lemeshow statistic =0.521).

The ROC curve of this new model and ELF are shown in Figure 5.11.

Figure 5.11 ROC curves of ELF and New Model at 6 years Liver-related outcomes



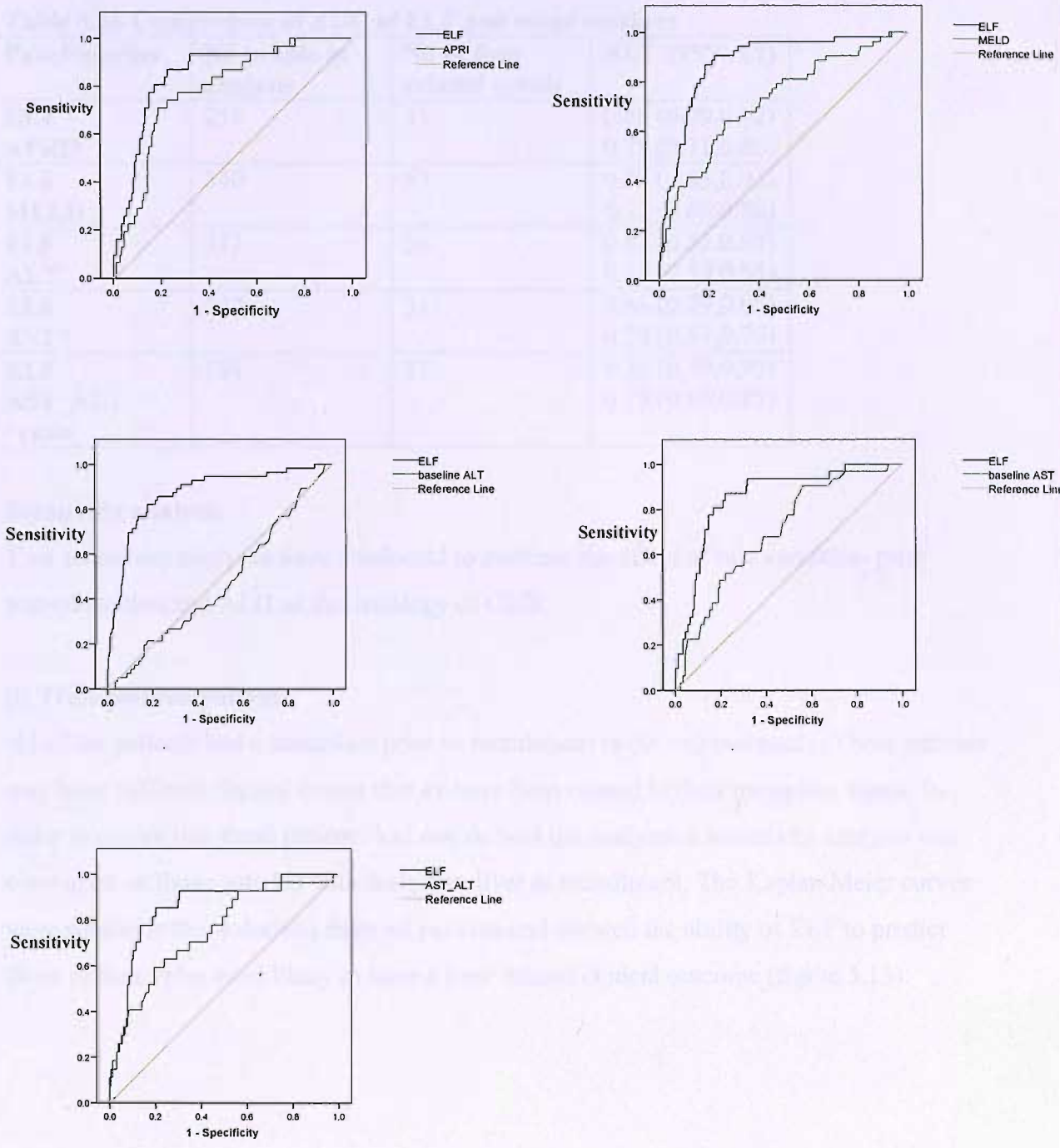
ROC curve of ELF panel compared to new model derived on n=498 in the prediction of liver – related outcomes at 6 years

The new model and ELF have the same AUC (0.87-95%CI 0.82, 0.92) although it may be seen from the figure that the shapes of the curve are slightly different with ELF having a better performance in the mid-range.

Comparison of ELF with other panels/clinical score

The performance of ELF has never been directly compared to other biomarkers. Available tests were used to conduct such a comparison in this cohort. For several marker panels (especially those involving AST) availability of constituent tests was limited and just over half of the cohort could be utilised. Comparisons were made with APRI (baseline AST/Upper limit of normal range/baseline platelets)*100), AST, ALT,AST_ALT ratio, and MELD score (MELD = 3.8[Ln serum bilirubin (mg/dL)] + 11.2[Ln INR] + 9.6[Ln serum creatinine (mg/dL)] + 0.643 (Figure 5.12).

Figure 5.12 ROC curves comparing ELF and other serum markers



ELF panel performance at prediction of live related outcomes at 6 years compared to APRI, MELD, AST, ALT, AST-ALT ratio.

Table 5.16 shows the AUC values and the number of patients. In each comparison ELF has a better AUC than all of the comparator markers and panels. It is not known whether those patients with available AST values are randomly or systematically different from those

with no AST values. Consequently there may be a bias in these values leading to under or over estimation of comparative performances.

Table 5.16 Comparison of AUC of ELF and other markers

Panel/marker	No people in analysis	No of liver related events	AUC (95% CI)
ELF	258	31	0.86 (0.79,0.92)
APRI*			0.79 (0.71,0.86)
ELF	340	53	0.88 (0.83,0.93)
MELD			0.72 (0.65,0.80)
ELF	372	56	0.87 (0.82,0.93)
ALT			0.47 (0.39,0.55)
ELF	217	31	0.86 (0.79,0.93)
AST*			0.70 (0.61,0.79)
ELF	191	27	0.86 (0.79,0.93)
AST_ALT			0.75 (0.65,0.85)
***ratio**			

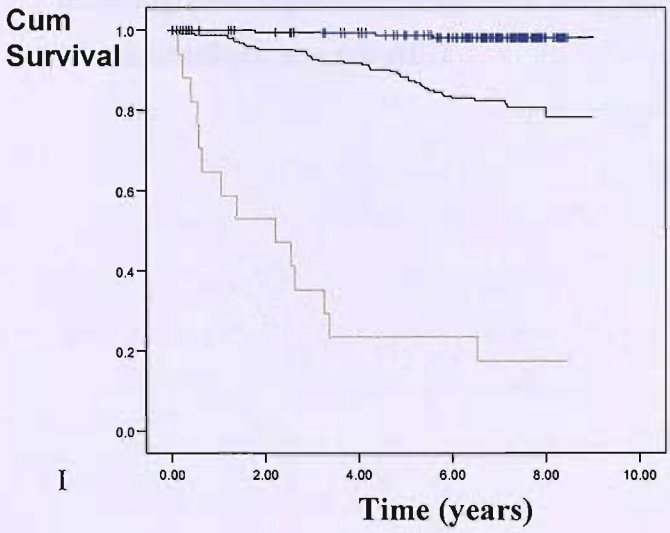
Sensitivity analysis

Two sensitivity analyses were conducted to evaluate the effect of two variables- prior transplantation and ALD as the aetiology of CLD.

(i) Transplanted patients

41 of the patients had a transplant prior to recruitment in the original study. These patients may have suffered clinical events that ay have been related to their transplant status. In order to ensure that these patients had not skewed the analyses a sensitivity analysis was conducted on those patients with their own liver at recruitment. The Kaplan Meier curves were similar to those derived from all patients and showed the ability of ELF to predict those patients who were likely to have a liver related clinical outcome (figure 5.13).

Figure 5.13 Kaplan Meier curve of survival from liver-related outcomes in patients that did not begin study with transplanted liver



Kaplan Meir survival curves of liver related outcomes by tertile of ELF excluding those patients who were recruited with a transplanted liver, showing that survival was better in lowest ELF tertile and worst in highest.

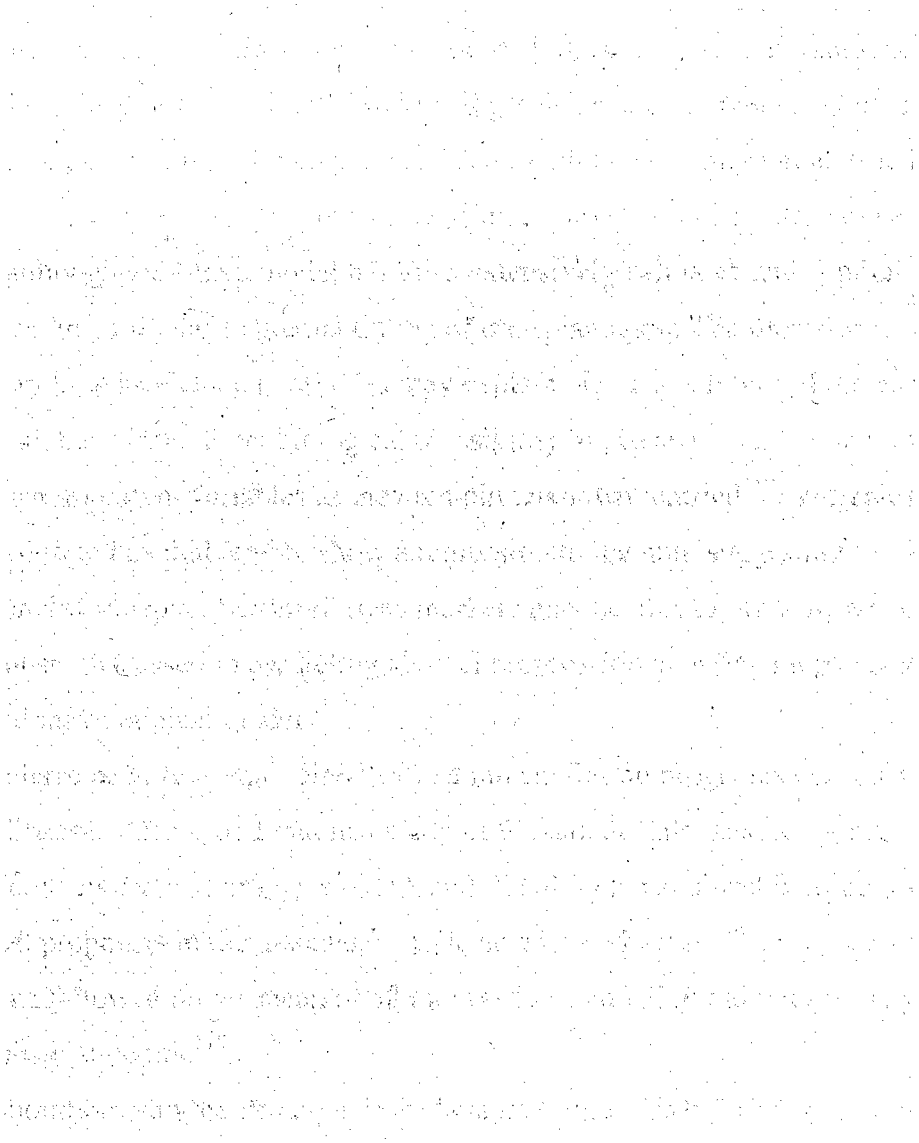
The differences between survival in the ELF tertiles were significant. AUC from a ROC analysis at 6 years for liver-related outcomes were 0.87 (95% CI 0.81, 0.92) (n=355) Cox proportional hazards analysis had an adjusted HR of 35.8 (95% CI 9.4, 137.1) and 6.22 95% CI 1.9, 20.8). Using logistic regression at 6 years for Liver–related outcomes the Odds Ratio was 2.0 (1.6, 2.6) adjusted for age, gender, aetiology, responder to treatment status, smoking and alcohol. This does differ materially from the logistic regression for all patients which also showed that a unit change in ELF was associated with a doubling of risk of liver-related clinical outcome.

(ii) Patients with ALD

There were 55 patients with ALD who had 38 (54%) of the liver related outcomes. Crude unadjusted analyses by Kaplan Meier plots showed that tertiles of baseline ELF score could predict liver outcomes and all-cause mortality, with those people having ELF scores in the highest tertile at baseline being significantly more likely to have had clinical

outcomes than those in the middle tertile, who in turn were more likely to have had outcomes than those with the lowest tertile score (log rank test (Mantel-Cox) p= <0.001). This was maintained when four groups of ELF are analyzed by dividing the middle tertile into 2 by score.

In a CPH analysis without patients with ALD the fully adjusted HR did not decrease. Similarly from logistic regression the fully adjusted OR was not reduced in those patients whose aetiology was not ALD.



5.6 The performance of ELF serum marker panel in predicting clinical outcomes in Primary Biliary Cirrhosis (STUDY 2)

5.6.1 Background

The diagnosis of PBC can often be made based on a positive anti-mitochondrial antibody (AMA) test in the appropriate clinical setting. However, liver biopsy is still used to assess the extent of liver fibrosis and provide prognostic information. There are several prognostic models that include liver function tests and clinical variables, that have been used in the prediction of short and long term survival in PBC. Some are time-fixed (Yale (bilirubin, age, hepatomegaly, cirrhosis,) European (bilirubin, age, albumin, cirrhosis, central cholestasis) , Mayo (age, bilirubin, albumin, PTT and peripheral oedema), Glasgow (bilirubin, age, ascites, varicael bleeding, marked cholestasis, hepatic fibrosis)) and some are time-dependent models Mayo, European (bilirubin, ascites, albumin, age, gastro-intestinal bleeding)) which give improved precision and are more suitable for longer term monitoring of disease. The only variable common to all of these models is bilirubin. There have been no direct comparisons of these models and no one model can be identified as the best, although the Mayo model has been extensively validated and is often use in clinical practice, in particular to predict timing of transplantation. The precision of these models is not very high as included variables may explain only a small part of the variation in survival, have little direct biological plausibility in disease processes and may not include the most important variables as they remain unknown /untried[189]. Variables at diagnosis or admission to hospital tend to show a regression to the null, suggesting that later values may be more informative. Serum fibrosis markers may be able to offer the advantages previously discussed in predicting clinical progression avoiding biopsy complications and limitations of clinical models.

There have been few small direct evaluation studies on prognosis using serum markers of liver fibrosis. One found that in a study of 97 patients allocated to treatment with ursodeoxycholic acid or placebo, HA and P3NP replaced bilirubin as predictors of poor clinical prognosis in the ursodeoxycholic acid treated group[190]. However in a study of 55 patients followed up for mean of 58 months HA and P3NP did not give any indication of prognostic outcome[175].

Collaboration with the Principle Investigators of the PUMPS RCT based in University of Texas Southwestern in Dallas Texas permitted access to paired samples of biopsies and sera taken from a longitudinal cohort of patient with PBC. The performance of ELF in the identification of fibrosis on biopsy has been shown to be good (results presented in Chapter

4 of this thesis). Study 2 was undertaken to evaluate serum fibrosis markers as predictors of clinical progression in large cohort of PBC patients and to compare with commonly used clinical models.

5.6.2 Aim of the study

To evaluate the prognostic performance of the ELF panel in patients with PBC in predicting liver-related outcomes.

5.6.3 Methods

161 patients with PBC were prospectively followed between 1993 and 2003 as part of a multi-centre US clinical trial that was designed to investigate whether low dose weekly methotrexate, when added to ursodiol, improved survival or delayed progression of PBC. Methotrexate was not found to affect the course of PBC, so patients from both treatment arms were combined for the purpose of this study. Combining treatment arms was also later justified by a sub-analysis demonstrating no effect of the treatment arm on the outcomes of this analysis. Inclusion criteria were positive AMA and an abnormal alkaline phosphatase or at least stage 1 disease on liver biopsy. Patients with decompensated cirrhosis were excluded.

Serial liver biopsies and serum samples were collected every 2 years, when endoscopy was performed to ascertain new varices, and abdominal ultrasound to ascertain new ascites. A single pathologist graded all histology using Ishak classification. Clinical progression was defined as development of one or more of the following events: new varices, ascites, encephalopathy, variceal bleed, liver transplant, or liver-related death. The ability of the simplified ELF panel, histological fibrosis, bilirubin, MELD, and Mayo Risk Score to differentiate between individuals who would experience clinical decompensation and those who would not was evaluated at different time points.

Statistical methods were the same as in Study 1, with crude unadjusted Kaplan Meier survival analysis being presented at 7 years follow up, Cox proportional hazards model at 7 years adjusting for age, and ROC analyses at specific time points. Comparisons with other prediction tests and clinical scores were made using ROC analyses at different time points. AUC values were compared using non-parametric approach.

5.6.4 Results

The median follow-up of patients was 7.3 years (0.4-10 years). 93% were female and 93% were Caucasian. 43 patients developed liver related clinical outcomes (27%).The mean scores of all prognostic tests were significantly higher at baseline in those that developed a clinical outcome. There were difficulties in determining the boundaries of the tertiles to be used for ELF – ideally those derived from the original ELF scores would be used (larger dataset and it would be ideal to have a reference tertile scoring system that could be applied to external populations). However, this created problems as the PBC population had a different ELF distribution (in PBC the median = 9.1 and the range was 6.83 to 13.02 compared to original ELF cohort median 8.48and the range was 4.16 to 16.67). The boundaries of the two methods of deriving tertile score boundaries are show in Table 5.17. Using the dataset driven boundaries (a) the Kaplan Meier survival curve for survival at 7 years for liver related events showed a similar pattern to that in the ELF follow-up cohort (figure 5.14). However when the Original cohort boundaries are used (b) the very few people in the top tertile affected the survival curve (figure 5.15).

Table 5.17(a) PBC dataset derived tertiles (n=161)

ELF Tertile (derived from PBC dataset)	**Clinical Event**	**No people**
1 (6.83-8.89)	7	62
2 (8.9-10.96)	26	85
3 (10.97-13.02)	10	14
Total	43	161

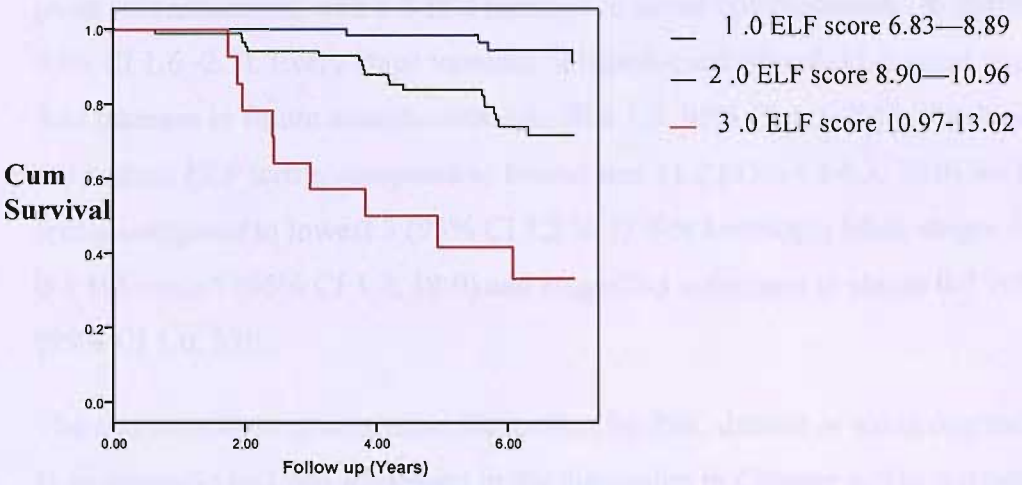
Table 17(b) Original ELF dataset derived tertiles (n=921)

ELF Tertile (derived from PBC dataset)	**Clinical Event**	**No people**
1 (4.16-8.33)	4	38
2 (8.34-12.51)	37	121
3 (12.51-16.67)	2	2
Total	43	161

Table 5.18 Clinical events and number of people by category of histology

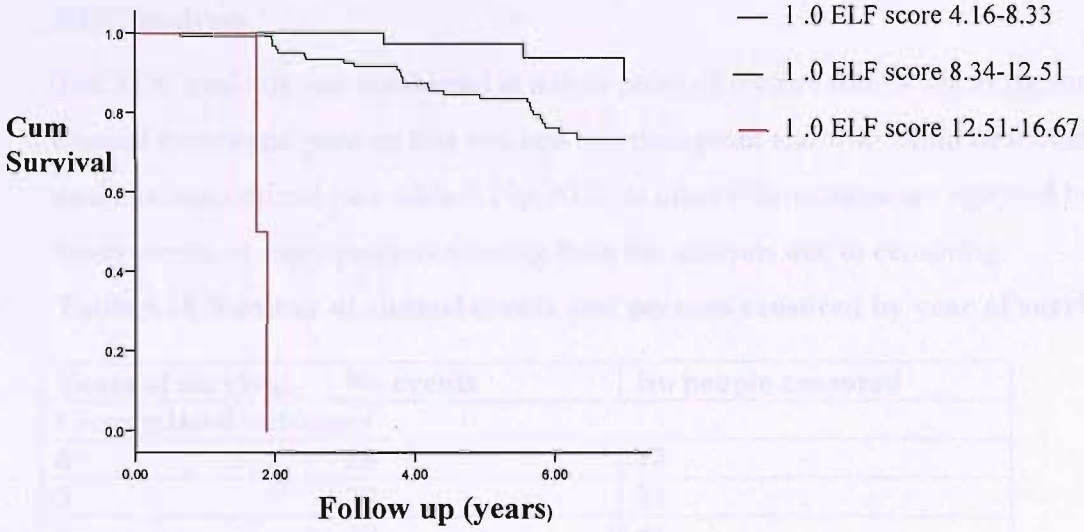
Histology	**Clinical Event**	**No people**
Stages 0-1	5	37
Stages 2-3	22	86
Stages 4-6	14	14
Total	41	152

Figure 5.14 Kaplan Meier curve of survival from liver-related outcomes for ELF tertiles
*using tertile values from PBC dataset



Kaplan Meir survival curves of liver related outcomes by tertile of ELF showing that survival was better in lowest ELF tertile and worst in highest.

Figure 5.15 Kaplan Meier curve of survival from liver-related outcomes for ELF tertiles (derived from original ELF cohort n=921)



Kaplan Meir survival curves of liver related outcomes by tertile of ELF (using the original ELF cohort to derive tertile score boundaries) showing that survival was better in lowest ELF tertile and worst in highest. The steep fall off in the highest tertile is due to the very few (n=2) persons in this tertile.

The log rank test comparing the three subgroups showed a highly significant difference ($p<0.0001$) in event-free survival depending on the baseline ELF score.
Survival curve for lowest and middle tertile show significant difference in survival from liver related clinical events.
In a Cox's Proportional Hazard model adjusting for age each increase in ELF score by 1 point was associated with a 2-fold increase in future complications (Relative risk 2.09, 95% CI 1.6 -2.7). Every stage increase of histological fibrosis (1-6 point scale) led to a 1.5-fold increase in future complication rate (RR 1.6, 95% CI 1.3 -2.1). The Hazard ratio for the highest ELF tertile compared to lowest was 11.2 (95% CI 4.3, 30.0) for ELF middle tertile compared to lowest 3 (95% CI 1.2 to 7). For histology, Ishak stages 4-6 compared to 0-1 HR were 5 (95% CI 1.7, 14.0) and stages 2-3 compared to stages 0-1 HR were 1.8 (95% CI 1.0, 5.0).

The choice of how to determine the tertiles, by PBC dataset or using original ELF dataset, is problematic and was addressed in the discussion in Chapter 4. The dataset driven tertiles allow models to be fitted to the PBC data whereas the original ELF dataset scores may reflect scores derived from a much larger dataset but a mixed aetiology where the PBC numbers were small and CHC made up the majority of the patients. This gives those analyses which use ELF as a continuous variable more external validity. Both are presented for comparison.

ROC analyses

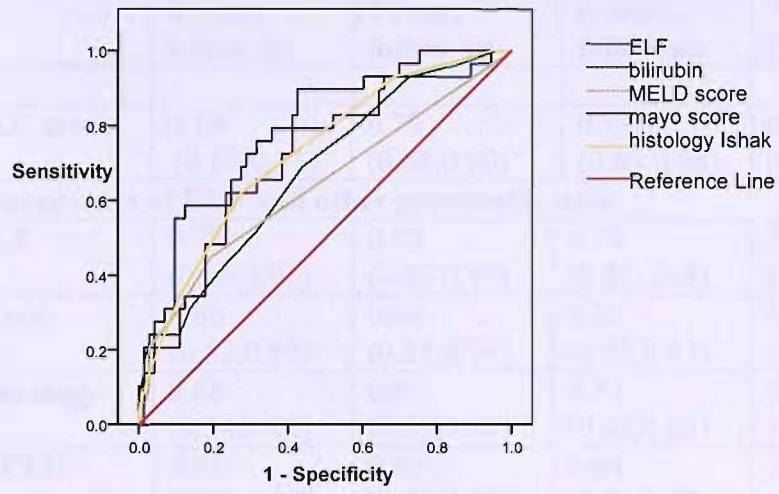
The ROC analysis was conducted at a time point of 6 years follow up, as the number of clinical events and persons that reached this time point and who could be included in the analysis was optimal (see table 5.19). AUC at other follow times are reported but have fewer events or more patients missing from the analysis due to censoring.

Table 5.19 Number of clinical events and persons censored by year of survival

Years of survival	No events	No people censored
Liver related outcomes		
4	**18**	**13**
5	**22**	**25**
6	**30**	**46**
7	**33**	**68**
8	**39**	**103**

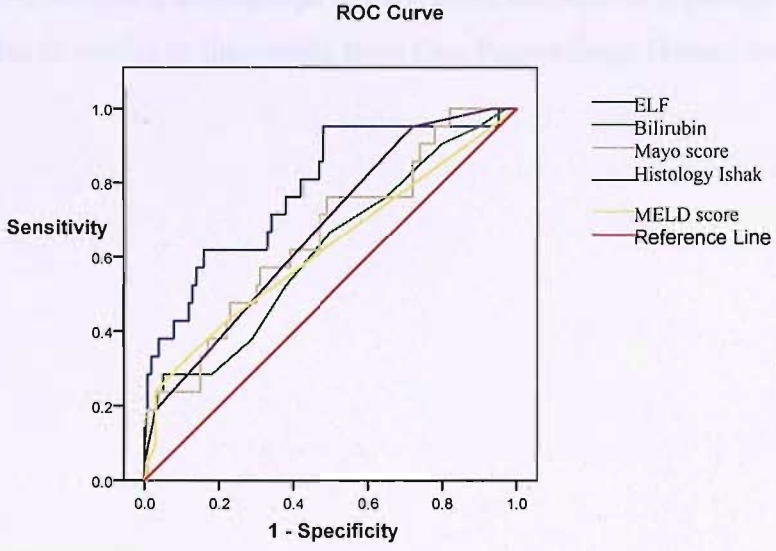
Figure 5.16 and 5.17 show ROC curves for ELF compared to different biomarkers.

Figure 5.16 ROC curve at 6 years follow up of patients with PBC comparing prognostic tests in prediction of liver –related events



ROC curves of ELF panel compared to bilirubin, MELD score, Mayo score and histology in the prediction of liver related events at 6 years showing ELF has a better performance than the other non-invasive markers

Figure 5.17 ROC curve at 5 years follow-up of patients with PBC comparing prognostic tests in prediction of liver–related events



ROC curves of ELF panel compared other biomarker scores in the prediction of liver related events at 5 years showing ELF has a better performance than the other non-invasive markers

The AUC values of the comparisons of ELF and other biomarkers in the prediction of liver related outcomes at various time points are shown in Table 5.20. ELF alone has the best performance at all time points compared to commonly used biomarkers.

Table 5.20 AUC values with 95% CI comparing ELF and other prognostic tests/clinical scales at different years of follow up

	4 years follow up	5 years follow up	6 years follow up	7 years follow up	8 years follow up
ELF alone	0.79 (0.67,0.91)	0.79 (0.68,0.90)	0.79 (0.69,0.88)	0.78 (0.67,0.88)	0.83 (0.71,0.95)
Comparison of ELF and other prognostic tests					
ELF	0.78 (0.66,0.91)	0.78 (0.67,0.89)	0.78 (0.67,0.88)	0.79 (0.69,0.90)	0.82 (0.70,0.95)
Mayo	0.66 (0.52,0.80)	0.66 (0.53,0.79)	0.72 (0.62,0.83)	0.66 (0.54,0.78)	0.74 (0.59,0.90)
Histology	0.68 (0.54,0.82)	0.67 (0.55,0.80)	0.73 (0.62,0.83)	0.73 (0.62,0.84)	0.71 (0.57,0.86)
MELD	0.61 (0.45,0.77)	0.62 (0.48,0.77)	0.64 (0.51,0.77)	0.59 (0.46,0.72)	0.60 (0.44,0.75)
Bilirubin	0.64 (0.49,0.78)	0.62 (0.48,0.76)	0.68 (0.56,0.79)	0.59 (0.46,0.72)	0.71 (0.57,0.87)

Logistic Regression

From Logistic regression the Odds ratio for ELF (continuous) adjusted for age was 2.2 (1.4, 3.5)-for a unit change in ELF score the odds of a clinical event more than doubled. This is similar to the results from Cox Proportional Hazard model.

5.7 Discussion

Statement of main findings

Follow up clinical outcomes data were collected on 498 patients in England and 404 in Europe originally recruited to a diagnostic study. Robust information on mortality and morbidity in those lost to follow up or discharged was possible in England, and very incomplete in Europe. The all-cause mortality rate in England centres was 15% (n=73) and 13.5% had liver related outcomes (n=67). The prognostic ability of the baseline ELF panel and biopsy were evaluated in England. Crude unadjusted analyses by Kaplan Meier plots showed that tertiles of baseline ELF score can predict liver outcomes and all-cause mortality, with those people having ELF scores in the highest tertile at baseline being significantly more likely to have clinical outcomes than those in the middle tertile, who in turn were more likely to have outcomes than those with the lowest tertile score (log rank test p= <0.001). Histology was also predictive of outcome when classed as mild, moderate or severe fibrosis. Analyses using Cox Proportional Hazards model show that ELF remained predictive when fully adjusted for possible confounders (p=<0.0001); the adjusted hazard ratio (HR) for the middle tertile ELF score versus lowest tertile ELF score= 8.3 (95% CI 2.5-26.5), and for highest tertile ELF score versus lowest tertile, HR=57 (95% CI 15-215). The hazard ratio decreased after adjustment for potential confounding. Adjusted odds ratio from logistic regression at 6 years showed a doubling of risk with unit change in ELF score. AUC from ROC analysis showed ELF had a very good performance in predicting liver related outcomes at 6 years (0.88 (0.83,0.93). At 6 years follow up ELF was better at predicting all-cause mortality than biopsy (0.82 vs. 0.70).

A similar trend in results was obtained by a follow up study of 161 patients with PBC where ELF was shown to be predictive of clinical outcomes. Crude unadjusted analyses by Kaplan Meier plots showed that tertiles of baseline ELF score can predict liver outcomes and all-cause mortality, with those people having ELF scores in the highest tertile at baseline being significantly more likely to have clinical outcomes than those in the middle tertile, who in turn were more likely to have outcomes than those with the lowest tertile score (log rank test p= <0.001). ELF score as a continuous variable in a Cox Proportional Hazard model showed a fully adjusted Hazard Ratio of 2.1. ROC analyses showed that ELF was the best predictor of clinical events compared to other accepted clinical scores, with an AUC at 6 years follow up of 0.78 (95% CI .67,.88). Adjusted odds ratio from

logistic regression at 6 years showed more than a doubling of risk with unit change in ELF score.

Strengths and Limitations

Study 1 This study was conducted in large number of patients representative of patients referred to hepatologists for investigation of liver disease by biopsy. It involved many centres and reflected general hepatology practice. In the English centres despite high attrition rates from recruiting centres, robust follow up was achieved in a traditionally hard to reach population. The excellent response rate from General Practitioners regarding the liver morbidity contributed to this, permitting the ascertainment of clinical outcomes in most of those failing to attend or discharged from recruiting centres, with only one reported clinical event. There was useful control of confounders, and the derived relative risks were large enough to discount chance and bias. This makes this study one of the most robust prognostic studies that has been conducted in this field.

In contrast there were difficulties in ascertainment of data from the European centres. This varied between centres but there were some common themes. The most significant was the problem of ascertainment of outcomes in those patients lost to follow up from the recruiting centre. In three out of the four countries there was no system of national mortality that could be interrogated regarding the mortality status of these patients. The mortality rate that could be determined was much lower than that found in the English centres (2.5% vs. 14%). The disease spectrum was similar in Europe and England so assuming that the rate of progression of disease in patients in England and Europe are similar, it is likely that many deaths had not been ascertained from Europe. The non-fatal event rate in those lost to follow up is similarly unknown. Use of such data with so much missing information would have compromised validity. Consequently the available numbers for analysis was reduced, although power was still adequate as evidenced by Hazard ratio results. Sensitivity analysis using the data from Europe did not change the trend of the results. Kaplan Meier plots showed that tertiles of baseline ELF score can predict liver outcomes and all-cause mortality, and that histology could predict outcome.

Other local issues in data collection in Europe also contributed to the decision to exclude those centres. In some countries the referral of patients is not restricted to locality catchment area, with patients attending from all over the country for assessment and initial management, following which they were looked after by more local centres. Hepatology expertise also increased over the decade following recruitment leading to dissemination of

patients countrywide. This problem also pertained to those centres that were large regional transplant centres.
Loss of clinical records (including by fire and flood in 2 centres) was a problem in all centres but electronic records in some centres provided follow-up information.

Ideally prognostic studies should be evaluated using an inception cohort with disease at same stage. This latter condition is difficult in CLD due to the asymptomatic nature of most of the natural history stages of liver disease with clinical symptoms only appearing at the end stage of disease (cirrhosis). Therefore patients present to health services at different stages depending on the trigger for investigation. The cohort in this study was not a true inception cohort (where all patients are recruited at the same point of disease) and this may make interpretation difficult and there may be survivor bias. The cohort was established primarily as a cross-sectional cohort to derive and validate diagnostic tests to identify fibrosis on biopsy. Inclusion criteria did not demand that the recruited patients were incident cases and therefore it may not be a perfect inception cohort. If the cohort were not an inception cohort bias may have occurred. Results from an analysis in one centre showed that most patients had spent less than four months under the care of hepatology before recruitment to the study, implying that the cohort may be more like an inception cohort that not.

The original study recruited consecutive patients at the point of biopsy, with few patients at the severest end of the disease and most at mild or moderate stages. Thus the numbers of patients in the highest ELF tertile or with cirrhosis on biopsy were small. Nevertheless this is likely to represent clinical practice in hepatology centres. Data collection was conducted by one researcher per patient records. Double data collection would have been ideal to remove any measurement bias, but due to constraints of time and availability of clinical records this was not possible. Such measurement bias would have tended to direct estimates of effect to the null.

Survival analysis gives a broad picture of the average survival probability for a patient group. As the characteristics of individual patients may vary from the average it does not describe individual patients. Experts in prognostic study methodology have emphasized that prognostic models may only explain some of the variation in survival between

patients, and have advised that they be used as guidance to prognosis in individual patients in the context of the overall clinical picture that is being considered by a clinician[192].

Study 2

Study subjects were not consecutive patients attending in the clinical setting but were recruited to a treatment trial. This may have led to some selection bias.

The choices of clinical event are the same as in Study 1 although incident diagnosis of oesophageal varices was included. This may introduce bias as endoscopy practice was not standardised and different clinicians in different centres may have varying practice in which patient is endoscoped, and how often. This is likely to be non-differential as clinicians were blind to ELF. Also practice may vary in what is reported, e.g. the size of varies considered to be significant as smaller varices may not be reported systematically.

This was not an inception cohort leading to limitations outlined above. There was loss to follow up which leads to selection bias, and loss of power leading to less precise results. However patients with PBC are recognised to be a more compliant population with respect to clinic attendance and therapy than other aetiology of CLD and overall the length of follow up was good in this cohort.

ELF predicts outcomes

The results of both of these prognostic studies showed that the enhanced liver fibrosis panel of biomarkers established after external validation studies (Chapter 4) can predict those people who would progress to clinical outcomes over the next 8 years. Crude unadjusted analyses by Kaplan Meier plots showed that tertiles of baseline ELF score can predict liver outcomes and all-cause mortality, with those people having ELF scores in the highest tertile at baseline being significantly more likely to have clinical outcomes than those in the middle tertile, who in turn were more likely to have outcomes than those with the lowest tertile score. Histology was also predictive of outcome when classed as mild, moderate or severe fibrosis, although less reliable when individual stages of fibrosis were analysed.

The ability of ELF to predict all-cause mortality was greater than biopsy. The hazard ratio for a unit increase in ELF remained high (1.60) when adjusted for biopsy suggesting that it has something to offer in addition to knowing the histology. This is in contrast to the Ngo study where the AUC performance of both Fibrotest and biopsy was less in predicting all-cause mortality than liver related outcomes. The reasons for this are unclear. It could be

that ELF which includes biomarkers involved in extracellular matrix formation and breakdown is reflecting additional morbidity in the body that contributes to the mortality rates[187].

Attempts using logistic multivariable regression analyses to derive a more effective panel of markers to predict clinical outcomes using this dataset were unfruitful. Such panels did not add any significant improvement to the performance of ELF in predicting clinical outcomes.

Role in clinical management

The ability of ELF to predict clinical outcome may be a very useful additional tool in the management of patients with CLD. It allows patients to be identified who are more likely to suffer serious clinical outcomes within 6 years with a high degree of accuracy. Patients could then be offered surveillance for oesophageal varices and HCC, and could be better prepared for future transplantation. There is published evidence that such surveillance is associated with a reduction in morbidity and mortality making ELF useful to the clinician and patient in the identification of serious asymptomatic liver disease (see Chapter 1.2). This use of ELF is also likely to improve the cost effectiveness of management of patients, who often have protracted and frequent hospital admissions with clinical complications of decompensated cirrhosis. This needs further research to provide robust evidence to confirm such a hypothesis.

ELF is not a replacement for biopsy. Biopsy offers opportunities to diagnose co-existent causes of CLD which may alter patient management, and provides direct visualisation of any distortion of liver architecture which may provide clues to additional co-pathology. In this chapter ELF has been shown to be effective in the provision of information with which to predict future clinical outcomes, and potentially it may have a place in the more regular monitoring of disease progression/regression where the role of biopsy is restricted. It may also be used when patient preference is for a non-invasive method of evaluation of prognosis and where biopsy is difficult to perform –such as in prison.

Due to a high NPV (97%) ELF can also accurately exclude those patients who are unlikely to have clinical events in the next 6 years. It may help clinicians in primary and secondary care to decide the setting for optimal follow-up of patients. ELF may also be used in the evaluation of therapy in patients with CLD, both in tackling the underlying causes such as

viral hepatitades, and development of anti-fibrotic drugs. Care needs to be taken in extrapolation to primary care as currently no studies have evaluated performance of ELF in this setting.

There are public health applications in the elaboration of the epidemiology of CLD. With the burgeoning epidemics of hazardous drinking, obesity and hepatitis C it is imperative that the prevalence of CLD especially serious fibrosis/cirrhosis in the general population/primary care settings is determined. ELF is a tool to explore the epidemiology in this silent population allowing the optimization of management of people with CLD and offering opportunities for modification of lifestyle and effective therapeutic interventions. This aspect is explored further in the overall discussion (Chapter 6).

Further research is needed to evaluate the ability of ELF to predict clinical outcomes in independent populations, especially those with a higher proportion of ALD and NAFLD as the importance of these liver diseases become ever more important in the light of rising obesity and harmful drinking.

5.8 Conclusion

ELF score can independently predict liver related clinical outcomes in patients with chronic liver disease, at least as well as liver biopsy, and is better than biopsy at predicting all-cause mortality. ELF performs better than tools currently used for prediction of outcome in patients with PBC. It is likely to be a useful prognostic tool in clinical practice.

CHAPTER 6
DISCUSSION

6.1 Statement of main findings

This thesis set out to assess the role of non-invasive biomarkers in chronic liver disease, and to evaluate the diagnostic and prognostic performance of one particular panel of serum markers-ELF. Chronic liver disease (CLD) is an important cause of death in middle-age in the UK, with increasing mortality and morbidity over the past 3 decades. This trend is likely to continue with rising prevalence of major risk factors such as harmful alcohol consumption and obesity. The asymptomatic nature of CLD for most of its natural history means that that there are no accurate estimates for the prevalence of CLD and there are few opportunities to identify those people at risk of developing serious fibrosis and cirrhosis in general and high risk populations. Non-invasive biomarkers with high performance could provide a better tool than those currently available, to identify accurately those people with risk factors for CLD who are at risk of progression to serious liver disease. They may also be able to add considerably to the management of people with CLD, contribute to understanding natural history and epidemiology, and aid the assessment of anti-fibrotic and anti viral therapeutics.

The systematic reviews in chapter 3 showed the current scope and performance of available serum markers explored so far in the diagnosis of liver fibrosis severity. The markers have been most robustly evaluated in CHC, and least in ALD. The performance of all markers is better at the severe fibrosis/cirrhosis end of the spectrum than in identification of no or mild fibrosis, and panels of markers perform better than single markers. The performance of some panels of markers, including the ELF panel, seems to be good at identification of significant fibrosis in reasonable quality studies, with AUC values of more than 0.80. Eight external validation studies of the ELF panel in cohorts of patients with CHC, NAFLD, PBC and HCV-HIV co-infection presented in Chapter 4 showed that ELF was able to identify moderate/severe and severe fibrosis defined by biopsy with a high NPV and good sensitivity and specificity. Access to these external validation cohorts allowed the whole of the original ELF cohort to be utilised to simplify the original published ELF panel through the removal of age whilst maintaining its diagnostic performance.

Liver biopsy acts as a surrogate for the prediction of clinical outcomes, in that those patients deemed to have more severe stages of fibrosis are more likely to progress to clinical outcomes. Histology obtained from biopsy is used as a reference standard to determine the accuracy of new tests to assess fibrosis and therefore clinical outcomes. Biopsy is an imperfect reference test and may lead to misrepresentation of index tests such as the ELF panel. Direct evaluation of the ability of non-invasive markers to predict clinical outcomes may offer a method of reducing such inaccuracies.

Having established the ability of ELF to identify fibrosis on biopsy, the ability of the panel to directly predict clinical outcome was studied and results were presented in chapter 5. The ELF panel was shown to be very good at predicting liver related outcomes at 6 years survival (AUC 0.87) and all-cause mortality (AUC 0.82) in patients with mixed aetiology CLD (n=498) and in a smaller (n=161) cohort of patients with PBC (AUC 0.79). Survival analysis showed that tertiles of baseline ELF were able to identify those patients who would go on to have clinical outcomes, with those in the lowest tertile (ELF score range 4.16 to 8.33) having significantly greatest event free survival time, and those in middle tertile (ELF score range 8.34 to 12.51) significantly longer event free survival than those in the highest tertile (ELF score range 12.52 to 16.67). From logistic regression at 6 years there was a doubling of the odds of a liver related clinical event with each unit change in ELF score. The predictive performance of ELF was confirmed in a long term follow up of a cohort of patients with PBC presented in Chapter 5 where ELF had the best performance profile compared to other prognostic biomarkers.
The results of these prognostic studies strongly suggest that ELF is able to predict clinical outcome, and may offer a valuable tool in patient management.

6.2 Strengths and limitations

The research presented in this thesis has strengths and limitations which are important to discuss in order to facilitate robust distillation of the results, and to identify and plan future research.

Strengths

This thesis has addressed an important topic (use of diagnostic tests to identify liver fibrosis and predict clinical outcomes) in a disease (chronic liver disease) with rising prevalence and urgent public health implications, with obesity, harmful alcohol consumption and injecting drug use representing major concerns in the UK. A wide range

of statistical analyses have been used, including summative analyses and development of clinical utility models which attempt to place such tests in a clinical context.
Three novel systematic reviews of the published literature were conducted. These reviews have added to the available knowledge on serum markers in CLD. The SROC of diagnostic tests in this field was also novel and offered summary measures of diagnostic performance in biomarkers in CHC.

Collaboration with many national and international centres permitted external validation of ELF panel in a large number (n=941) of patients which confirmed that it is able to identify significant fibrosis in many different types of liver disease with a good diagnostic performance. This adds considerably to the knowledge of ELF and permits its clinical use to be elaborated for direct patient benefit. In addition, these validation studies make ELF one of the most externally validated panels in the public domain. The extensive validation work also begins to allow ELF to be used as a reference standard/proxy outcome for further research in place of biopsy where this is not feasible, to use in large studies, to improve recruitment, to offer potentially more cost effectiveness to the NHS and reduce patient morbidity/mortality. The collaborations with centres and external validation allowed the use of the entire cohort in which ELF was derived to conduct analysis that lead to the simplification of ELF without reduction in performance.

In this thesis much has been written of the limitations of liver biopsy as a reference standard and the need to use clinical outcomes as outcome measures in the evaluation of tests in predicting future morbidity and mortality. The prognostic studies had large sample sizes and provided robust data on two cohorts of patients on the ability of serum markers to predict clinical outcomes. Currently this has a very small body of published evidence and research conducted and presented in this thesis represents a considerable addition to this. In England the ability to track all mortality and the excellent response from General Practitioners in determining morbidity in those patients who had defaulted or discharged from the recruiting centre meant that the information collected and used in analyses was 92% complete. This is an excellent follow up rate rarely achieved in cohort studies. Willing cooperation from all participating centres and local help in getting clinical records meant that in most centres data were able to be extracted direct from clinical records for most of the 498 people followed up. For cohort studies loss to follow-up and poor access to records compromise analysis and in this study both of these were largely avoided. The

PBC cohort offered an opportunity to confirm the predictive ability of ELF in a single aetiology cohort, and also to compare performance with clinical scales in current use as prognostic indictors. Data collection and analysis were performed using methodology designed to reduce bias. For example, the liver related outcomes were ascertained independent of ELF and the analyses controlled for confounders. The relative risks derived were large, making chance and bias unlikely as alternative explanations. The overall result derived from this study, demonstrating that ELF can predict clinical outcomes will be of benefit in management of CLD to clinician and patient.

Limitations

The systematic reviews were focused in their objective of identification of liver fibrosis severity. They did not address the ability of serum markers to identity inflammation, in particular alcoholic hepatitis and necro-inflammation and steatosis in CHC and NAFLD. Inflammation and fibrosis are on a disease continuum with many common pathological processes and it is likely that common markers will be able to identify both histological conditions. There is evidence also that necro-inflammation on the index biopsy predicts fibrosis reinforcing this premise. There have been studies that have looked at serum markers ability to predict inflammation and fibrosis.

Only serum markers were evaluated and not all modalities such as liver elasticity (Fibroscan) and hepatic vein transit time (microbubble ultrasound). These tests do have identified limitations such as reproducibility, and technical difficulties in execution in obese patients. However in the last few years evaluations of these tests and comparison with certain serum markers have shown promise. In particular, using them in combination with serum markers showed good diagnostic performance (AUC 0.88 for identification moderate /severe fibrosis)[191]. Development of such combinations and further validations are a promising area for future research.

Literature searches of prognostic studies are more likely to miss studies than searches for RCTs[192], and the search conducted for this thesis may have missed publications. However it did identify studies in different liver diseases, at different calendar times, conducted in small and larger populations.

Transplanted patients were included in the cohort (n=42) who may have behaved differently with respect to disease processes and development of clinical outcomes to

people with birth livers. In sensitivity analyses conducted by removing these patients (see Chapter 5), trends in the results were the same as those when they were included, with adjusted OR from logistic regression and ROC at 6 years being the same. Adjusted HR from Cox Proportional Hazard analyses differed slightly but the direction of results was the same.

The ability of ELF to predict clinical outcomes was conducted in a mixed cohort of patients with CLD. Individual aetiology groups were too small to carry out analysis in each separate aetiology, and it may be that ELF will have a differing performance at different thresholds in each of the different aetiologies of CLD. This was partly explored with the sensitivity analysis where those patients with ALD were removed, and the study in 161 patients with PBC where ELF had similar performance in prediction of outcomes.

Data was extracted from the clinical records by more than one observer for pragmatic operational reasons, and no record was double extracted. This may have led to measurement bias but this was reduced by use of a standard data collection form, and commonality of training for data collection. In addition any bias would have tended to direct results to the null.

Whilst follow up of those patients who had defaulted or discharged form the recruiting centre was excellent in England, this was not the case in the continental European centres. In only one centre (Sweden) was it possible to interrogate the national central records for mortality. In all others the outcomes in these patients could not be ascertained. Efforts were made to contact last recorded Primary Care physician in one centre but the results from this were sporadic and not systematic and many GPs contacted no longer had the patient under their care. Defaulters and discharge rates are high in this patient group making data from these centres incomplete and not robust. Sensitivity analyses combining English and continental European data showed that the trends in predictive performance of ELF were the same as when the English data were analysed independently with little change in AUC (although more precise estimates were possible with narrower 95% confidence intervals). This may indicate that the losses to follow up were non-differential-i.e. the associations between serum markers and outcome were similar in those followed up and those lost to follow up. These data are presented in Appendix 6.

The mean follow-up time was 6.5 years in English centres and this may not be long enough for those patients with mild disease to develop clinical outcomes. Repeat data collection for clinical outcomes in the future may provide these data.

Some of these limitations are consequent on follow-up of a cohort not originally established for prospective study. In addition there have been changes in the research ethical environment in the years since the original study was conceived. In the past information regarding mortality could be accessed in England via flagging with the Office of National Statistics, but such data now requires specific patient consent, and tracking for morbidity requires that this is set up at the start of the study and consent obtained in advance for direct contact with patients. It seems logical that when studies are conceived and set up consideration should be given to the potential advantage of long term follow-up so that appropriate consent can be obtained.

In Europe the patterns and processes of patient referral often differ from those operating in England, in that patients can be referred to any centre regardless of place of residence. This meant that many patients were referred for assessment and treatment of CHC and then discharged back to local healthcare facilities which could be hundreds of miles away. It was a minority of patients that continued to receive care for CLD in the recruiting centre. Healthcare systems also differed in that patients could be referred to more than one service within the same hospital for the same condition at different times, with little collaboration/coordination or cross referral resulting in different sets of notes raised, stored in individual departments/offices, different protocols of care, and different clinical disciplines involved with clinical care. Such problems of dispersal of patient care may be a problem in multicentre studies as it tends to be large secondary /tertiary referral centres that express most interest in participating and they may have responsibility for patient care for a certain time and then the patient is cared for more locally (although transplant centres tend to share care for many years). Also trends in the last 5 years have been to increase skills and technology in smaller more local hospitals for the management of patients with CLD. For example, one continental European centre at the time of recruitment to the original study was the major liver centre in the region. Over the intervening 5 years it trained many hepatologists who were then able to set up satellite units in the region capable of managing all but the most complex patients. All of these factors contributed to

the challenge of ascertaining the clinical outcomes in those patients defaulted/discharged from the recruiting centres in Continental Europe.
Finally there is always the problem of loss of clinical records; accidental (fire and flood), within the vagaries of personal/hospital filing systems, or routine culling. Different hospital protocols exist for destruction of paper notes and the processes for microfilming clinical notes- which data are conserved, how well this is done and where these are stored. This posed some limitations on the collection of data on clinical outcomes from clinical records.

One central laboratory assayed the nine serum markers so that consistency was maintained for these tests. The simple markers (ALT, AST, bilirubin, platelets etc.) were all assayed locally in different laboratories at each of the centres. This may limit the use of these simple markers, in that different assays may have been used, different upper limits of normal selected, and measurement performed using different units. However, harmonisation of units of measurement was possible using standard conversion matrices. Different centres tended to order different tests, for example, not all centres routinely measured AST and the clotting parameters differed between centres (prothrombin time/INR). This reduced the number of patients available for analyses, for example constructing new models and using standard tests to perform comparison with published panels of markers (APRI) and clinical scores (MELD, Mayo) to compare with ELF. There was only a baseline value of ELF for most subjects so the ability to evaluate the predictive performance of a change in ELF was limited.

Even when mortality was ascertained in England the information on the death certificate may have been inaccurate. The definition for liver-related death that was used was any mention of liver disease in Part I. The cause of death on the certificate in some cases was very vague, and in very similar clinical cases the information detailed in Parts I and II in one case, were reported as Part 2 and Part 1 in others. In a few cases supplemental data from clinical records allowed clarification (for example, gastrointestinal haemorrhage entered Part I of the certificate when bleeding varices were detailed as the terminal events in the clinical records), but in many cases such elaboration was not possible (for example where the information entered was "cause of death unknown", "bronchopneumonia" with no other details). This may have led to an under-estimation of liver related deaths. The inaccuracies of death certification are common problems encountered in studies of

mortality, with the potential for bias in studies relying on certificates highlighted by other authors[193;194]. In a study in a teaching hospital in Canada, major errors were found in 33% of death certification[195]. Such errors may be particularly important in the case of CLD, where the underlying causes are associated with social stigma including alcohol consumption, injecting drug use, and even obesity. Studies have reported that 40% of people with biopsy proven ALD had no mention of alcohol on the death certificate[196]. The methodology of such studies has been criticised in a recent review citing problems of only taking main cause of death and not taking into account underlying conditions, and for the lack of transparency in reporting methods[197].

Bias may have occurred in the prognostic studies, in particular partial verification bias and spectrum bias. Partial verification bias occurs when patients with negative test results are not evaluated with the gold standard test. This may be relevant to this study in that patients were recruited at biopsy and their pathways into this test were non-standardized. For many patients the presence of an initial positive test such as an abnormal liver function test would lead to biopsy leaving those with negative tests not biopsied. To avoid this, a study should include consecutive patients at risk for a particular disease, and not only a subset who underwent definitive testing with the reference and index tests. This bias may have been reduced in that consecutive patients were recruited at the point of biopsy.

A frequently ignored problem is that of spectrum effect/bias. This is the phenomenon of the sensitivity and/or specificity of a test varying with characteristics of the patient sample. It occurs when the disease-test relationship is heterogeneous across patient subgroups (for example differences in test performance estimates by sex or age or cause of CLD), and the study draws preferentially from a limited portion of the patient spectrum[198]. There is also an impact of disease severity distribution on sensitivity and specificity, and of conditions that appear similar to the index disease on specificity. For example, in the hospital setting the prevalence of severe disease is higher than in a population in primary care, which in turn is likely to have a higher proportion of milder cases than a population recruited in the hospital setting. Yet both may have the same prevalence of the disease. Assuming that the prevalence of disease is constant in all population samples studied, if a test is better at identifying severe disease, the sensitivity will be less in primary care. The distribution of the results for such a test when applied to subjects with the condition in primary care is likely to be skewed towards the milder end of the disease severity spectrum and further

away from the global diseased population mean value for the test, resulting in more primary care diseased subjects with false negative results. If some of the healthy subjects share some of the same characteristics as those who are diseased (such as symptomatology), then the false positive rate is likely to increase and specificity will be lower. Thus the sensitivity and specificity of a test vary not with the disease prevalence but are related to the distribution of disease severity in the affected people or those without disease who have similar symptoms to those in the diseased population. The term bias implies that there has been some systematic error in the study design that invalidates the study results. Variations in test accuracy between subgroups as described above may actually be true variations and so the term "effect" may be more appropriate[199]. The primary problem resulting from spectrum effect is generalisability. If the test performance varies substantially with sex there is little clinical utility in using the test performance estimates in a mixed sex population. Strategies to identify and deal with spectrum effect include simple stratification procedures, which are limited by sample size and precision of the estimates.

The original recruitment to the ELF diagnostic cohort and the PBC cohort trials had strict inclusion and exclusion criteria and participants were not consecutive attendees in hepatology clinical settings. The exclusion criteria for the ELF study included those patients with any known extra-hepatic fibrosis including rheumatic and lung disease, significant cardiovascular disease and cancer, advanced cirrhosis with evidence of decompensation (Child-Pugh Class C), drug induced liver disease or hepatocellular carcinoma. For the PBC trial inclusion criteria included asymptomatic patients having more Stage 1 on biopsy, exclusion criteria included advanced disease, cancer patients, and those patients taking immunosuppressive drugs. This leads to a reduction in generalisability.

Where the reference standard itself is subject to error, (for example the liver biopsy), then non-differential misclassification is possible leading to underestimation of the test performance.

The case definitions of the clinical outcomes employed in the prognostic study were established at the beginning of the study and attempts were made to make them as objective as possible. Bleeding varices had to be diagnosed with endoscopy; jaundice by clinical assessment and raised bilirubin, ascites by peritoneal tap/ultrasound/clinical

assessment and encephalopathy by standard measurement scale/clinical assessment. The use of clinical assessment in some cases rather than formal tests may lead to non-differential misclassification of the outcome, which would tend to bias to null. However in these cases the clinical picture as a whole was considered by the data collectors who were clinically experienced, including other test results to reduce the impact of this limitation. Presence of varices on endoscopy was not included as an outcome in Study 1, as this may have been biased as it is dependent on endoscopy practice in addition to the true prevalence of varices.

There were several overarching methodological issues that arose from the work presented in this thesis and these, along with the place of non-invasive markers in CLD, are discussed in the sections below.

6.3 Reference test for liver fibrosis in CLD

The determination of a robust reference test for liver fibrosis in CLD against which to evaluate new diagnostic tests is problematic, as is the elaboration of which end-point of disease should be identified. The current reference standard is histology derived from liver biopsy. A reference standard should correctly reflect the true disease status but it is rare that reference standards are error free. Biopsy is no exception and liable to inaccuracy arising from both failure to acquire a truly representative specimen of liver, and in the interpretation of the histology. This may lead to misclassification of the index test with sensitivity and specificity biased in either direction depending on the error attributable to the biopsy. If the errors in the reference test and index tests are not correlated then the misclassification of the reference standard tends to underestimate test accuracy. The degree of underestimation is non-linear. Sensitivity is underestimated most when the prevalence of the target condition is low. The estimated sensitivity of the test will be closer to the true sensitivity with increasing prevalence. Specificity is underestimated most when prevalence is high. Estimated specificity of the test will be closer to the true specificity when the prevalence of target condition is low. Odds ratios are underestimated most when prevalence is at either extreme and tend to move towards the null. If errors in the reference standard are correlated with test errors the effects are more difficult to predict.

One methodological approach to the lack of a robust reference standard has been the use of latent class analysis, where a mathematical model is constructed and a variable is assumed

to represent the true disease status. The model is then used to estimate error for different index tests. However assumptions have to be made about the inter-dependence of the errors from each test and as these cannot be measured the method has limitations. It does not have real-world relevance in a clinical setting as it assigns a mathematical function to represent "disease" rather than a clinical case definition[200;201 202] Another approach is discrepant resolution. This is used in cases where the results of index and reference tests are different and a further test (the "resolver") is used to ascertain which is correct. As only those discrepant cases are subjected to the resolver test it biases towards improving the apparent accuracy of the index test[203]. In addition there may be errors in the resolver test which, if similar to those in the index test, may again overestimate the accuracy of the index test. When the index test is better than the reference standard the false test results may in fact be true. The index test may eventually be shown to have greater predictive ability by identification of groups with different outcomes, establishing over time, the index test as the new reference standard.

A recent review of methods of evaluation of diagnostic tests when there is no "gold standard", suggested methods of adjustment when the reference standard is imperfect. These included adjusting accuracy estimates using information on the degree of imperfection of the reference test and the correlation of the errors between the index and reference tests, or sensitivity analysis to examine the impact of the imperfect standard. However their main conclusions were that when the reference standard accuracy was very poor, or where the magnitude of the imperfection is not well documented, the concept of clinical test validation may be helpful. This is when clinicians and scientists examine, using a number of different methods, whether the results of an index test are meaningful in practice. This takes time and relies on a consensus in academic and clinical settings on thresholds, and a point in the validation process where the data gathered would be sufficient to allow clinical use of the test with confidence[204]. This latter suggestion is promising and the research findings in this thesis contribute to the evidence base of this approach. More studies using this methodology are required.

In addition to methodological flaws as a reference test, biopsy is an invasive test with morbidity (3%) and mortality (0.03%), and may lead to only those with a likelihood of significant disease (identified with prior testing with, for example with abnormal traditional liver function tests) and not all patients seen by the clinician being biopsied.

This will result in partial verification bias which tends to overestimate sensitivity and underestimate specificity.

How well the biopsy actually reflects the true fibrosis in the liver is limited by sampling error (a very small part of liver is assessed, and fibrosis may not be homogenously distributed in the liver), intra and inter-observer variability in interpretation of histology, and by its nature the biopsy yields histology which gives a single snapshot in time of the architecture of the liver.
Use of ordinal categorical classifications for liver histology artificially forces a biological process into a rigid structure which does not reflect the dynamic processes involved in fibrosis, nor accurately describes the non-linear nature of fibrosis (stage 2 is not twice as much fibrosis as stage 1)[43;205]. The quantity of fibrosis itself may not be the only factor in determining the staging of fibrosis as the location, distortion of liver architecture and vascular involvement also contribute to which stage a pathologist may allocate the biopsy. This is supported by morphometry studies which show that there is only a 12% increase in the quantity of fibrosis between Stages 0-4. It is likely that the vascular disruption in the later stages of liver disease where the portal vein and central vein are connected by fibrosis tissue leading to shunting of blood flow in the liver, may have a major role in the loss of liver function and thus the prognosis.

The quality of liver biopsy (length of sample and the number of portal tracts) has been shown to be important in improving the biopsy as a reference standard, with those with 100mm having a 30% coefficient of variation compared to 55mm which had a 55% coefficient of variation[60]. Classification can be improved by using biopsy lengths of 25mm (75% correct classification). This has practical considerations in that current practice of obtaining biopsies under ultrasound guidance and using narrow gauge needles, such lengths are rarely obtained. Some authors have suggested the use of a discordant analysis. Where biopsy and serum marker disagree (discordant pairs) and the biopsy is less than 10mm long the case is classified as a biopsy failure (false positive rather than true positive). As in discordant analysis using a resolver test this variation has methodological problems. These include evidence that more severe disease may result in shorter more fragmented biopsies resulting in selection bias so that those with severe disease are classed as false positive reducing the prevalence of severe disease[206]. However it cannot be assumed that the score assigned to a short biopsy is always inaccurate. If both biopsy and

serum markers agree (concordant pairs) and the biopsy is less than 10mm should both be classed as inaccurate? Such discordant analysis is therefore flawed.

The biopsy remains a valuable test for fibrosis despite these imperfections as a reference standard, the limitations of acquisition and practical issues of its invasive nature. It can provide information on the structure and architecture of the liver, suggest the possible aetiology in those with unknown cause of CLD or identify concomitant aetiologies which may otherwise remain undetected, and supply an indication of the length of duration of the injury[207]. There is evidence that severity of fibrosis on biopsy can independently predict liver-related mortality in a cohort of 3,000 patients with CHC[208]. However the limitations of acquisition, difficulties in interpretation and repeatability of biopsy still remain.

The use of other tests as a reference standard for liver fibrosis has been suggested. These include automated morphometry, which uses computer programmes and digital analysers to quantify the amount of fibrosis in a biopsy. Problems remain in acquisition of the sample, and intra-sample variations between readings[209;210].

6.4 Role of biomarkers

Non-invasive biomarkers are a candidate for evaluating liver fibrosis rather than biopsy as they can be repeated easily, have a low coefficient of variance, have biological plausibility, are continuous measures reflecting the biological processes involved in fibrosis, do not require senior professional experts to acquire, and are likely to be cheaper.
They are measures of the structure or function of the liver in liver fibrosis. Albumin (manufactured by the liver), platelets and ALT flag up the failing hepatocytes (function). Other tests such as HA, P3NP, TIMP, MMP are considered to be direct markers of extracellular matrix deposition (structure). However they may be a reflection of both structure (ECM is laid down and degraded in a dynamic process leading to structural alteration of liver architecture) and function (as the vascular structure is disrupted there may be reduced resorption of the markers due to intra-hepatic shunting. This may explain why ELF panel is very good at identifying cirrhosis with most people producing a lot of matrix (active disease), and some people having a balance between depositing and degrading ECM (still high matrix production but "inactive" disease), with decreased reabsorption of components of ELF. As such biomarkers are measured from the serum it is possible that they are also measuring other pathophysiological functions of the body. The

most obvious confounding process would be fibrosis outside of the liver. The original ELF cohort excluded patients with known fibrotic conditions and so it is likely that the liver was the predominant site of fibrosis. However the external validation cohorts comprised of patients less carefully selected with respect to fibrosis and this was not an exclusion criterion in any of them. ELF performed as well if not better in the external validation as in the original study in identifying significant fibrosis on biopsy, so it may be that concern over extra-hepatic fibrosis is relatively unimportant. Studies directed at this specific question are required.

There have been reports in the literature of the effect of current heavy alcohol consumption on circulating serum markers which may limit their performance in identifying the chronic effect of alcohol on fibrosis in patients who may be current drinkers. The mode of action of alcohol on the markers is unclear but studies have shown that some are more susceptible to influences of acute consumption (tenascin, laminin), some are unaffected (PIIINP, TIMP1), and some very variable (HA). Some small studies have explored the diurnal variation in hyaluronic acid levels. In one study, HA measurements taken in a group of healthy people and a group with rheumatoid arthritis upon rising from bed in the morning were greater than later in the day. The authors suggested three to four hours after rising might be a suitable standard time to assess levels. Patients with PBC and a group of hospitalised patients did not demonstrate this same diurnal pattern. The suggestion for standardised times of phlebotomy was supported by a small study in patients with osteoarthritis in USA[211-213]. This variation in serum markers needs further research by assessment of levels at different times of the day and on different days.

The maintenance of performance between training, internal and external validation populations is unusual as performance of serum markers panels in external independent populations is expected to be lower than that in the population in which the panel was derived. It is possible however that ELF is measuring another function in the body. This is hinted at in the fact that ELF is a good predictor of all-cause mortality at 6 years (AUC is 0.82 (95% CI 0.75,0.88) whereas biopsy is not (AUC is 0.70 (95% CI) 0.62 0.79). Evidence is beginning to appear that in patients with NAFLD there are cardiovascular changes (such as carotid artery intimal thickening) which may link NAFLD via insulin resistance/metabolic syndrome to extra-hepatic disease[214]. A hypothesis could be that ELF

components are a surrogate indicator of these processes which may lead to non-liver related death.

The use of serum markers as prognostic tests to directly predict clinical outcome rather than using biopsy to predict fibrosis severity marks a step forward in the management of patients with CLD. The limitations of biopsy as reference standard become less important as it is substituted with clinical outcomes which are much more easily measured.

Highly accurate identification of patients who are likely to suffer a liver related outcome within 6 years offers opportunities to streamline care for patients at greatest risk. This will allow early instigation of therapy for oesophageal varices (an intervention of established benefit), screening for HCC and better preparation of suitable candidates for transplantation.

6.5 Performance indicators in diagnostic tests

Summary measures of diagnostic accuracy are reported in different ways in the published literature. The inclusion criteria for the systematic reviews presented in chapter 3 used studies that reported test results in terms of sensitivity and specificity, predictive values or DOR or LR which can be derived from a 2x2 table with participants classified as True Positive, False Positive, False Negative, True Negative. This excluded some of the poorer quality studies found in the systematic reviews searching which reported the means of tests in patients with different fibrosis severity but no other performance parameters. Further summary measures are ROC curves which present sensitivity and specificity of continuous variables at different thresholds. However, the day-to-day use and interpretation of AUC in the clinical setting is difficult and whilst AUC offer a means of comparing two or more tests, the clinical meaning of small differences are not well understood. The use of likelihood ratios is more clinically powerful in reporting test accuracy for dichotomous outcomes than sensitivity and specificity and can be used to convert pre-test probability to post-test probability of disease in the clinical setting. Reliance on parameters other than LR may tend to over estimate the performance of a test. Whilst these parameters are considered by methodological experts to be the preferred performance indicators for diagnostic tests, the reporting of such measures is inconsistent in the literature. A systematic review of accuracy in diagnostic literature suggested that there was a need for

consensus to support change as such poor reporting may result in readers of diagnostic literature misinterpreting test accuracy[215].

How clinicians make diagnoses is complex, with many constraints and drivers operating simultaneously and not all of them are robust, scientific considerations. In a survey conducted with 300 practicing physicians in USA, fewer than 25% considered sensitivity or specificity before ordering tests and LRs were almost never used. The main reasons were that the measures of accuracy were not available in the literature, not easily accessible within a short time and unless instantly available to the doctor they were unlikely to be used. In addition formal training in understanding diagnostic test evaluation and accuracy of common tests was lacking or inadequate, and often when the diagnostic accuracies were located, the published indices were established in a population that was dissimilar from that of the doctors[216] . Other factors impact on the decision to use a test in clinical practice such as cost, the anticipated risks and benefits of the test both to patient and physician, physician legal liability, and patient preferences [217] . The physicians reported that they empirically used informal "direct" methods for evaluating test performance in their population; that is they used a test after eliciting history, symptoms and signs (arriving at a pre-test probability) and then evaluated who did and did not have the disease. This obviously can lead to bias, including partial verification bias and review bias (subjective interpretations may be biased by prior knowledge and can occur when a new test is used and the disease status from the reference standard is known). In general, in order to determine the performance, value and place of a test in clinical practice there is a need to have sufficient numbers of each condition to obtain reasonably precise estimates of diagnostic performance, a need for a "gold standard" test and accurate recall of diagnostic accuracies in various diseases and sub-groups of patients all of which may be lacking.

Other attributes of a test that may have clinical applicability in the management of individual patients are the interpretation of positive and negative results of a test (the predictive values). These are the most intuitive attributes to use in a clinical setting with individual patients. However, the published predictive values of any test are specific to that clinical setting as predictive values vary with prevalence of background disease. Generally whilst AUCs and DORs are often used to compare tests in the literature they are more difficult to translate to the clinical setting. Clinical utility models, as shown in chapter 3 with CHC and NAFLD, use theoretical cohorts of 1,000 people and can be used to show

the error rates and the proportion of people for whom a test result is not attainable at robust levels of sensitivity/specificity/predictive values. These models can permit informed choices about the use of tests in individual clinical settings. The values that correspond to acceptable error rates for patients and clinicians have not been ascertained in formal studies, and tolerance of false positive/negative rates may vary with clinical setting and between individual clinicians, and according to the clinical consequence of false positives and false negative results. Where testing is important for identification of eligible people for treatment, clinicians may employ lower false negative levels, and for surveillance in severe fibrosis/cirrhosis they may have a reduced tolerance of false positives. Selection of threshold for a test is therefore difficult. Is it more important to minimise false positives or false negatives? Clinically if the cost of missing a diagnosis is great, and treatment (even inappropriate treatment of a healthy person) is safe, the use of a threshold which has a high true positive rate and most of those people with disease will be treated appropriately at the cost of treating the people who are false positives. Conversely, if the risks of therapy are grave, and therapy has a relatively low efficacy, a threshold could be taken where true positives may be missed but many unaffected people will not be harmed (low false positives).

Communication of test results to patients is not always easy, and explanation of the likelihood of disease requires knowledge of test attributes and skills in interpretation for an individual patient and specific clinical situation. Patient preferences for risk-taking and understanding test choice and results need to be explored. Physicians do not tend to use single tests in isolation. Selection of tests tends to be a phased, sequential and hierarchical process beginning with history and a clinical examination. After each test the results are converted from a pre-test probability, often implicitly, into probability of disease which can then direct the selection of further tests or instigation of therapy. Physicians need to know which subsequent tests change the probability and to what extent. Diagnostic studies (as presented in this thesis) tend to focus on the single test and its diagnostic accuracies in a study population which may or not be generalisable to individual clinician practice and they generally do not evaluate the test contribution to estimate probability of disease. The future use of markers of liver fibrosis is likely to be in this context of real-world use of such tests. Such sequential testing with explicit diagnostic accuracy values, is starting to appear in the literature and is discussed below in the implications of the findings reported in this thesis.

6.6 Implications of research findings for clinical practice

The findings of the work conducted in this thesis are of direct relevance to clinical practice and have the potential for incorporation into patient management and in further research. Overall the systematic reviews found that panels of markers perform better than single markers. This may be due to synergy between biomarkers producing a better performance than each of them alone. Markers may be used in combinations that represent different liver processes such as liver synthetic function and fibrosis which may explain this synergy. Other combinations may complement each others biological functions. There are considerations in the use of marker panels in clinical practice which include difficulties in translating differences in performance indicators between single markers and panels (such as AUC values) to what is clinically meaningful, and relative cost of panels which may be more expensive than single markers to the healthcare provider.

The simplification of ELF without loss of diagnostic accuracy is a development of the panel permitting the more seamless automation, and the use of age in analyses. The extensive validation of ELF in many different liver patient groups will allow it to be used in conjunction with other tests including biopsy to identify those patients with any fibrosis, those with moderate/severe fibrosis and those with compensated cirrhosis in up to 64% of patients. The use of an algorithmic test pathway approach to non-invasive markers in clinical practice in order to reduce the number of biopsies has started to appear in the published literature. Sebastiani et al have proposed the use of a panel which uses simple indirect markers of liver function (APRI) as a screening test followed by Fibrotest in APRI non-classified cases for the identification of significant fibrosis, restricting biopsy to patients classified as F0-1 by the non-invasive tests. The number of biopsies was reduced by 50% whilst retaining accuracy of 94% (with Fibrotest alone 43% of biopsies could be avoided) but biopsy could not be avoided altogether. Different algorithmic pathways were put forward for different sub-groups of patients, for example patients with persistently normal ALT and those with compensated cirrhosis. The authors claim cost effectiveness, an increased proportion of patients that can be assigned a diagnosis by the test, and increased accuracy of allocation to fibrosis stage as reasons for this approach[166]. Further studies have expanded this theme using different markers and modalities including Fibrotest and Fibroscan where authors suggested if the tests were in agreement then no

biopsy was needed, if the tests do not agree on the stages of fibrosis then biopsy would be required[162]. These studies were in patients with CHC and there is a need to extend such research into other biomarkers and in different populations such as ALD and NAFLD.

The numbers of people in the UK population who are potentially at risk of liver disease with risk factors such as obesity, harmful drinking and injecting drug use may well overwhelm existing hepatology services. The identification and investigation of people with risk factors, who may be screened using biomarkers (such as ELF) in general practice may be a solution to this epidemic. In those who are found to have a low ELF value significant fibrosis can be excluded (ELF has a high NPV). If they are identified as having any fibrosis, monitoring could be conducted in primary care. Those identified as having significant fibrosis using non-invasive biomarkers could refer to secondary care for therapy and surveillance. Using non-invasive markers in this way is novel and further research investigating this use is needed urgently.

Demonstration that ELF can predict clinical outcomes in patients is very useful for clinicians. It will allow patients to be identified who require closer monitoring and may act as a motivator for lifestyle change. Further research is needed to test such a hypothesis. With the extensive external validation of ELF as predictor of fibrosis on biopsy and of clinical outcomes, the time approaches when it may be considered for use as the directly measured end point for therapeutic trials and epidemiological studies.

6.7 What is the most important end-point in diagnostic test studies in CLD?

Selection of the most critical clinical outcome in chronic liver disease is a matter for debate. In the systematic reviews almost all studies chose to evaluate the performance of serum markers in the identification of "significant fibrosis" (METAVIR F2-4) or cirrhosis (F4). This outcome was chosen in CHC as guidelines for treatment with anti-viral drugs required that moderate/severe fibrosis be present. Identification of patients with this severity of liver disease often in the absence of symptoms was problematic and spurred on the search for a non-invasive biomarker with a good diagnostic performance. However, over the course of this thesis national guidelines have changed in the light of published evidence on the cost-effectiveness of treating patients with mild disease. Future studies of biomarkers in patients with CHC, in common with patients with NAFLD, may thus seek to distinguish those patients who have any fibrosis versus those with no fibrosis. These patients are more likely to progress to cirrhosis and clinical symptoms. The clinical

imperative for patients with NAFLD and ALD is to identify those with high risk behaviours (hazardous drinking or morbid obesity) who have any fibrosis. The numbers are great and rising and it is important to be able to identify those with a greater risk of progression to serious liver fibrosis. This will permit focused interventions to change lifestyle choices and will allow monitoring of CLD. Serum markers with good performance at this lower end of the disease spectrum will be a valuable tool for the clinician.

While successful liver transplantation can extend life expectancy in all cases, limited supply of organs means that the identification of appropriate subjects and optimal timing of transplantation are essential. Use of these serum markers can better identify those at most risk of developing clinical outcomes and better prepare them for transplantation.

6.8 Lessons learned

The cohort of patients that was studied was not established *de novo* for long term follow-up at the initial design stage. When recruiting patients to any study, due consideration may need to be given to possible follow-up and should this potentially yield productive data, then patients should be considered for consent for follow up with respect to flagging for mortality and morbidity using central national records or contact by GP. In addition further contact for possible evaluation with repeat samples should be considered at the outset of the study. However PIAG's interpretation of the legal aspects of data protection and patient follow-up has tended to obstruct such studies.

Multicentre studies where principal investigators change/move to different centres and corporate memory of a study falls away with time, pose difficulties which may be surmountable by having a central repository of study information which is kept up to date. The use of one laboratory for all tests involved in the study would have provided consistency of units and values and covariance which could have helped in the comparison of the simple liver tests in the prognostic study.

Whilst all principal investigators were seen singly during this follow-up of the original study, a group meeting face-to-face with all collaborators and identified staff from each centre who would help in the operationalising of the study, may have contributed to a sense of study ownership, cohesion and direction. It must be recognised that financial constraints prevented this.

6.9 Future research

The research presented in this thesis has shown that biomarkers including ELF panel are very promising as tests which can be used in the clinical management of patients CLD in a hospital setting. They are particularly good at identifying people with serious liver fibrosis which can help to stratify those patients at risk of developing HCC and other symptoms of clinical decompensation. Indeed the work presented has shown directly that ELF can predict those people who are more at risk of clinical outcomes and for whom therapies and harm reduction measures can be directed. As the obesity and harmful drinking prevalence increase and the numbers of people at risk of CLD rises, more identification and triage of those people at risk of serious disease will have to occur in primary care. There are no data in the published literature on performance of biomarkers in this setting where prevalence of serious fibrosis is likely to be less common and the imperative is to identify those with any fibrosis as well as those with significant fibrosis. Future work should include the evaluation of ELF and other biomarkers in primary care.

Although collaborations were extensive and eight external validation studies were conducted to evaluate the performance of ELF panel in the identification of fibrosis in biopsy, each of these cohorts consisted of patients with a single CLD aetiology. It was not possible to combine them as they were heterogeneous with respect to inclusion criteria. More external validation of simplified ELF in mixed aetiology cohorts is needed to further establish the place of ELF in routine clinical practice.

Systematic reviews of non-invasive markers have shown that there is a gap in the evidence base of the evaluation of markers in ALD with poorer quality, smaller, older studies forming the majority of published research. Further robust assessment of biomarkers in large cohorts of patients with ALD is needed as their numbers are increasing. Impact of reporting the presence of fibrosis/deteriorating biomarker values suggesting worsening of liver disease to hazardous /harmful drinkers with respect to modification of drinking behaviours needs to be investigated. Whilst these patients are difficult to recruit and retain in such studies it is vital that such studies are performed.
Similarly assessment of ELF is required in other high risk groups such as patients with Type 2 diabetics/morbid obesity where the Metabolic Syndrome contributes to a high level of NAFLD. Such studies will help to evaluate the prevalence of NAFLD in these patients and determine which are at risk of liver fibrosis and future clinical events.

The research presented in this thesis focused on the performance of serum markers in the identification of fibrosis. It did not evaluate inflammatory activity. This may be important markers for progression of disease in CLD and this work should be extended to include this. The role of extra-hepatic inflammation and links with metabolic syndrome are speculative and require further investigation. Potential studies could include evaluation of the role of ELF in predicting future secondary cardiovascular events in obese patients who have had revascularization procedures; and the relationship between ELF and measures of insulin sensitivity, whole body fat, features of metabolic syndrome and estimates of cardiovascular risk in patients with obesity/Type 2 diabetes.

There are scant data on repeat ELF measures over time and studies utilizing sequential testing of ELF to monitor disease progression/regression as a result of natural history or as a response to intervention would be a valuable addition to understanding how biomarkers can be used in practice.

The work presented has shown that ELF can predict future clinical events in this cohort. This work should be replicated in other populations. For example there is a cohort study of mixed aetiology patients recruited in gastroenterology Offices in three counties in USA which has excellent data on clinical follow up in one of these centres (Kaiser Permanente n~500). This would be an ideal population to further validate the prognostic performance of ELF and new predictive models derived in the original diagnostic cohort.
Lack of robust non-invasive markers of fibrosis has been cited as a major barrier to the clinical trial development of new drugs in hepatology in particular those directed at regression of fibrosis[29] . Use of biomarkers instead of biopsy to monitor progression/regression of liver damage would be invaluable in the evaluation of anti-viral therapeutic agents, anti-obesity drugs and in the development of anti-fibrotic drugs. The results presented in this thesis support the use of ELF in this capacity and studies can be designed to evaluate therapeutics in all of these areas.

Extension of ELF for use in particular single aetiologies of CLD for example NAFLD and CHC is a further area of future research. We have already conducted some work in NAFLD using simple markers in addition to ELF to derive newer models in a cohort of patients which seem to improve the identification of fibrosis of different severity. This has been recently published[150]. Use of ELF with other non-invasive modalities such as

Fibroscan/MRI/microbubble is an exciting area of future research with some preliminary collaborative research being conducted on a small sample of patients with CHC.
All of the research outlined above concentrate on the use of biomarkers and ELF as tests for direct patient benefit. They are also key tools in epidemiological research to investigate CLD at a population level. Access to large birth cohorts with stored sera could allow the prevalence of high ELF values to be ascertained indicating significant fibrosis in a population, and its association with factors such as age, gender, cardiovascular disease, lifestyle factors (smoking and alcohol consumption), and anthropometric factors (BMI, waist circumference). Such datasets do exist in the UK (1958 Birth Cohort and the Health Survey for England) and application has been made to use such data to extend the work in this thesis to evaluate CLD in the UK.

Much of the research in this thesis has been concerned with validity of biomarkers-the degree to which a measurement measures what it purports to measure. However other attributes of a test are important such as reliability (degree of stability shown when a measurement is repeated under identical conditions thus the degree to which the results obtained by a measurement procedure can be replicated). The assay reliability of ELF has been evaluated using standard laboratory procedures and is very good (95% covariance). However the intra-person reliability where ELF is repeated under the same conditions but at a different time of day/on a different day/week is yet to be conducted. The closest time interval available is 0-3 months, and this intra-class correlation coefficient is reported in chapter 5 and is within acceptable limits.

Direct comparison between biomarkers is beginning to be performed but often on small numbers and more often in CHC. Most comparisons exist between those panels that use readily available simple blood tests (often direct liver function markers) such as APRI or Forns, and Fibrotest some which incorporated direct markers such as HA[218-222] (See Appendix 4 for table of comparison of these markers). ELF has not yet been directly compared to these panels (except in n~240 as reported in Chapter 5). It would be ideal for the research community if a reference population could be established that could be used to compare biomarkers under the same conditions and in the same subjects. Recent recommendations from the Cochrane Diagnostic Test Accuracy Working Group included support for large multi-centre studies with standardised recruitment and diagnostic test

pathways and analysis, using individual patient data rather than combining numerous small heterogeneous studies.

6.10 Overall conclusions

Liver biopsy has practical and methodological limitations as a reference standard for fibrosis staging of patients with liver disease and due to the former, cannot be used to closely monitor patients nor investigate the epidemiology of liver disease in high risk or general populations. There have been major developments in non-invasive biomarkers over the past decade especially in CHC. Amongst these markers the ELF panel using direct markers of ECM production has been shown in extensive validation studies to have good accuracy in predicting fibrosis on biopsy and will have practical use in clinical practice, reducing the need for biopsy. ELF is able to predict serious clinical outcomes at least as well if not better than biopsy, and this will enable clinicians to identify those patients most at risk of complications.

APPENDIX 1

(a) Histopathological Scores of fibrosis

SCORING SYSTEM	STAGE	DESCRIPTION
METAVIR	0	No fibrosis
	1	Stellate enlargement of portal tracts but without septa formation
	2	Enlargements of portal tracts with rare septa formation
	3	Numerous septa without cirrhosis
	4	Cirrhosis
SCHEUER	0	No fibrosis
	1	Enlarged fibrotic portal tracts
	2	Periportal or port-portal septa but intact architecture
	3	Fibrosis with architectural distortion but no obvious cirrhosis
	4	Probable or definite cirrhosis
ISHAK	0	No fibrosis
	1	Fibrous expansion of some portal areas ± short fibrous septa
	2	Fibrous expansion of most portal areas ± short fibrous sep
	3	Fibrous expansion of most portal areas with occasional portal bridging
	4	Fibrous expansion of most portal areas with marked portal-portal + portal-central bridging
	5	Marked bridging portal-portal and/or portal –central with occasional nodules(incomplete cirrhosis)
	6	Established cirrhosis probable or definite
KLEINER	0	No fibrosis
	1	Perisinusoidal or periportal fibrosis
	1a	Mild zone 3 perisinuoidal
	1b	Moderate zone 3 perisinuoidal
	1c	Portal/periportal
	2	Perisinusoidal and portal/periportal fibrosis
	3	Bridging fibrosis
	4	Cirrhosis
BATTS LUDWIG	0	No fibrosis
	1	Fibrous portal expansion
	2	Periportal fibrosis with short septa extending into lobules or rare porto-portal septa (intact architecture)
	3	Fibrous septa reaching adjacent portal tracts and terminal hepatic venule (architecture distortion but no obvious cirrhosis)
	4	Diffuse nodular formation

(b) Clinical Scores in the prognosis of CLD

Clinical Score	Parameter measured	Components of score and equation
Child Turcotte Pugh	Disease severity	*Scores calculated for 5 variables and allocation to Class A (mild) B or C (most severe)
Model End Stage Liver Disease (MELD)	Disease severity to predict survival & to prioritise liver transplantation Maximum value =40	MELD = 3.8[Ln serum bilirubin (mg/dL)] + 11.2[Ln INR] + 9.6[Ln serum creatinine (mg/dL)] + 0.643
Mayo clinical score	Disease severity to predict survival to prioritise liver transplantation	1.209 * $\log_e$(bilirubin in mg/dl) + -3.304 * $\log_e$albumin in gm/dl) + 0.051 * age in years + 2.754 * $\log_e$(prothrombin time in sec) + 0.675 * oedema

Child Pugh Calculation of Score

*Parameter	Points assigned		
	1	2	3
Ascites	Absent	Slight	Moderate
Bilirubin, mg/dL	</= 2	2-3	>3
Albumin, g/dL	>3.5	2.8-3.5	<2.8
Prothrombin time * Seconds over control * INR	 1-3 <1.8	 4-6 1.8-2.3	 >6 >2.3
Encephalopathy	None	Grade 1-2	Grade 3-4

A total score of 5-6 is considered grade A (well-compensated disease); 7-9 is grade B (significant functional compromise); and 10-15 is grade C (decompensated disease). These grades correlate with one- and two-year patient survival.

Grade	Points	One-year patient survival (%)	Two-year patient survival (%)
A: well-compensated disease	5-6	100	85
B: significant functional compromise	7-9	80	60
C: decompensated disease	10-15	45	35

APPENDIX 2: SEARCH STRATEGIES

serum markers.mp. or exp Biological Markers/
2. limit 1 to (human and english language and yr=1990)
3. YKL 40.mp.
4. exp LAMININ/
5. (MMP-2 or TIMP 1).mp.
6. PIIINP.mp.
7. hyaluron$.mp.
8. (MMP$ or TIMP$ or type$ collagen).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
9. (tenascin or $globulin).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
10. non-invasive marker.mp.
11. 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12. aspartate transaminase.mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
13. alanine transferase.mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
14. aminotransferase.mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
15. (ALT or AST).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
16. liver fibrosis marker.mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
17. 12 or 13 or 14 or 15 or 16
18. 11 or 17
19. exp "PREDICTIVE VALUE OF TESTS"/
20. (receiver operat$ adj2 curve).ab,ti.
21. (prognos$ or predict$ or course$).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
22. diagnostic test.mp.
23. exp MORTALITY/
24. exp ROC Curve/
25. exp "Sensitivity and Specificity"/
26. exp Follow-Up Studies/
27. 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26
28. 18 or 27
29. limit 28 to (human and english language and all adult <19 plus years> and yr=1980-2004)
30. alcoholic liver disease.mp. or exp Liver Diseases, Alcoholic/
31. 29 and 30

1 exp ROC CURVE/ (
2. roc.ab. or roc.ti.)
3 sroc.ab. or sroc.ti.
4 accuracy.ab,ti.
5 false negativ$.ab,ti.
6 false positiv$.ab,ti.
7 predictive value$.ab,ti.
8 exp Predictive Value of Tests/
9 specifict$.ab,ti. (
10 sensitivit$.ab,ti.
11 receiver operat$ characteristic$.ab,ti.
12 (receiver operat$ adj2 curve).ab,ti. (
13 roc curve$.ab,ti.
14 diagnos$.ab,ti.
15 exp Cohort Studies/
16 exp INCIDENCE/
17 exp MORTALITY/
18 exp Follow-Up Studies/
19 (prognos$ or predict$ or course$).mp. [mp=title, abstract, name of substance, mesh subject heading]
20 1 or 2 or 3 or4 or5.mp. or 6 or7.mp. or 8 or 9 or 10 or 11 or 12 or 13 or 14 [mp=title, abstract, name of substance, mesh subject heading]
21 15 or 16 or 17 or 18 or 19
22 exp Liver Diseases, Alcoholic/
23 exp Fatty Liver/
24 non alcoholic fatty liver.mp.
25 non-alcoholic steatohepatitis.mp.
26 20 and 22
27 21 and 22
28 exp Hepatitis C/
29 20 and 28

APPENDIX 3: QUADAS TOOL TO ASSESS QUALITY

Item
1. Was the spectrum of patients representative of the patients who will receive the test in practice?
2. Were selection criteria clearly described?
3. Is the reference standard likely to classify the target condition correctly?
4. Is the time period between reference test and index test short enough to be reasonably sure that the target condition did not change between the two tests?
5. Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis?
6. Did patients receive the same reference standard regardless of index test result?
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?
8a. Was the execution of the index test described in sufficient detail to permit its replication?
8b. Was the execution of the reference standard described in sufficient detail to permit its replication
9a. Were the index test results interpreted without knowledge of the results of the reference standard?
10 Were the same clinical data available when test results were interpreted as would be available when the test is used in clinical practice?
11 Were uninterpretable /intermediate test results reported?
12. Were withdrawals from the study explained?

Additional questions were posed in the context of this review:

13a. Was the composition of the panels of serum markers reported in full?

13b. Was any formula derived for the panel of serum markers reported in full?

14. Was there validation in a separate cohort of patients of the panel of serum markers performed.

NAFLD Systematic Review of serum markers of liver fibrosis QUADAS

Author	Q1 Rep sample	Q2 Select criter	Q3 Ref Test Approp	Q4 Ref/index test time short	Q5 Verific	Q6 Verific with same Ref Test	Q7 Ref/Index tests Indep	Q8 Ref Test Reprod	Q9a Inde Test Blind	Q9b Ref Test Blind	Q10 Data same as in pract	Q11 Results report	Q12 Withdra-wal explained	Q13a Index tests comp	Q13b Index Score	Q14 Valid of score
Angulo	Yes	Yes	Yes	Yes	Yes	Yes	Yes	unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	no
Rosenberg	Yes	Yes	Yes	Yes	Yes	Yes	Yes	unclear	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Sakugawa	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	No	Yes	Yes	n/a	No
Albano	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	No
Mofrad	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	n/a	n/a
Shimada	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Dixon	No	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Beymer	No	Yes	Yes	Yes	Yes	Yes	Yes	unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Bugianesi	Yes	Yes	Yes	Yes	Yes	Yes	Yes	unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Dixon	No	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Hui	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	yes	n/a	n/a
Guidorizzi	Yes	Yes	Yes	Yes	Yes	yes	Yes	Unclear	Yes	Yes	Yes	Yes	No	Yes	n/a	n/a
Suzuki	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Angulo	No	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Marchesini	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Hashimoto	Yes	Yes	Yes	Yes	Yes	Yes	Yes	unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Ong	No	No	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Ledinghen	Yes	Yes	Yes	Yes	Yes	Yes	Yes	unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Ratziu	No	Yes	Yes	Yes	Yes	Yes	Yes	unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	no
Sorrentino	No	Y es	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Crespo	No	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Fierbinteanu-Braticevici	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Loguercio	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Santos	No	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a

Yesilova	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Koruk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Hartleb	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a
Chitturi	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a	n/a
Brunt	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	n/a	n/a	n/a

Chronic hepatitis C Systematic Review of panels of serum markers of liver fibrosis QUADAS

Author:	Select criter:	Ref Test Approp:	Ref/ index Test Time Short:	Verific	Verific With Same Ref Test:	Ref / Index Tests Indep:	Ref Test Reprod:	Index Test Blind:	Ref Test Blind:	Data Same as in practice	Results report:	Withdr explained:	Index compon tests:	Index Score of formula	Valid Of score:
Poynard	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes - not all ref	yes	yes	unclear	no
Wai	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes - not all ref	no	yes	yes	yes
Leroy	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes - not all ref	yes	yes	yes	no
Sud	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	yes	yes	yes	yes
Rossi	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	no	yes	no	yes
Myers	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	yes	yes	no	no
Forns	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	yes	yes	no	yes
Myers	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	yes	yes	no	no
Poynard	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	yes	yes	no	no
El-Shorgaby	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	yes	yes	no	no
Thabut	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	no	yes	no	yes
Calvez	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	no	yes	no	yes
Kaul	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	yes	yes	yes	yes
Patel	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	yes	yes	no	yes
Rosenberg	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	yes	yes	yes	yes
Imbert-Bismut	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	yes	yes	no	yes
Poynards	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	no	yes	no	yes
Fortunato	yes	yes	yes	yes	yes	yes	unclear	yes	yes	n/a	yes- not all ref	no	yes	yes	yes

APPENDIX 4: DATA NOT PRESENTED IN CHAPTER 3 SYSTEMATIC REVIEWS

(a) Summary of Serum marker panels components

Serum marker panel	Constituents of panel
Fibroindex	AST, platelets, gamma globulin,
Forns	Age, platelets, AST cholesterol
Fibrotest	Bilirubin, GGT, apolipoprotein A1, a2macroglobulin, haptoglobin
APRI	AST, platelets
Hepascore	Bilirubin, GGT, HA, A2 macroglobulin, age, sex
Fib-4	Platelet, age, AST, ALT
Fibrospect II	HA, TIMP1, A2 macroglobulin
NAFLD fibrosis score	Age,hyperglycemia, BMI, platelets, albumin, AST/ALT
Leroy index	MMP1, P3NP
Cirrhosis Discriminant score (CDS)	Platelets, ALT/AST, INR
Pohl	AST/ALT, platelets
SHASTA	Platlets, Prothrombin index, AST,
Fibrometer	HA, AST, Albumin, A2 macroglobulin, HA, urea, age

(b) Purposive literature review of studies of serum markers in the diagnosis of liver fibrosis >2004

Study year (journal and vol)	Number	Serum marker	Patient population	% sig fibrosis	Diagnostic performance AUC	Threshold	Sens	Spec
Lackner 2005 (Hepatol 41)	194	APRI AST/ALT CDS Platelet count	**CHC consecutive**	50	**0-2 v 3-6** APRI 0.80 AST/ALT 0.57 CDS 0.71 Pohl platelet **0-4 v 5-6** APRI 0.90 AST/ALT 0.73 CDS 0.91 Platelet 0.89	1.5 2 n/r >8 <130 2 1 n/r <150	 n/r 46 41 57 78 n/r 100	 n/r 98 99 100 93 97 n/r 100
Iacobellis 2005 Am J Gastroenterol 100)	1,252	APRI AST/ALT Forns index Platelets	**CHC retrospective**	58	**0,1v 2-4** APRI AST/ALT Forns Platelets	 >1.5 >1 >6.9 <140,00	 59.5 26 79.3 70.6	 88.3 87.5 86.3 86.5
Adams 2005 (Clin Chem 51	221 (117 tr; 104 val)	Hepascore	**CHC prospective**	57	F01 v 2-4 O.85 F0-2 v 3-4 0.96 F0-3 v 4 0.94	0.5 0.5 0.84	63 88 71	89 74 89
Varaut 2005 (Transplantation 80)	110	Fibrotest	**CHC in renal dialysis and renal transplant**	46	**F01 v 2-4** 0.47 dialysis 0.71 transplantation **F0-2 v 3-4** 0.66 dialysis 0.72 transplantation *32% avoid biopsy*	0.3 0.5	84	 93
Lok 2005 (Hepatol 42)	1141 (783 tr; 358 val)	AST/ALT; platelets; INR APRI	**CHC**	34 cirrhosis	F0-4 v 5-6 0.81 0.75,0.86 0.79 0.74,0.85 *(50% avoid biopsy)*	0.5 0.2	54 92	85 30

Study	Number	Serum marker	Patient population	% sig fibrosis	Diagnostic performance AUC	Threshold	Sens	Spec
Bourliere 2006 (J Viral Hep 13)	235	Fibrotest Forns APRI	**CHC consecutive prospective**	42	**F01 v 2-4** Fibrotest 0.81 0.76,0.86 APRI 0.71 0.67,0.79 Forns 0.76,0.70,0.82 **F0-3 v 4** Fibrotest 0.82 APRI 0.81 *81% avoid biopsy*	 0.1 0.6 0.5 1.5 4.2 6.9 n/r 1 2	 97 55 70 22 80 30 n/r 69 38	 20 90 55 95 54 96 n/r 82 96
Yilmaz 2007 (Int J Clin Practice 61)	70	HA	**CHC 13% CHB 84%**	40 cirrhosis	F 0 v1-4 0.86 F0-3 v 4 1.0	63 154 ng/ml	63 90	100 100
Christiensen 2006 (J Viral Hep 13)	139	Fibrospect II	**CHC**	38	**0-2 v 3-6** 0.85 0.79 0.92 *44% avoid biopsy*	0.42 0.9	92 52	64 91
Trocme 2006 (JViral Hep 13)	79	Leroy index	**CHC**	66	**F01 v 2-4** 0.77 **F0-2 v 3-4** 0.81	n/r n/r	n/r n/r	n/r n/r
Halfon 2006 (Am J Gastroenterol 101)	519	Fibrotest	**CHC prospective**	46	**F01 v 2-4** 0.79 0.75,0.82 **F0-2 v 3-4** 0.80 0.76, 0.83	0.36 0.44	73 76	72 70
Atallah 2006 (Clin Chima Acta 369)	455	Albumin;platelets; Alk Phosph;AST/ALT	**CHC**		Cirrhosis	0		97
Sterling 2006 (Hepatol 43)	832 (555 tr; 277 val)	Fib-4	**HIV-HCV**	22	F0-3 v 4-6 0.77 *(71% avoid biopsy)*	1.45 3.25	70	 97

Study	Number	Serum marker	Patient population	% sig fibrosis	Diagnostic performance AUC	Threshold	Sens	Spec
Sebastiani 2006 (J Hepatol 44)	190	Fibrotest Forns APRI	**CHC** consecutive	70	F01 v 2-4 Fibrotest 0.81 0.68,0.90 Forns 0.79 0.68 0.90 APRI 0.69 0.54 0.85 F0-3 v 4 Fibrotest 0.71 APRI 0.61 *50-70% avoid biopsy*	 ? 4.2 6.9 0.5 1.5 ? 2	 65 80 24 84 30 50 39	 81 61 98 77 94 93 87
Parise 2006 (Liver Int 26)	206	HA APRI	**CHC prospective consecutive**	42	**F01 v 2-4** HA 0.88 0.83,0.93 APRI 0.0.82 0.77, 0.90 **F0-3 v 4** HA 0.91 0.87,0.95 APRI 0.84 0.77, 0.90	 34.2 0.70 78.6 1.5	 85 85 91 73	 71 66 81.5 81
Sebastiani 2006 (World J Gastroenterol 13)	125	APRI Fibrotest sequential	**CHB**	71	**F01 v 2-4** **0.78**	0.47	80	63
Zaman 2007 (Am J Med 120)	108	Fibrospect II	**CHC consecutive prospective**	36	F01 v 2-4 0.83	42	72	74
Angulo 2007 (Hepatol 45)	733 (480tr 253 val)	NAFLD fibrosis score	**NAFLD**	17	F0-2 v 3-4 0.82 0.76, 0.88 *75% avoid biopsy*	-1.455 0.676	77 43	71 96

Study	Number	Serum marker	Patient population	% sig fibrosis	Diagnostic performance AUC	Threshold	Sens	Spec
Koda 2007 (Hepatol 45)	360 (240 tr: 120/162 val)	Fibroindex Forns APRI	**CHC** consec with biopsy	50	F01 v 23 FI 0.83 0.78 0.88 Forns 0.79 0.73 0.85 APRI 0.79 0.74 0.85 F0-2 v 3 FI 0.81 0.76 0.87 Forns 0.77 0.70 0.83 APRI 0.80 0.74 0.86 F0-1 v 2-4 FI 0.86 0.81 0.92 Forns 0.84 0.77 0.90 APRI 0.82 0.76 0.89 F0-2 v 3-4 FI 0.85 0.79 0.91 Forns 0.83 0.77 0.89 APRI 0.81 0.74 0.88 *(35% avoid biopsy)*	 2.25: 1.25 8.70; 4.5 0.36; 0.85		
Grigorescu 2007 (J Gastrointesti Liver Dis 16)	206	Fibrotest	**CHC**	63	**F01 v 2-4** **0.78**	0.47	80	63
Snyder 2007 (Clin Chim Acta 381)	93	APRI Fibrospect II	**CHC** **Consecutive prospective**	54	**01v2-4** **APRI** **Fibrospect II** **APRI_Fibrospect II**	 >1.2 <55	 96 82	 62 77

APPENDIX 5: DIAGNOSTIC PERFORMANCE OF ELF (DATA NOT PRESENTED IN CHAPTER 4)

What do the serum markers measure?

1. Breakdown products of extracellular matrix and the enzymes regulating their production
    a. Glycoproteins
        i. Antibodies to **hyaluronic acid**
        ii. **Laminin**
        iii. **Type IV collagen**
        iv. **tenascin**

    b. Propeptides from cleavage of ECM molecules as incorporated into scar
        i. Propeptides of collagen I
        ii. Propeptide collagen III
        iii. Proppeptide collagen IV

    c. Enzymes involved in ECM synthesis
        i. Lysyl oxidase
        ii. Prolyl hydoxylase
        iii. Lysyl hydoxylase

TIMP regulates Metalloproteinases which are involved in the breakdown of collagen.
Hyaluronic Acid is a constituent of ECM.

2. Liver cell synthesis
    a. ά 2 macrogloulin
    b. haptoglobulin
    c. gamma glutaml transpeptidase
    d. γ globulin
    e. apolipoprotein
3. Liver cell function
    a. Bilirubin

Table 1 Diagnostic performance of the Enhanced Liver Fibrosis Panel (ELF) at different thresholds in the combined cohort (1&2) of NAFLD

Stage of Fibrosis	ELF threshold	Sens	Spec	PPV	NPV	LR+	LR-
0 versus 1/2/3/4 (any fibrosis)	8.3467	100	4	60	100	1.04	0.25
	8.7991	95	19	63	71	1.17	0.26
	8.97187	90	27	64	66	1.23	0.37
	9.3585	80	56	72	66	1.82	0.36
	9.793	61	80	81	79	3.05	0.49
	10.2112	45	90	86	53	4.50	0.60
	10.3272	43	95	93	54	8.60	0.60
	11.6454	13	100	100	45	13.00	0.87
0/1 versus 2/3/4 (moderate fibrosis)	8.5783	100	7	42	100	1.08	0.14
	8.9309	95	22	45	86	1.22	0.23
	9.3254	90	50	54	88	1.80	0.20
	9.6375	80	67	62	84	2.42	0.30
	9.8932	70	80	70	80	3.50	0.38
	10.3145	56	90	78	75	5.60	0.49
	10.5734	45	95	85	72	9.00	0.58
	12.2859	9	100	100	62	9.10	0.91
0/1/2 versus 3/4 (severe fibrosis)	8.7587	100	12	26	100	1.14	0.08
	9.2879	98	42	34	98	1.69	0.05
	9.5816	96	57	41	98	2.23	0.07
	9.8932	90	75	52	96	3.60	0.13
	10.3576	80	90	71	94	8.00	0.22
	10.8139	62	95	78	89	12.40	0.40
	11.6454	29	99	87	82	29.00	0.72
	12.2858	16	100	100	80	16.00	0.84

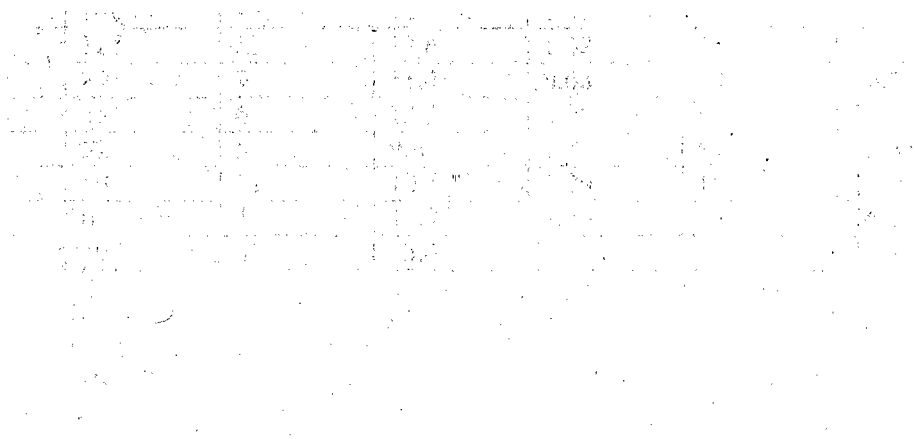
Table 2 Diagnostic performance of the Enhanced Liver Fibrosis Panel (ELF) at different thresholds in the NAFLD Cohort 3 (paediatric)

Fibrosis identified	ELF threshold	Sens	spec	PPV	NPV	LR+/LR -	FP (%)	FN (%)	% cases allocated
Any fibrosis (0 vs 1a-4)									
	8.07 **8.28** 8.53 high & low	95 **88** 79	60 **86** 95	86 **93** 97	86 **78** 69	2.3/0.08 6.3/0.14 16/0.22	4.5 1.8	8 3	100 66
0, 1a vs 1b-4									
	8.09 8.53 8.96 high & low	95 84 51	60 89 96	62 83 88	95 89 74	2.4 7.6 13	6.7 2.7	6.6 1.8	100 61
0-=1b vs 1c-4									
	8.36 8.53 9.1 high & low	92 85 46	77 82 98	72 76 94	93 89 73	4 4.7 23	8 1	9 4	100 65
0-1c vs 2-4									
	10.09 10.18 10.30	100 94 82	88 93 100			8.3 13.4 82	6.2 0	0.8 0	100 88
0-2 v 3,4									
	10.51 10.78 11.56	100 50 25	98 99 100			50 50 25	1.8 0	0 0	100 93

Table 3 AUC values for ELFa and ELF for patients with PBC

Fibrosis	**AUC (95% CI) PBC cohort (baseline time zero & 2 yrs) (n=147)**		**AUC (95% CI) PBC at time zero (n=110)**	
	ELFa	**ELF**	**ELFa**	**ELF**
% 4-6 **	19		20	
0 v 1-6	0.74 (.58,.88)	0.72 (.55,.86)	0.83 (.62,1.0)	0.80 (.57,1.0)
0,1 v 2-6	0.72 (.62,.79)	0.71 (.60,.78)	0.76 (.66,.86)	0.75 (.65,.86)
0-2 v 3-6	.76 (.68,.84)	0.76 (.68,.84)	0.76 (.66.85)	076 (.66,.85)
0-3 v 4-6	0.84 (.77,.92)	0.84 (.76,.92)	0.85 (.77,.93)	0.84 (.76,.92)
0-4 v 5,6	0.86 (.77,.94)	.85 (.76,.94)	0.85 (.75,.96)	0.85 (.74,.95)

ELF has higher AUC values for the identification of more severe disease. ELFa and ELF have the same values at each cut of fibrosis identified.



APPENDIX 6: DATA FROM ELF PROGNOSTIC STUDY

SECTION 1. Analyses for results reported in Chapter 5

A.ELF in centres in England

(i) Rationale for variables used for adjusting in Cox proportional Hazards and Logistic regression analyses

Tables 1a-2 show that prognosis varies by aetiology of CLD and by centre.

Table 1a Liver related clinical outcomes by aetiology of CLD

Aetiology CLD	Fraction of cohort (%)	Liver related outcomes			p value
		Observed	Expected	O/E	
HCV	50	13	33.5	0.4	<0.05
ALD	10	36	6.7	5.4	<0.05
NAFLD	8	1	5.4	0.2	ns
PBC	5	3	3.4	0.9	ns
Other	27	14	18.1	0.8	ns
All	100	67	67		

Table 1b Liver related mortality by aetiology of CLD

Aetiology CLD	Fraction of cohort (%)	Liver related mortality			p value
		Observed	Expected	O/E	
HCV	50	9	22.0	0.4	<0.05
ALD	10	27	4.4	6.1	<0.05
NAFLD	8	0	3.5	0	ns
PBC	5	1	2.2	0.5	ns
Other	27	7	11.9	0.6	ns
All	100	44	44		

Table 2

Centre	Subjects (n)	Liver related outcomes		Relative risk	All cause mortality		Relative risk
		n	%				
1	102	13	12.7	1.33	13	12.7	1
2	125	22	17.6	1.42	16	12.8	1.00
3	79	9	11.4	0.68	13	16.5	1.20
4	45	6	13.3	0.99	6	13.3	1.05
5	52	5	9.6	1	8	15.4	1.21
6	89	11	12.4	1.29	16	18.0	1.41
7	6	1	16.7	1.73	1	16.7	1.31
All	498	67	13.4		73	14.7	

(ii) Table 3 Logistic regression analysis with ELF as continuous variable

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
							Lower	Upper
age1	.012	.016	.561	1	.454	1.012	.981	1.044
abstinent			4.850	3	.183			
Light drinker	-.362	.393	.847	1	.358	.696	.322	1.506
Moderate drinker	-1.157	.660	3.070	1	.080	.315	.086	1.147
Heavy drinker	.252	.496	.258	1	.612	1.286	.487	3.400
HCV			7.340	4	.119			
ALD	.984	.446	4.880	1	.027	2.676	1.117	6.409
NAFLD	-18.924	6286.991	.000	1	.998	.000	.000	.
PBC,PSC,AIH	-.713	.980	.529	1	.467	.490	.072	3.349
Others	.167	.481	.121	1	.728	1.182	.461	3.031
responder(1)	-.810	.820	.977	1	.323	.445	.089	2.217
gender(1)	.198	.386	.262	1	.608	1.219	.572	2.596
Never smoked			1.892	2	.388			
Past smoker	.677	.514	1.732	1	.188	1.967	.718	5.389
Current smoker	.375	.394	.906	1	.341	1.455	.672	3.152
Low ELF tertile			25.553	2	.000			
Mid ELF tertile	2.172	.630	11.895	1	.001	**8.777**	2.554	30.160
High ELF tertile	4.495	.889	25.553	1	.000	**89.604**	15.680	512.045
Constant	-4.578	1.008	20.645	1	.000	.010		

(iii) Table 4 Logistic regression analysis with ELF as categorical variable

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
							Lower	Upper
Age (yr)	.010	.017	.326	1	.568	1.010	.977	1.044
abstinent			2.293	3	.514			
Light drinker	-.309	.421	.539	1	.463	.734	.321	1.676
Moderate drinker	-.827	.702	1.388	1	.239	.437	.111	1.731
Heavy drinker	.177	.535	.109	1	.741	1.194	.418	3.404
elf2	**.779**	**.119**	**42.902**	**1**	**.000**	**2.179**	**1.726**	**2.751**
HCV			3.610	4	.461			
ALD	.742	.474	2.449	1	.118	2.100	.829	5.317
NAFLD	-18.722	6357.686	.000	1	.998	.000	.000	.
PBC,PSC,AIH	-.484	1.000	.235	1	.628	.616	.087	4.373
Others	.092	.497	.034	1	.853	1.096	.414	2.905
responder(1)	-.687	.847	.659	1	.417	.503	.096	2.643
gender(1)	.354	.416	.723	1	.395	1.425	.630	3.219
Never smoked			2.606	2	.272			
Past smoker	.698	.561	1.546	1	.214	2.010	.669	6.038
Current smoker	.637	.428	2.215	1	.137	1.891	.817	4.378
Constant	-2.494	.994	6.300	1	.012	.083		

(iv) Table 5 Logistic regression analysis with biopsy as categorical variable

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
							Lower	Upper
age1	.023	.018	1.638	1	.201	1.023	.988	1.060
abstinent			3.599	3	.308			
Light drinker	-.053	.443	.014	1	.905	.948	.398	2.258
Moderate drinker	-1.240	.897	1.911	1	.167	.289	.050	1.679
Heavy drinker	.557	.592	.884	1	.347	1.745	.547	5.568
HCV			5.646	4	.227			
ALD	1.027	.495	4.294	1	.038	2.792	1.057	7.373
NAFLD	-18.409	6352.122	.000	1	.998	.000	.000	.
PBC,PSC,AIH	-.470	.950	.245	1	.621	.625	.097	4.026
Others	.310	.526	.348	1	.555	1.364	.486	3.823
responder(1)	-1.891	1.107	2.919	1	.088	.151	.017	1.321
gender(1)	.223	.431	.267	1	.605	1.249	.537	2.905
Never smoked			1.876	2	.391			
Past smoker	.691	.623	1.228	1	.268	1.995	.588	6.768
Current smoker	.552	.454	1.479	1	.224	1.737	.713	4.227
Ishak 0-1			34.847	2	.000			
Ishak 2-3	2.062	.540	14.549	1	.000	7.858	2.724	22.666
Ishak 4-6	3.104	.526	34.840	1	.000	22.276	7.948	62.430
Constant	-5.208	1.137	20.976	1	.000	.005		

B. PBC

(i) Table 6 Logistic regression analysis with biopsy as categorical variable

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Low ELF tertile			26.345	2	.000			
Mid ELF tertile	1.059	.446	5.634	1	.018	2.885	1.203	6.918
High ELF tertile	2.422	.491	24.311	1	.000	11.273	4.304	29.529
age	.015	.019	.617	1	.432	1.015	.978	1.052

(ii) Table 7 Logistic regression analysis with biopsy as categorical variable

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
age	.018	.018	1.014	1	.314	1.018	.983	1.054
Ishak 0-1			12.548	2	.002			
Ishak 2-3	.591	.502	1.390	1	.238	1.806	.676	4.828
Ishak 4-6	1.602	.526	9.279	1	.002	4.962	1.770	13.908

SECTION 2. Data from Continental European centres (n=404)

(i) Table 8 Follow up by Europe centre

	State of follow up					
Centre	**Active**	**Lost**	**Discharged**	**Discharged to other hospital**	**Death**	**Total**
1 (n)	25	9	8	1	3	46
Row %	54.3%	19.6%	17.4%	2.2%	6.5%	100.0%
Col %	16.0%	4.0%	66.7%	50.0%	30.0%	11.4%
2 (n)	39	35	4	1	1	80
Row %	48.8%	43.8%	5.0%	1.3%	1.3%	100.0%
Col %	25.0%	15.6%	33.3%	50.0%	10.0%	19.8%
3 (n)	24	67	0	0	1	92
Row %	26.1%	72.8%	.0%	.0%	1.1%	100.0%
Col %	15.4%	29.9%	.0%	.0%	10.0%	22.8%
4 (n)	68	113	0	0	5	186
Row %	36.6%	60.8%	.0%	.0%	2.7%	100.0%
Col %	43.6%	50.4%	.0%	.0%	50.0%	46.0%
Total (n)	156	224	12	2	10	404
Row %	38.6%	55.4%	3.0%	.5%	2.5%	100.0%
Col %	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

(ii) **Table 9 Values for available simple blood tests Continental Europe Centres (n=404)**

	INR	Creat/U LN	ALT	GGT	AST	Gamma glob	Alb	Bili	Alk Phosp	Platelets
Number available	464	482	488	489	490	2	476	488	486	491
Mean (SD)	1.02 (0.1)	0.87 (0.2)	79.3 (86)	76.4 (124)	48.7 (58)	1.35 (0.35)	71.1 (108)	17.5 (29.3)	188 (164)	196 (66)
Median(IQR)	1 (0.95,1.1	0.84 (0.95,1.1	56 (30, 92)	41 (19,81)	32 20, 55	1.35 (1.1,1.6)	44.7 (41,47)	13 (9,17)	147 (105,2 04)	192 (152, 237)
Range (min-max)	0.6-1.65	0.4-1.8	4-809	3 - 1893	6- 851	6-851	24- 674	.93- 495)	43- 1456	59- 556

(iii) Table 10 Values for available serum marker tests Continental Europe (n=404)

	HA	P3NP	TIMP1	Laminin	MMP2	MMP9_timp	Tenascin	Coll IV	Coll VI
Mean (SD)	75.8(12 4)	6.0 (5)	714 (429)	19 (20)	617 (214)	487 (266)	425 (247)	150 (82)	4.4 (1.9)
Median IQR	35.3 (17,71)	4.6 (3,7.0)	605 (452,852)	15 (8,24)	598 (479,735)	451 (309,627)	247 (276,500)	131 (105,167)	4.1 (3,5)
Range	1.9- 1131	1.3-51	3.8-3004	0.6-270	3.9-1857	0.1-1853	83.8-2217	3.1-833	0.6-17

(iv) Table 11 Severity of liver fibrosis on biopsy (Ishak) by centre (Continental Europe)

Centre	**Baseline Fibrosis Score (Ishak) n (%)**							**Total n**
	0	**1**	**2**	**3**	**4**	**5**	**6**	
1	12 (15)	15 (19)	15 (19)	12 (15)	7 (9)	4 (5)	13 (17)	78
2	16 (36)	9 (20)	7(16)	5(11)	3(7)	3 (7)	2(4)	45
3	20 (25)	19(24)	12(15)	8(10)	6 (8)	9(11)	5(6)	79
4	15 (17)	18(20)	17(19)	16(18)	7 (8)	8(9)	8(9)	89
5	47 (26)	28(16)	26(14)	24(13)	16 (9)	15(8)	24(13)	180
Total *(%) within Biopsy)*	110(23)	89*(19)*	77*(16)*	65*(14)*	39*(8)*	39*(8)*	52*(11)*	471

Figure 1 ELF score by aetiology Age ,gender and response to treatment

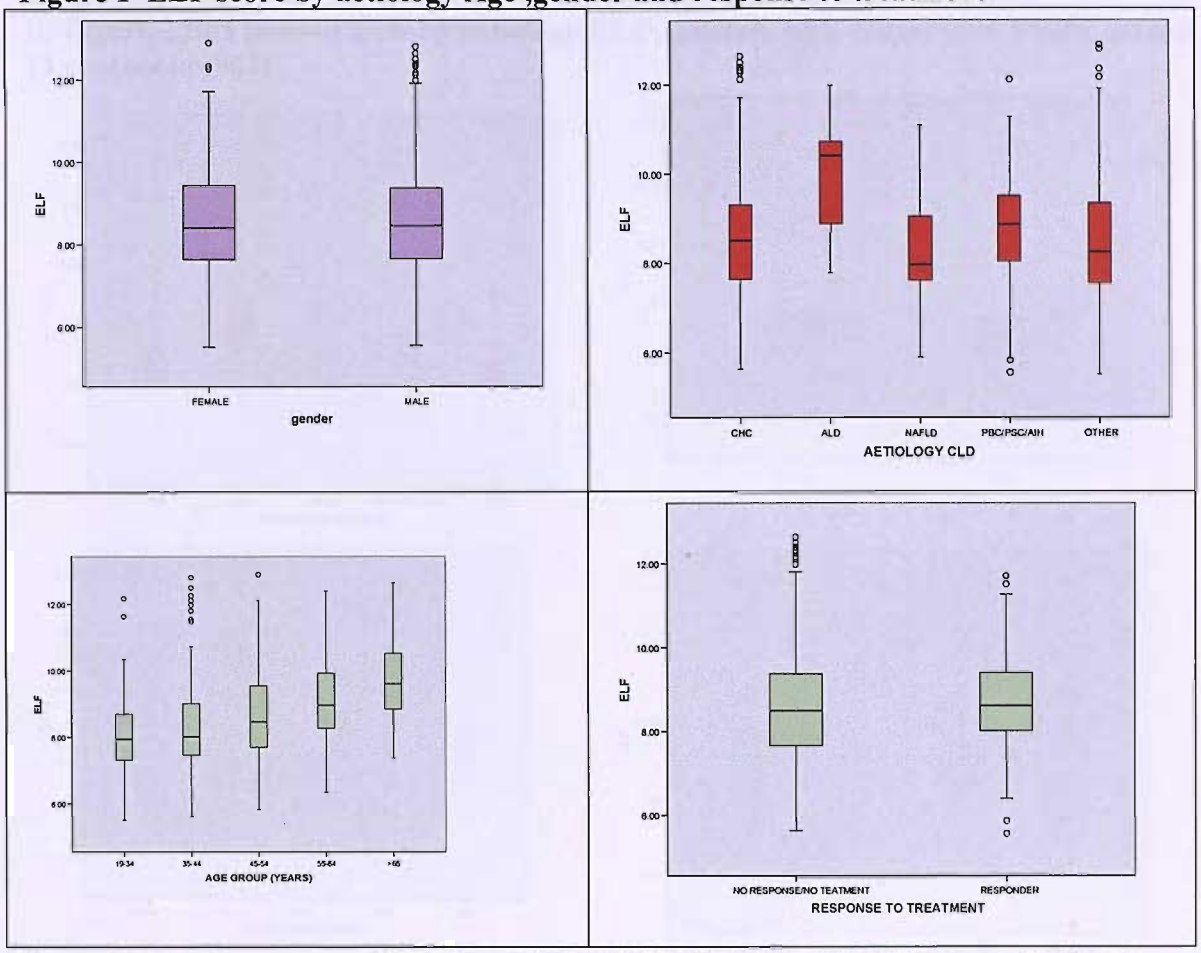
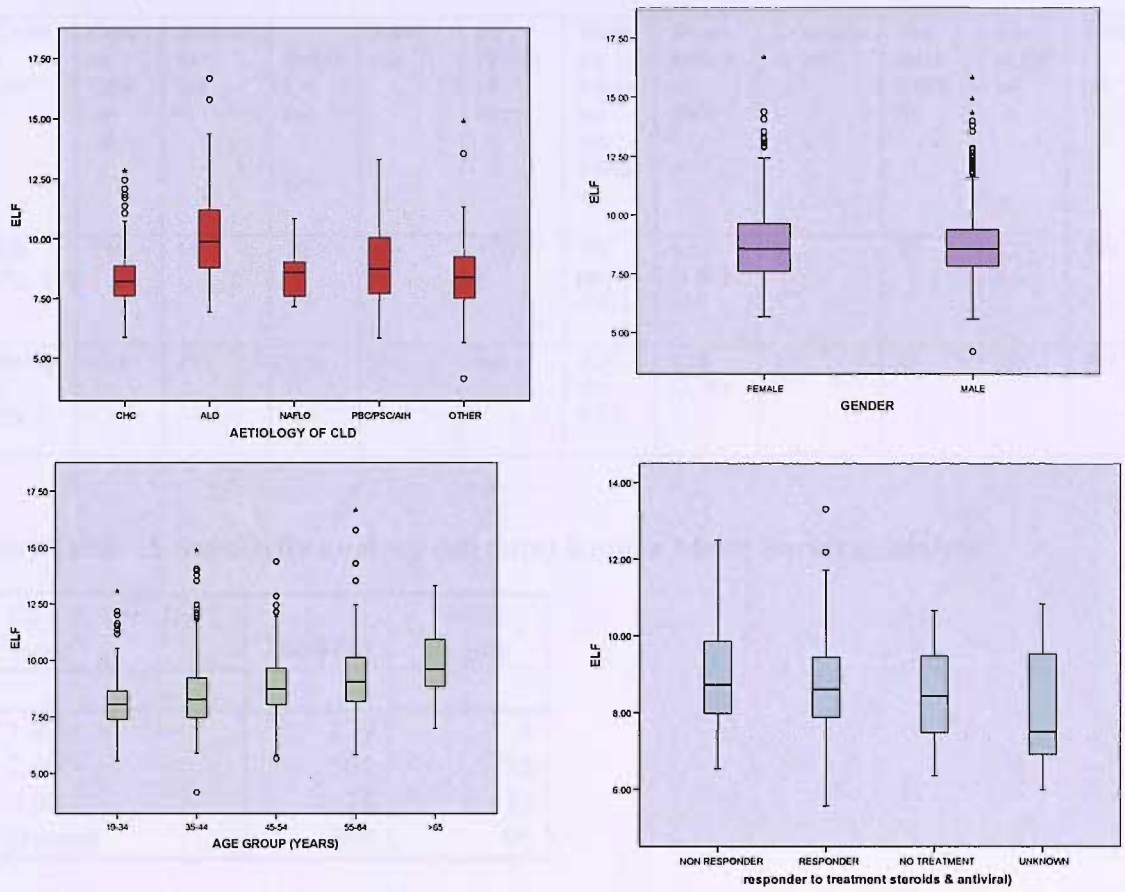


Table 12 Median values of ELF for gender, CLD aetiology, age group, responder to treatment for centres in Continental Europe

		Median	IQR (min-max range)
Gender	**Male**	8.46	1/73 (5.56-12.80)
	Female	8.41	1.82 (5.52-12.89)
Aetiology CLD	**CHC**	8.52	1.68 (5.63-12.64)
	ALD	10.41	2.43 (7.78-11.99)
	NAFLD	7.98	1.69 (5.90-11.09)
	PBC,PSC,AIH	8.87	1.54 (5.56-12.12)
	others	8.26	1.78 (5.52-12.89)
Age group	**19-34**	7.95	1.41 (5.52-12.17)
	35-44	8.02	1.57 (5.62-12.80)
	45-54	8.47	1.86 (5.83-12.89)
	55-64	8.97	1.74 (6.35-12.40)
	>65	9.61	1.72 (7.37-12.64)
Responder to treatment	**No/no treatment**	8.81	1.87 (6.61-12.50)
	responder	8.62	1.42 (5.56-11.71)

C. Sensitivity analysis for all centres in Continental Europe

(i) Figure 2 Box plots of ELF by aetiology CLD, gender, Age, response to treatment (all 11 centres (n=902)



(ii) Table 13 Mean (SD) values of ELF for gender, CLD aetiology, age group, responder to treatment for All centres

		Median	IQR (min-max range)
Gender	Male	8.55	1.55 (4.16-15.78)
	Female	8.58	2.04 (5.66-16.67)
Aetiology CLD	CHC	8.25	1.25 (5.88-12.83)
	ALD	9.90	2.53 (6.95-16.67)
	NAFLD	8.63	1.44 (7.16-10.84)
	PBC,PSC,AIH	8.77	2.46 (5.84-13.31)
	others	8.42	1.73 (4.16-14.89)
Age group	19-34	8.06	1.26 (5.56-13.06)
	35-44	8.27	1.75 (4.16-14.89)
	45-54	8.74	1.61 (5.66-14.37)
	55-64	9.06	1.96 (5.84-16.67)
	>65	9.61	2.10 (7.01-13.31)
Responder to treatment	No/no treatment	8.72	1.91 (6.53-12.50)
	responder	8.60	1.58 (5.56-13.31)

(iii) Table 14 Follow up destinations of patients in England and Europe

Centre (n)	Lost to follw up (n)*	D/charge (n)	Deaths (n)	LRO (n)	GP returned (n)	Median follow up yrs (range)	Mean Follow up (SD)	Transplant (n)	Non fatal LRO (n)	Lost to GP (n)	Total (n)
UK (7)	146	65	70	67	171/198	7.0 (0.1-9.01)	6.53 (1.87)	6	17	12/170	498
Europe (4)	224	14	10	30	n/a	6.36 (0-8.84)	4.98 (2.96)	11	11	n/a	404

(iv) Table 15 Sensitivity analysis (all data) Kaplan Meier Survival analysis

ELF tertiles by score	Total No	No of Events
1.00	379	6
2.00	502	76
3.00	18	14
Overall	899	96

Biopsy	Total No	No of Events
1.00 (0-1)	398	14
2.00 (2-3)	226	15
3.00 (4-6)	224	60
Overall	848	89

Figure 3 Kaplan Meier curve of survival from liver-related outcomes for ELF tertiles (All centres)

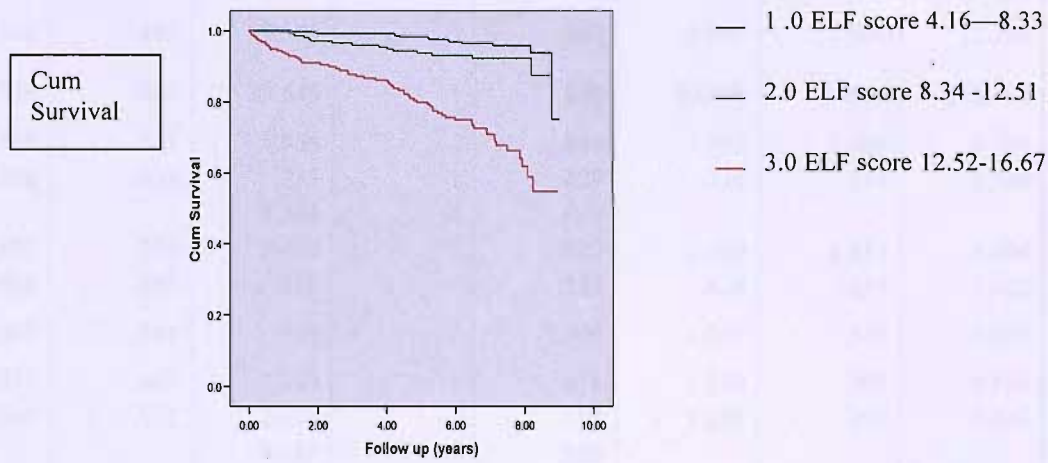


Figure 4 Kaplan Meier curve of survival from all-cause mortality for ELF tertiles all centres

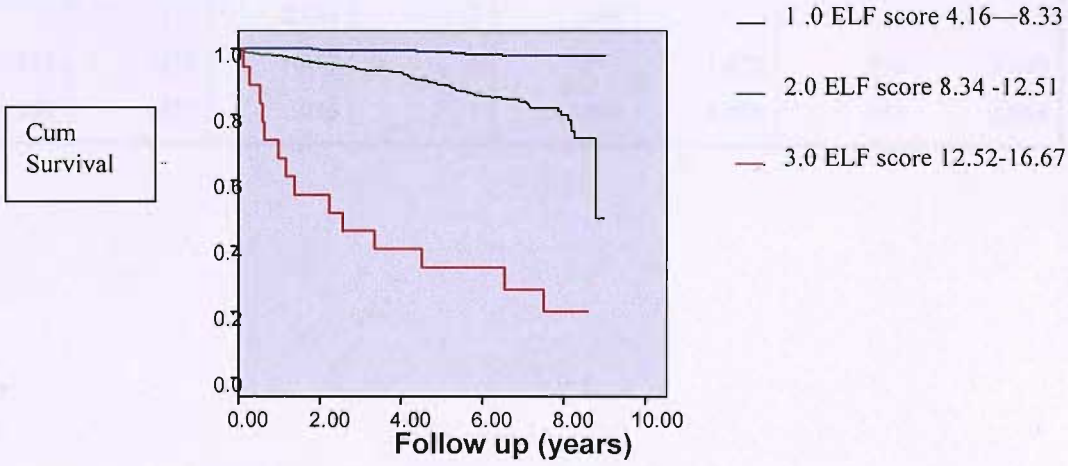


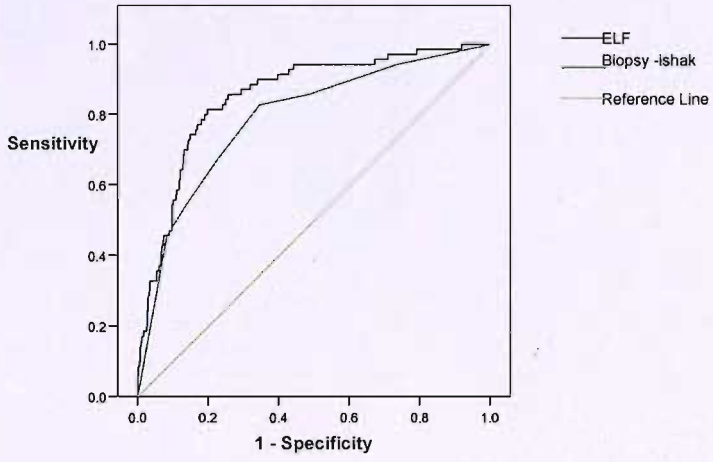
Table 16 Cox Proportional Hazards model for liver related outcomes –All centres

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Low ELF tertile			44.025	2	.000			
Mid ELF tertile	1.546	.489	9.995	1	.002	4.695	1.800	12.246
High ELF tertile	3.929	.624	39.645	1	.000	50.860	14.969	172.798
gender	.689	.351	3.854	1	.050	1.992	1.001	3.964
responder	-.306	.634	.233	1	.629	.736	.213	2.549
HCV			8.364	4	.079			
ALD	.875	.375	5.452	1	.020	2.400	1.151	5.004
NAFLD	-.468	.786	.355	1	.552	.626	.134	2.922
PBC,PSC,AIH	.007	.554	.000	1	.990	1.007	.340	2.982
Others	.231	.463	.249	1	.618	1.260	.509	3.119
age1	.020	.012	2.626	1	.105	1.020	.996	1.046
abstinent			9.087	3	.028			
Light drinker	-.570	.330	2.972	1	.085	.566	.296	1.081
Moderate drinker	-1.135	.546	4.314	1	.038	.321	.110	.938
Heavy drinker	.289	.374	.598	1	.439	1.336	.641	2.781
Never smoked			2.133	2	.344			
Past smoker	.513	.376	1.857	1	.173	1.670	.799	3.489
Current smoker	.064	.337	.036	1	.849	1.066	.551	2.065

Table 17 Logistic regression at 6 years follow up

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
							Lower	Upper
Age	.012	.018	.447	1	.504	1.012	.977	1.050
ELF	.823	.134	37.643	1	.000	2.277	1.751	2.961
HCV			5.074	4	.280			
ALD	.933	.504	3.420	1	.064	2.542	.946	6.832
NAFLD	-.126	.900	.020	1	.888	.881	.151	5.139
PBC,PSC,AIH	-.244	.737	.110	1	.740	.783	.185	3.319
Others	.268	.594	.204	1	.652	1.308	.408	4.191
gender(1)	.762	.473	2.593	1	.107	2.143	.847	5.418
abstinent			5.715	3	.126			
Light drinker	-.767	.453	2.865	1	.091	.465	.191	1.129
Moderate drinker	-1.168	.752	2.415	1	.120	.311	.071	1.357
Heavy drinker	.223	.569	.153	1	.696	1.249	.410	3.809
Never smoked			4.532	2	.104			
Past smoker	1.043	.583	3.204	1	.073	2.838	.906	8.891
Current smoker	.905	.481	3.541	1	.060	2.471	.963	6.339
responder	-.118	.839	.020	1	.888	.889	.172	4.603
Constant	-2.992	1.107	7.303	1	.007	.050		

Figure 5 ROC analyses at 6 years follow up



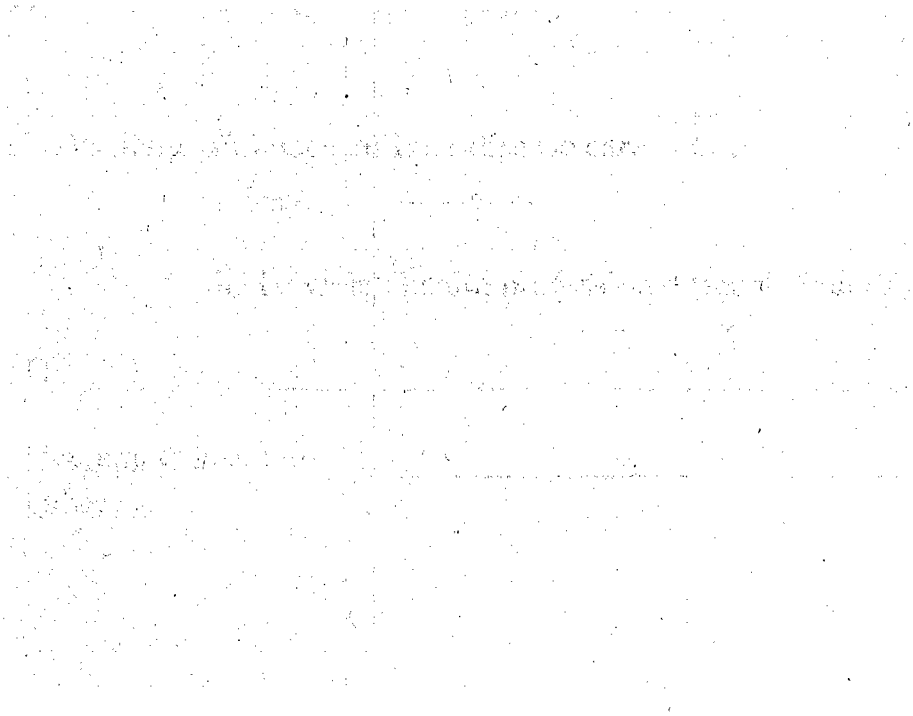
	AUC	95% Confidence Interval	
ELF	.85	.806	.90
Biopsy (Ishak)	.79	.731	.848

Section C ELF Performance of Key Assays

Table 18 ELF Performance of Key Assays

		Coefficient of variation (%)				
Assay	Range	Interassay variation	total	Min detectable Level (ng/ml)	Linearity (%)	Ref range (ng/ml) mean±SD (no)
P3NP	0-151	1.0-5.3	1.9-5.3	0.5	90.5-100.1	5.84 ± 3.26(199)
TMP1	0-300	3.0-2.8	4.7-3.2	<2	98.5-101.8	619 ± 111.7 (225)
HA	0-1000	<10	10-12	5.07	94.7-100.8	8.16 ± 8.21 to 34.6±22.47(382)

The performance of each assay was determined using samples obtained from healthy donors of both sexes aged 18-75 years. The range, minimum detectable level and coefficient of variation presented for interassay variation as well as total variation are shown. The results for linearity as the range of variation around 100%.



APPENDIX 7: Data extraction form Prognostic study

PROGNOSTIC PERFORMANCE OF SERUM MARKERS OF LIVER FIBROSIS

1. Date completed: ☐☐/☐☐/☐☐
   d d m m y y

2. Cause of Liver Disease ----------

3. Date last known alive ☐☐/☐☐/☐☐

4. Follow up status by your hepatology team(tick)

   (5a) Currently under review by hepatology team ☐

   (5b) Lost to follow up by hepatology team ☐ Date -- --/-- --/-- --

   (5c) Discharged/transfer of liver disease care ☐

   i. Date -- --/-- --/-- --

   ii. To which health professional (name & address)

- GP _____
- Hospital Consultant _____
- Other

6. Liver biopsy **AFTER RECRUITMENT TO STUDY**
(if cirrhosis on biopsy look carefully for any signs of decompensation in notes)

Biopsy data	Date	Reason	Result
1. no.portal tracts -------- Length (mm) ---------- Not available --------- 2. no.portal tracts -------- Length (mm) ---------- Not available ---------			

7. Has the patient had any clinical signs of the following **since recruitment?**
Please tick if yes and give date of FIRST event

Varices ☐ Date ☐☐/☐☐/☐☐

Variceal bleeding: ☐ Date ☐☐/☐☐/☐☐

Ascites* ☐ Date ☐☐/☐☐/☐☐

Encephalopathy ☐ Date ☐☐/☐☐/☐☐

Hepatocellular Cancer ☐ Date ☐☐/☐☐/☐☐

Cholangiocarcinoma ☐ Date ☐☐/☐☐/☐☐

Other/comment:

_____

_____

_____

*please state which method was used to diagnose ascites:
Clinical Ultrasound Paracentesis

8 . **Treatment**

If Hepatitis C:

a. HCV anti virals used Yes ☐ No ☐

b. Antiviral medication used _____

c. Date started _____ Duration treatment _____ mths

d. Outcome treatment i)Responder ☐ SVR follow up time ____ mths

ii)Non-responder ☐

iii)Relapser ☐ treatment to relapse _____mths

If Hepatitis B:

e. HBV treatment Yes ☐ No ☐

f. Antiviral medication used _____

g. Date started _____ Duration treatment _____ mths

h. Sero convert Yes ☐ No ☐

If other liver disease:

i. Oral steroids Date started _____ Duration treatment _____ mths

j. URSO Date started _____ Duration treatment _____ mths

Other (liver disease relevant)

_____

_____

9. Final outcome

a) Death Yes ☐ No ☐

If Yes; Date of death ☐☐/☐☐/☐☐

What was the cause of death?

.....

.....

.....

Occurred since recruitment:

b) Transplant Yes ☐ No ☐ **Date transplant** ----------

10. Summary of case (state liver disease and clinical condition at recruitment and consequent course of disease)

Other comments ----------

University of Southampton

ID study number: ☐☐☐☐

Predictive accuracy of serum markers of liver fibrosis study

Date questionnaire completed: ----/----/----
dd mm yy

Q1. When was the patient last known to be alive?
(e.g last surgery attendance/out-patients/blood test etc)

----/----/----
dd mm yy

Q2. Is this patient currently registered with your practice?

YES ☐ NO ☐

Q3. Has the patient had any complications of cirrhosis-ascites, bleeding varices, encephalopathy hepatocellular carcinoma?

☐ ☐ ☐

YES NO Don't Know

If yes please give date of 1st episode recorded

----/----/----
dd mm yy

Thank you very much for your time in completing this questionnaire. Please return the questionnaire in the pre-paid envelope provided.

February 2007

Study title: Predictive accuracy of serum markers of liver fibrosis MREC 98/6/08

RE: Patient Name DOB

Dear Dr

We are conducting a study to evaluate the ability of non-invasive serum markers to predict serious clinical outcomes in chronic liver disease. This would reduce the need for liver biopsy and provide a valuable tool in the management of patients with liver disease. We are following up patients recruited to a diagnostic study in 1998-2000 in which serum markers were compared to biopsy. We have follow-up for mortality, but we need to ascertain whether they have had any episodes of decompensated liver disease such as oesophageal variceal bleeding, ascites, encephalopathy or hepatocellular cancer. We have MREC approval to review their medical notes including those in primary care. The patient named above participated in the original diagnostic study but has been lost to follow up by the recruiting centre and your practice has been identified as the last practice in which they were registered. We would be very grateful if you could review their notes and complete the very short questionnaire attached to this letter. All questionnaires have been anonymised to maintain confidentiality. Complete follow up on as many patients as possible is vital for a robust reliable result and we would be very grateful for your help in this important study. We would be glad to answer any queries you may have.

Yours sincerely

Dr Julie Parkes MRC Clinical training Fellow/Lecturer Public Health Medicine
Tel: 02380 794206 jules@soton.ac.uk
Dr Paul Roderick Reader Public Health Medicine pjr@soton.ac.uk
Professor William Rosenberg Professor of Hepatology wmr@soton.ac.uk

APPENDIX 9: PUBLISHED PAPERS & PRESENTED/ACCEPTED CONFERENCE ABSTRACTS ARISING FROM RESEARCH IN THIS THESIS

Parkes J., Guha N., Rosenberg W., Roderick P. J Hepatol 44(2006) 462-474
Systematic review of panels of serum markers for liver fibrosis in Hepatitis C

I N Guha, **J Parkes**, P R Roderick, S Harris, and W M Rosenberg **Gut** 2006; 55: 1650-1660
Non-invasive markers associated with liver fibrosis in non-alcoholic fatty liver disease (NAFLD)

IN Guha, **J Parkes**, P Roderick, and others. Non-invasive markers of fibrosis in non-alcoholic fatty liver disease: validating the European Liver Fibrosis panel and exploring simple markers. *Hepatology* 47(2): 455 - 460. February 2008

W.Rosenberg and **J. Parkes.** Biomarkers of Liver Fibrosis-The Enhanced Liver Fibrosis Markers Clinical Laboratory International 2007

Conferences

***Presidential Poster (Top 10% abstracts)**

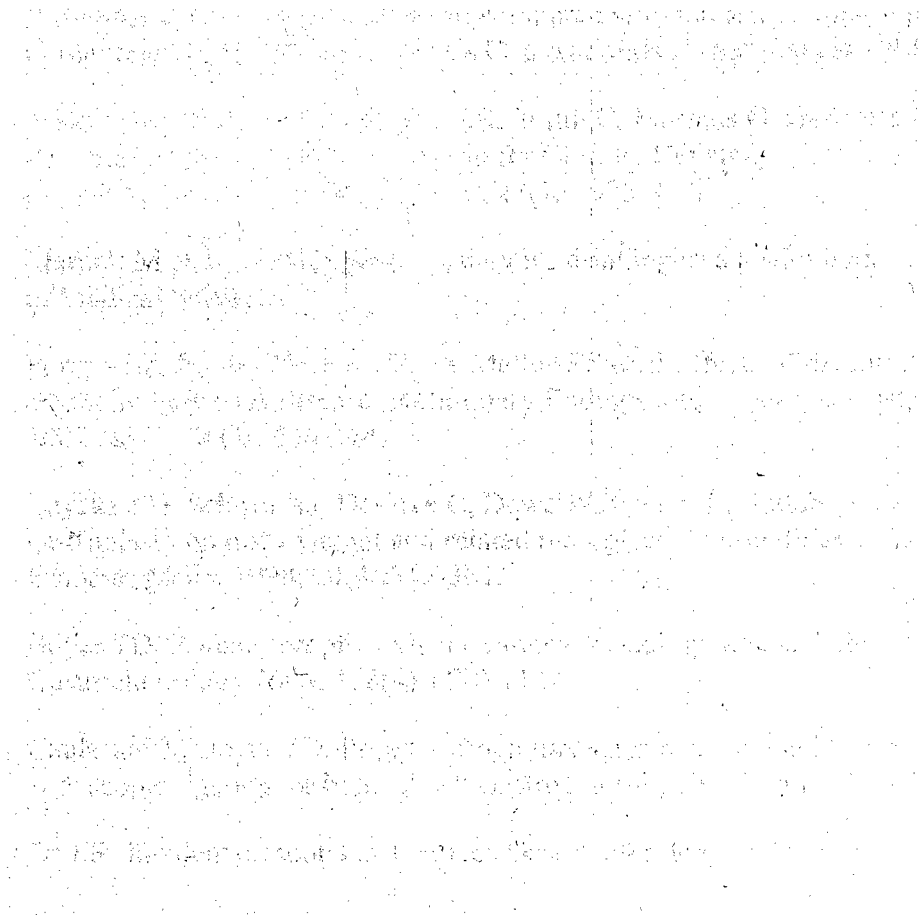
- *Presidential Poster. Annual Conference of American Association for the Study of Liver Disease ELF serum markers accurately distinguishes fibrosis severity in primary biliary cirrhosis: An external validation study. Parkes et al. Boston 2006
- *Presidential Poster presentation. Annual Conference of American Association for the Study of Liver Disease. ELF serum markers accurately distinguish fibrosis severity in non-alcoholic fatty liver disease: An external validation study. **Parkes et al.** Boston 2006
- *Presidential Poster presentation. Annual Conference American Association for the Study of Liver Disease European Liver Fibrosis (ELF) panel of serum markers can predict clinical outcome in a cohort of patients from England with mixed aetiology chronic liver disease 2007 **Parkes et al** Boston 2007
- *Award winning poster presentation. Analytical and clinical evaluation of the Enhanced Liver Fibrosis panel W Rosenberg, J Parkes,et al Annual Conference of Analytical Chemistry Chicago 2007
- Poster presentation Annual Conference of American Association for the Study of the Liver Systematic review of panels of surrogate non-invasive markers of liver fibrosis in Hepatitis C. **J Parkes** et al. San Fransisco November 2005
- Poster presentation Annual Conference of European Association for the Study of the Liver Vienna 2006 Evaluation of ELF Serum markers of liver fibrosis in Chronic Hepatitis C **J. Parkes et al.**
- Poster presentation Annual Conference of American Association for the Study of the Liver ELF serum markers accurately distinguish fibrosis severity in chronic hepatitis C: An external validation study. **Parkes et al.** Boston 2006
- Poster presentation Annual Conference of American Association for the Study of the Liver Serum markers predict future clinical decompensation better than biopsy

bilirubin or Mayo score in Primary Biliary Cirrhosis Mayo M., **Parkes J. et al.** Boston 2006

Accepted at conferences

- Annual Conference of British Society of Gastroenterology Birmingham March 2008: **Oral presentation** European Liver Fibrosis (ELF) panel of serum markers can predict clinical outcome in a cohort of patients from England with mixed aetiology chronic liver disease

- Annual Conference of British Society of Gastroenterology Birmingham March 2008: Poster presentation: Systematic review of the diagnostic performance of serum markers of liver fibrosis in Alcoholic Liver Disease **Parkes J;** Guha IN; Harris S; Rosenberg W; Roderick P;



Reference List

(1) Department of Health. On the State of the Public Health. Annual Report of the Chief Medical Officer. 2001. London, Department of Health.

(2) Leon DA, McCambridge J. Liver cirrhosis mortality rates in Britain, 1950 to 2002 35. Lancet 2006; 367(9511):645.

(3) Department of Health. Hepatitis C Strategy for England. Department of Health Publications, editor. 2002. London.

(4) Sweeting MJ, De AD, Brant LJ, Harris HE, Mann AG, Ramsay ME. The burden of hepatitis C in England J Viral Hepat 2007; 14(8):570-576.

(5) Neuschwander-Tetri BA, Caldwell SH. Nonalcoholic steatohepatitis: summary of an AASLD Single Topic Conference Hepatology 2003; 37(5):1202-1219.

(6) Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet 2001; 358:958-965.

(7) Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL et al. Peginterferon Alfa-2a plus Ribavirin for Chronic Hepatitis C Virus Infection. The New England Journal of Medicine 2002; 347(13):975-982.

(8) Marmot M et al. Calling time: the nation's drinking as a health issue. 2004. Academy of Medical Sciences.

(9) Furuya CK Jr., de Oliveira CP., de Mellos ES et al Effects of bariatric surgery on non-alcoholic fatty liver disease: preliminary findings after 2 years. J Gastrenterol Hepatol 2007 April; 22 (4): 510-514.

(10) Luyckx FH, Scheen AJ, Desaive C, Dewe W, Gielen JE, Lefebvre PJ. Effects of gastroplasty on body weight and related biological abnormalities in morbid obesity. Diabetes Metab 1998; 24(4):355-361.

(11) Boyer TD. Primary prophylaxis for variceal bleeding: are we there yet? Gastroenterology 2005; 128(4):1120-1122.

(12) Chalasani N, Boyer TD. Primary prophylaxis against variceal bleeding: beta-blockers, endoscopic ligation, or both? Am J Gastroenterol 2005; 100(4):805-807.

(13) De FR. Incidental esophageal varices Gastroenterology 2004; 126(7):1860-1867.

(14) Jalan R, Hayes PC. UK Guidelines on the management of variceal haemorrhage in cirrhotic patients. British Society of Gastroenterology Guidelines in Gastroenterology 2000 June:[1-1.5] Available from: URL:www.bsg.org.uk

(15) Thalheimer U, Bosch J, Burroughs AK. How to prevent varices from bleeding: shades of grey--the case for nonselective beta blockers Gastroenterology 2007; 133(6):2029-2036.

(16) Tanaka H, Nouso K, Kobashi H, Kobayashi Y, Nakamura S, Miyake Y et al. Surveillance of hepatocellular carcinoma in patients with hepatitis C virus infection may improve patient survival Liver Int 2006; 26(5):543-551.

(17) Barbara L, Benzi G, Gaiani S, Fusconi F, Zironi G, Siringo S et al. Natural History of Small Untreted Hepatocellular Carinoma in Cirrhosis: A Multivariate Analysis of Prognostic Factors of Tumour Growth Rate and Patient Survival. Hepatology 1992; 16:132-137.

(18) Colombo M, De FR, Del NE, Sangiovanni A, De FC, Tommasini M et al. Hepatocellular carcinoma in Italian patients with cirrhosis N Engl J Med 1991; 325(10):675-680.

(19) Taura N, Hamasaki K, Nakao K, Ichikawa T, Nishimura D, Goto T et al. The impact of newer treatment modalities on survival in patients with hepatocellular carcinoma Clin Gastroenterol Hepatol 2006; 4(9):1177-1183.

(20) Zhang BH, Yang BH, Tang ZY. Randomized controlled trial of screening for hepatocellular carcinoma J Cancer Res Clin Oncol 2004; 130(7):417-422.

(21) Toyoda H, Kumada T, Kiriyama S, Sone Y, Tanikawa M, Hisanaga Y et al. Impact of surveillance on survival of patients with initial hepatocellular carcinoma: a study from Japan Clin Gastroenterol Hepatol 2006; 4(9):1170-1176.

(22) Yang B, Zhang B, Xu Y, Wang W, Shen Y, Zhang A et al. Prospective study of early detection for primary liver cancer J Cancer Res Clin Oncol 1997; 123(6):357-360.

(23) Di Bisceglie AM. Issues in screening and surveillance for hepatocellular carcinoma Gastroenterology 2004; 127(5 Suppl 1):S104-S107.

(24) Wong GL, Wong VW, Tan GM, Ip KI, Lai WK, Li YW et al. Surveillance programme for hepatocellular carcinoma improves the survival of patients with chronic viral hepatitis Liver Int 2008; 28(1):79-87.

(25) Ryder SD. Guidelines for the diagnosis and treatment of hepatocellular carcinoma (HCC) in adults GUT 2003; 52 Suppl 3:iii1-iii8.

(26) Thompson CJ, Rogers G, Hewson P, Wright D, Anderson R, Cramp M et al. Surveillance of cirrhosis for hepatocellular carcinoma: systematic review and economic analysis Health Technol Assess 2007; 11(34):1-206.

(27) Foucher J, Chanteloup E, Vergniol J, Castera L, Le BB, Adhoute X et al. Diagnosis of cirrhosis by transient elastography (FibroScan): a prospective study Gut 2006; 55(3):403-408.

(28) Lim AK, Taylor-Robinson SD, Patel N, Eckersley RJ, Goldin RD, Hamilton G et al. Hepatic vein transit times using a microbubble agent can predict disease severity non-invasively in patients with hepatitis C Gut 2005; 54(1):128-133.

(29) Talwalkar JA, Yin M, Fidler JL, Sanderson SO, Kamath PS, Ehman RL. Magnetic resonance imaging of hepatic fibrosis: emerging clinical applications Hepatology 2008; 47(1):332-342.

(30) Rosenberg WM, Voelker M, Thiel R, Becka M, Burt A, Schuppan D et al. Serum markers detect the presence of liver fibrosis: a cohort study. Gastroenterology 2004; 127(6):1704-1713.

(31) Friedman SL, Arthur MJ. Reversing fibrosis. Sci Med 8, 194-205. 2002.

(32) Friedman SL. Liver fibrosis -- from bench to bedside. J Hepatol 2003; 38 Suppl 1:S38-S53.

(33) Shiratori Y, Imazeki F, Moriyama M, Yano M, Arakawa Y, Yokosuka O et al. Histologic improvement of fibrosis in patients with hepatitis C who have sustained response to interferon therapy Ann Intern Med 2000; 132(7):517-524.

(34) Desmet VJ, Roskams T. Cirrhosis reversal: a duel between dogma and myth. J Hepatol 2004; 40(5):860-867.

(35) Tome S, Lucey MR. Review article: current management of alcoholic liver disease Aliment Pharmacol Ther 2004; 19(7):707-714.

(36) Gramenzi A, Caputo F, Biselli M, Kuria F, Loggi E, Andreone P et al. Review article: alcoholic liver disease--pathophysiological aspects and risk factors Aliment Pharmacol Ther 2006; 24(8):1151-1161.

(37) Quero JC, Schalm SW. Subclinical hepatic encephalopathy Semin Liver Dis 1996; 16(3):321-328.

(38) Amodio P, Quero JC, Del PF, Gatta A, Schalm SW. Diagnostic tools for the detection of subclinical hepatic encephalopathy: comparison of standard and computerized psychometric tests with spectral-EEG Metab Brain Dis 1996; 11(4):315-327.

(39) Velazquez RF, Rodriguez M, Navascues CA, Linares A, Perez R, Sotorrios NG et al. Prospective analysis of risk factors for hepatocellular carcinoma in patients with liver cirrhosis Hepatology 2003; 37(3):520-527.

(40) Faivre J, Forman D, Esteve J, Obradovic M, Sant M. Survival of patients with primary liver cancer, pancreatic cancer and biliary tract cancer in Europe. EUROCARE Working Group Eur J Cancer 1998; 34(14 Spec No):2184-2190.

(41) Brunt EM. Grading and staging the histopathological lesions of chronic hepatitis: the Knodell histology activity index and beyond. Hepatology 31, 241-246. 2006.

(42) Kleiner DE, Brunt EM, Van NM, Behling C, Contos MJ, Cummings OW et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease Hepatology 2005; 41(6):1313-1321.

(43) Standish RA, Cholongitas E, Dhillon A, Burroughs AK, Dhillon AP. An appraisal of the histopathological assessment of liver fibrosis GUT 2006; 55(4):569-578.

(44) Poynard T, Mathurin P, Lai CL, Guyader D, Poupon R, Tainturier MH et al. A comparison of fibrosis progression in chronic liver disease. Journal of Hepatology 2003; 38(3):257-265.

(45) Ryder SD Progression of hepatic fibrosis in patients with hepatitis C: a prospective repeat liver biopsy study. GUT 2004; 53:451-455.

(46) Lagging LM, Westin J, Svensson E, Aires N, Dhillon AP, Lindh M et al. Progression of fibrosis in untreated patients with hepatitis C virus infection. Liver 2002; 22(2):136-144.

(47) Yano M, Kumada H, Kage M, Ikeda K, Shimamatsu K, Inoue O et al. The long-term pathological evolution of chronic hepatitis C Hepatology 1996; 23(6):1334-1340.

(48) Kage M, Shimamatu K, Nakashima E, Kojiro M, Inoue O, Yano M. Long-term evolution of fibrosis from chronic hepatitis to cirrhosis in patients with hepatitis C: morphometric analysis of repeated biopsies Hepatology 1997; 25(4):1028-1031.

(49) Boccato S, Pistis R, Noventa F, Guido M, Benvegnu L, Alberti A. Fibrosis progression in initially mild chronic hepatitis C J Viral Hepat 2006; 13(5):297-302.

(50) Angulo P, Keach JC, Batts KP, Lindor KD. Independent predictors of liver fibrosis in patients with nonalcoholic steatohepatitis Hepatology 1999; 30(6):1356-1362.

(51) Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity Gastroenterology 1999; 116(6):1413-1419.

(52) Sweeting MJ, De AD, Neal KR, Ramsay ME, Irving WL, Wright M et al. Estimated progression rates in three United Kingdom hepatitis C cohorts differed according to method of recruitment J Clin Epidemiol 2006; 59(2):144-152.

(53) D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies J Hepatol 2006; 44(1):217-231.

(54) Aube C, Oberti F, Korali N, Namour MA, Loisel D, Tanguy JY et al. Ultrasonographic diagnosis of hepatic fibrosis or cirrhosis. Journal of Hepatology 1999; 30(3):472-478.

(55) Cadranel JF, Rufat P, Degos F, Tremolada F, Diodati G, Almasio P. Practices of liver biopsy in France: results of a prospective nationwide survey. Hepatology 32, 477-481. 2006.

(56) Abdi W, Millan JC, Mezey E. Sampling variability on percutaneous liver biopsy Arch Intern Med 1979; 139(6):667-669.

(57) Regev A, Berho M, Jeffers LJ, Milikowski C, Molina EG, Pyrsopoulos NT et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection Am J Gastroenterol 2002; 97(10):2614-2618.

(58) Siddique I, El Naga HA, Madda JP, Memon A, Hasan F. Sampling variability on percutaneous liver biopsy in patients with chronic hepatitis C virus infection. Scandinavian Journal of Gastroenterology 2003; 38(4):427-432.

(59) Rousselet MC, Michalak S, Dupre F, Croue A, Bedossa P, Saint-Andre JP et al. Sources of variability in histological scoring of chronic viral hepatitis. Hepatology 2005; 41(2):257-264.

(60) Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C Hepatology 2003; 38(6):1449-1457.

(61) Colloredo G, Guido M, Sonzogni A, Leandro G. Impact of liver biopsy size on histological evaluation of chronic viral hepatitis: the smaller the sample, the milder the disease. Journal of Hepatology 2003; 39:1-6.

(62) Griffiths C, Rooney C, Brock A. Leading causes of death in England and Wales -how should we group causes? Health Statistics Quarterly 28, 6-17. 2005.

(63) Bosetti C, Levi F, Lucchini F, Zatonski WA, Negri E, La VC. Worldwide mortality from cirrhosis: an update to 2002 J Hepatol 2007; 46(5):827-839.

(64) Clark JM, Brancati FL, Diehl AM. The prevalence and etiology of elevated aminotransferase levels in the United States Am J Gastroenterol 2003; 98(5):960-967.

(65) Bellentani S, Saccoccio G, Masutti F, Croce LS, Brandi G, Sasso F et al. Prevalence of and risk factors for hepatic steatosis in Northern Italy Ann Intern Med 2000; 132(2):112-117.

(66) Browning JD, Szczepaniak LS, Dobbins R, Nuremberg P, Horton JD, Cohen JC et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity Hepatology 2004; 40(6):1387-1395.

(67) Butland B. Tackling obesities Future choices project. Government Office for Science, editor. 2007.

(68) Health Protection Agency, SCIEH, National Public Health Service for Wales, CDSC Northern Ireland, CRDHB, UASSG. Shooting up:infections among injecting drug users in the United Kingdom 2005. 2006. London, Health Protection Agency.

(69) World Health Organisation. Weekly epidemiological record. Weekly epidemiological record Releve epidemiologique hebdomadaire 2002; 77(No 6):41-48.

(70) Law MG, Dore GJ, Bath N, Thompson S, Crofts N, Dolan K et al. Modelling hepatitis C virus incidence, prevalence and long-term sequelae in Australia, 2001. International Journal of Epidemiology 2003; 32(5):717-724.

(71) Zou S, Tepper M, El Saadany S. Prediction of hepatitis C burden in Canada. Canadian Journal of Gastroenterology. 14[(7)], 575-580. 2000.

(72) Deuffic S, Buffat L, Poynard T, Valleron A. Modeling the Hepatitis C epidemic in France. Hepatology 1999; 29:1598-1601.

(73) Heathcote J, Main J. Treatment of hepatitis C J Viral Hepat 2005; 12(3):223-235.

(74) Di Bisceglie AM. Natural history of hepatitis C: its impact on clinical management 99. Hepatology 2000; 31(4):1014-1018.

(75) Poynard T, Bedossa P, Opolon P, for the OBSVIRC, DOSVIRC groups. Natural history of liver fibrosis progression in patients with chronic hepatitis C. Lancet 1997; 349(March 22):825-832.

(76) Wright M, Grieve R, Roberts J, Main J, Thomas HC. Health benefits of antiviral therapy for mild chronic hepatitis C: randomised controlled trial and economic evaluation Health Technol Assess 2006; 10(21):1-113, iii.

(77) Grieve R, Roberts J, Wright M, Sweeting M, DeAngelis D, Rosenberg W et al. Cost effectiveness of interferon alpha or peginterferon alpha with ribavirin for histologically mild chronic hepatitis C Gut 2006; 55(9):1332-1338.

(78) National Institute for Clinical Excellence. Interferon alfa (pegylated and non-pegylated) and ribavirin for the treatment of chronic hepatitis C. Technology Appraisal Guidance 75. 2004.

(79) Hui JM, Hodge A, Farrell GC, Kench JG, Kriketos A, George J. Beyond insulin resistance in NASH: TNF-alpha or adiponectin? Hepatology 2004; 40(1):46-54.

(80) Chitturi S, Farrell G, Frost L, Kriketos A, Lin R, Fung C et al. Serum leptin in NASH correlates with hepatic steatosis but not fibrosis: a manifestation of lipotoxicity? Hepatology 2002; 36(2):403-409.

(81) Day CP. Non-alcoholic fatty liver disease. In: McDonald JW, Burroughs A, Feagan BG, editors. Evidence Based Gastroenterology and Hepatology. 2 ed. Blackwell; 2004.

(82) Day CP. Natural history of NAFLD: remarkably benign in the absence of cirrhosis Gastroenterology 2005; 129(1):375-378.

(83) Powell EE, Cooksley WG, Hanson R, Searle J, Halliday JW, Powell LW. The natural history of nonalcoholic steatohepatitis: a follow-up study of forty-two patients for up to 21 years Hepatology 1990; 11(1):74-80.

(84) Harrison SA, Torgerson S, Hayashi PH. The natural history of nonalcoholic fatty liver disease: a clinical histopathological study Am J Gastroenterol 2003; 98(9):2042-2047.

(85) Sanyal AJ, Banas C, Sargeant C, Luketic VA, Sterling RK, Stravitz RT et al. Similarities and differences in outcomes of cirrhosis due to nonalcoholic steatohepatitis and hepatitis C Hepatology 2006; 43(4):682-689.

(86) Ioannou GN, Weiss NS, Boyko EJ, Mozaffarian D, Lee SP. Elevated serum alanine aminotransferase activity and calculated risk of coronary heart disease in the United States Hepatology 2006; 43(5):1145-1151.

(87) Gebo KA, Herlong HF, Torbenson MS, Jenckes MW, Chander G, Ghanem KG et al. Role of liver biopsy in management of chronic hepatitis C: a systematic review. Hepatology 2002; 36(5 Suppl 1):S161-S172.

(88) Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness. CRD Report No 4. 2001. University of York.

(89) Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests BMJ 2001; 323(7305):157-162.

(90) Haynes RB, Wilczynski NL. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey BMJ 2004; 328(7447):1040.

(91) Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies Health Technol Assess 2004; 8(25):iii, 1-iii234.

(92) Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations Stat Med 1993; 12(14):1293-1316.

(93) Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method Med Decis Making 1993; 13(4):313-321.

(94) Smith-Bindman R, Kerlikowske K, Feldstein VA, Subak L, Scheidler J, Segal M et al. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities JAMA 1998; 280(17):1510-1517.

(95) Dukic V, Gatsonis C. Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds Biometrics 2003; 59(4):936-946.

(96) Poynard T, Imbert-Bismut F, Munteanu M, Messous D, Myers RP, Thabut D et al. Overview of the diagnostic value of biochemical markers of liver fibrosis (FibroTest, HCV FibroSure) and necrosis (ActiTest) in patients with chronic hepatitis C 21. Comp Hepatol 2004; 3(1):8.

(97) Imbert-Bismut F, Ratziu V, Pieroni L, Charlotte F, Benhamou Y, Poynard T. Biochemical markers of liver fibrosis in patients with hepatitis C virus infection: a prospective study Lancet 2001; 357(9262):1069-1075.

(98) Rossi E, Adams L, Prins A, Bulsara M, de Boer B, Garas G et al. Validation of the FibroTest biochemical markers score in assessing liver fibrosis in hepatitis C patients. Clinical Chemistry 2003; 49(3):450-454.

(99) Poynard T, McHutchison J, Manns M, Myers RP, Albrecht J. Biochemical surrogate markers of liver fibrosis and activity in a randomized trial of peginterferon alfa-2b and ribavirin. Hepatology 2003; 38(2):481-492.

(100) Wai CT, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C.[see comment]. Hepatology 2003; 38(2):518-526.

(101) Le CS, Thabut D, Messous D, Munteanu M, Ratziu V, Imbert-Bismut F et al. The predictive value of Fibrotest vs. APRI for the diagnosis of fibrosis in chronic hepatitis C Hepatology 2004; 39(3):862-863.

(102) Forns X, Ampurdanes S, Llovet JM, Aponte J, Quinto L, Martinez-Bauer E et al. Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. Hepatology 2002; 36(4:Pt 1):t-92.

(103) Thabut D, Simon M, Myers RP, Messous D, Thibault V, Imbert-Bismut F et al. Noninvasive prediction of fibrosis in patients with chronic hepatitis C [1] (multiple letters). Hepatology 2003;. 37(5):01.

(104) Sud A, Hui JM, Farrell GC, Bandara P, Kench JG, Fung C et al. Improved prediction of fibrosis in chronic hepatitis C using measures of insulin resistance in a probability index. Hepatology 2004; 39(5):1239-1247.

(105) Leroy V, Monier F, Bottari S, Trocme C, Sturm N, Hilleret MN et al. Circulating matrix metalloproteinases 1, 2, 9 and their inhibitors TIMP-1 and TIMP-2 as serum markers of liver fibrosis in patients with chronic hepatitis C: comparison with PIIINP and hyaluronic acid. American Journal of Gastroenterology 2004; 99(2):271-279.

(106) el Shorbagy E, Afefy AF, Ibrahem IA, Mangoud AM, Eissa MH, Sabee El et al. Non-invasive markers and predictors of severity of hepatic fibrosis in HCV patients at Sharkia Governorate, Egypt. Journal of the Egyptian Society of Parasitology 2004; 34(1:Suppl):Suppl-78.

(107) Patel K, Gordon SC, Jacobson I, Hezode C, Oh E, Smith KM et al. Evaluation of a panel of non-invasive serum markers to differentiate mild from moderate-to-advanced liver fibrosis in chronic hepatitis C patients J Hepatol 2004; 41(6):935-942.

(108) Kaul V, Friedenberg FK, Braitman LE, Anis U, Zaeri N, Fazili J et al. Development and validation of a model to diagnose cirrhosis in patients with hepatitis C. American Journal of Gastroenterology 2002; 97(10):2623-2628.

(109) Fortunato G, Castaldo G, Oriani G, Cerini R, Intrieri M, Molinaro E et al. Multivariate discriminant function based on six biochemical markers in blood can predict the cirrhotic evolution of chronic hepatitis Clin Chem 2001; 47(9):1696-1700.

(110) Sakugawa H, Nakayoshi T, Kobashigawa K, Yamashiro T, Maeshiro T, Miyagi S et al. Clinical usefulness of biochemical markers of liver fibrosis in patients with nonalcoholic fatty liver disease World J Gastroenterol 2005; 11(2):255-259.

(111) Albano E, Mottaran E, Vidali M, Reale E, Saksena S, Occhino G et al. Immune response towards lipid peroxidation products as a predictor of progression of non-alcoholic fatty liver disease to advanced fibrosis GUT 2005; 54(7):987-993.

(112) Mofrad P, Contos MJ, Haque M, Sargeant C, Fisher RA, Luketic VA et al. Clinical and histologic spectrum of nonalcoholic fatty liver disease associated with normal ALT values Hepatology 2003; 37(6):1286-1292.

(113) Shimada M, Hashimoto E, Kaneda H, Noguchi S, Hayashi N. Nonalcoholic steatohepatitis: risk factors for liver fibrosis Hepatol Res 2002; 24(4):429-438.

(114) Dixon JB, Bhathal PS, O'Brien PE. Nonalcoholic fatty liver disease: predictors of nonalcoholic steatohepatitis and liver fibrosis in the severely obese Gastroenterology 2001; 121(1):91-100.

(115) Beymer C, Kowdley KV, Larson A, Edmonson P, Dellinger EP, Flum DR. Prevalence and predictors of asymptomatic liver disease in patients undergoing gastric bypass surgery Arch Surg 2003; 138(11):1240-1244.

(116) Bugianesi E, Manzini P, D'Antico S, Vanni E, Longo F, Leone N et al. Relative contribution of iron burden, HFE mutations, and insulin resistance to fibrosis in nonalcoholic fatty liver Hepatology 2004; 39(1):179-187.

(117) Dixon JB, Bhathal PS, Jonsson JR, Dixon AF, Powell EE, O'Brien PE. Pro-fibrotic polymorphisms predictive of advanced liver fibrosis in the severely obese 4. J Hepatol 2003; 39(6):967-971.

(118) Guidorizzi de Siqueira AC, Cotrim HP, Rocha R, Carvalho FM, de Freitas LA, Barreto D et al. Non-alcoholic fatty liver disease and insulin resistance: importance of risk factors and histological spectrum Eur J Gastroenterol Hepatol 2005; 17(8):837-841.

(119) Suzuki A, Angulo P, Lymp J, Li D, Satomura S, Lindor K. Hyaluronic acid, an accurate serum marker for severe hepatic fibrosis in patients with non-alcoholic fatty liver disease Liver Int 2005; 25(4):779-786.

(120) Angulo P, Alba LM, Petrovic LM, Adams LA, Lindor KD, Jensen MD. Leptin, insulin resistance, and liver fibrosis in human nonalcoholic fatty liver disease J Hepatol 2004; 41(6):943-949.

(121) Marchesini G, Bugianesi E, Forlani G, Cerrelli F, Lenzi M, Manini R et al. Nonalcoholic fatty liver, steato-hepatitis and the metabolic syndrome. Hepatology 2003; 37:917-923.

(122) Hashimoto E, Yatsuji S, Kaneda H, Yoshioka Y, Taniai M, Tokushige K et al. The characteristics and natural history of Japanese patients with nonalcoholic fatty liver disease Hepatol Res 2005; 33(2):72-76.

(123) Ong JP, Elariny H, Collantes R, Younoszai A, Chandhoke V, Reines HD et al. Predictors of nonalcoholic steatohepatitis and advanced fibrosis in morbidly obese patients Obes Surg 2005; 15(3):310-315.

(124) de L, V, Combes M, Trouette H, Winnock M, Amouretti M, de MA et al. Should a liver biopsy be done in patients with subclinical chronically elevated transaminases? Eur J Gastroenterol Hepatol 2004; 16(9):879-883.

(125) Ratziu V, Giral P, Charlotte F, Bruckert E, Thibault V, Theodorou I et al. Liver fibrosis in overweight patients Gastroenterology 2000; 118(6):1117-1123.

(126) Sorrentino P, Tarantino G, Conca P, Perrella A, Terracciano ML, Vecchione R et al. Silent non-alcoholic fatty liver disease-a clinical-histological study J Hepatol 2004; 41(5):751-757.

(127) Crespo J, Fernandez-Gil P, Hernandez-Guerra M, Cayon A, Mayorga M, Dominguez-Diez A et al. Are there predictive factors of severe liver fibrosis in morbidly obese patients with non-alcoholic steatohepatitis? Obes Surg 2001; 11(3):254-257.

(128) Fierbinteanu-Braticevici C, Bengus A, Neamtu M, Usvat R. The risk factors of fibrosis in nonalcoholic steatohepatitis Rom J Intern Med 2002; 40(1-4):81-88.

(129) Loguercio C, De ST, D'Auria MV, De S, I, Federico A, Tuccillo C et al. Non-alcoholic fatty liver disease: a multicentre clinical study by the Italian Association for the Study of the Liver Dig Liver Dis 2004; 36(6):398-405.

(130) Santos VN, Leite-Mor MM, Kondo M, Martins JR, Nader H, Lanzoni VP et al. Serum laminin, type IV collagen and hyaluronan as fibrosis markers in non-alcoholic fatty liver disease Braz J Med Biol Res 2005; 38(5):747-753.

(131) Yesilova Z, Yaman H, Oktenli C, Ozcan A, Uygun A, Cakir E et al. Systemic markers of lipid peroxidation and antioxidants in patients with nonalcoholic Fatty liver disease Am J Gastroenterol 2005; 100(4):850-855.

(132) Koruk M, Savas MC, Yilmaz O, Taysi S, Karakok M, Gundogdu C et al. Serum lipids, lipoproteins and apolipoproteins levels in patients with nonalcoholic steatohepatitis J Clin Gastroenterol 2003; 37(2):177-182.

(133) Hartleb M, Kajor M, Gupinska-Kajor M, Nowak A, Rudzki K, Waluga M et al. Non-alcoholic fatty liver disease; who does not benefit from hepatic biopsy? Gastroenterologica Polska 12[3], 209-213. 2005.

(134) Brunt EM, Neuschwander-Tetri BA, Oliver D, Wehmeier KR, Bacon BR. Nonalcoholic steatohepatitis: histologic features and clinical correlations with 30 blinded biopsy specimens Hum Pathol 2004; 35(9):1070-1082.

(135) Paradis V, Perlemuter G, Bonvoust F, Dargere D, Parfait B, Vidaud M et al. High glucose and hyperinsulinemia stimulate connective tissue growth factor expression: a potential mechanism involved in progression to fibrosis in nonalcoholic steatohepatitis Hepatology 2001; 34(4 Pt 1):738-744.

(136) Samuel VT, Liu ZX, Qu X, Elder BD, Bilz S, Befroy D et al. Mechanism of hepatic insulin resistance in non-alcoholic fatty liver disease J Biol Chem 2004; 279(31):32345-32353.

(137) Gabrielli GB, Faccioli G, Casaril M, Capra F, Bonazzi L, Falezza G et al. Procollagen III peptide and fibronectin in alcohol-related chronic liver disease: correlations with morphological features and biochemical tests Clin Chim Acta 1989; 179(3):315-322.

(138) Poynard T, Aubert A, Bedossa P, Abella A, Naveau S, Paraf F et al. A simple biological index for detection of alcoholic liver disease in drinkers Gastroenterology 1991; 100(5 Pt 1):1397-1402.

(139) Li J, Rosman AS, Leo MA, Nagai Y, Lieber CS. Tissue inhibitor of metalloproteinase is increased in the serum of precirrhotic and cirrhotic alcoholic patients and can serve as a marker of fibrosis Hepatology 1994; 19(6):1418-1423.

(140) Oberti F, Valsesia E, Pilette C, Rousselet MC, Bedossa P, Aube C et al. Noninvasive diagnosis of hepatic fibrosis or cirrhosis Gastroenterology 1997; 113(5):1609-1616.

(141) Tran A, Hastier P, Barjoan EM, Demuth N, Pradier C, Saint-Paul MC et al. Non invasive prediction of severe fibrosis in patients with alcoholic liver disease. Gastroenterologie Clinique et Biologique 2000; 24(6-7):626-630.

(142) Plevris JN, Haydon GH, Simpson KJ, Dawkes R, Ludlum CA, Harrison DJ et al. Serum hyaluronan--a non-invasive test for diagnosing liver cirrhosis Eur J Gastroenterol Hepatol 2000; 12(10):1121-1127.

(143) Croquet V, Vuillemin E, Ternisien C, Pilette C, Oberti F, Gallois Y et al. Prothrombin index is an indirect marker of severe liver fibrosis. European Journal of Gastroenterology & Hepatology 2002; 14(10):1133-1141.

(144) Stickel F, Poeschl G, Schuppan D, Conradt C, Strenge-Hesse A, Fuchs FS et al. Serum hyaluronate correlates with histological progression in alcoholic liver disease. European Journal of Gastroenterology & Hepatology 2003; 15(9):01.

(145) Naveau S, Raynard B, Ratziu V, Abella A, Imbert-Bismut F, Messous D et al. Biomarkers for the prediction of liver fibrosis in patients with chronic alcoholic liver disease Clin Gastroenterol Hepatol 2005; 3(2):167-174.

(146) Cales P, Oberti F, Michalak S, Hubert-Fouchard I, Rousselet MC, Konate A et al. A novel panel of blood markers to assess the degree of liver fibrosis Hepatology 2005; 42(6):1373-1381.

(147) Lieber CS, Weiss DG, Morgan TR, Paronetto F. Aspartate aminotransferase to platelet ratio index in patients with alcoholic liver fibrosis Am J Gastroenterol 2006; 101(7):1500-1508.

(148) Guha IN, Parkes J, Roderick P, Chattopadhyay D, Cross R, Harris S et al. Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: Validating the European Liver Fibrosis Panel and exploring simple markers Hepatology 2008; 47(2):455-460.

(149) Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve Radiology 1982; 143(1):29-36.

(150) Habbema JD, Eijkemans R, Krijnen P, Knottnerus JA. Analysis of data on the accuracy of diagnostic tests. The evidence base of clinical diagnosis. 2002. 117.

(151) Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003; 138(1):W1-12.

(152) Bossuyt P. The quality of reporting in diagnsotic tests is getting better, still not optimal. Clin Chem. 50, 465-466. 2004.

(153) Buck AA, Gart JJ. Comparison of a screening test and a reference test in epidemiologic studies. I. Indices of agreement and their relation to prevalence Am J Epidemiol 1966; 83(3):586-592.

(154) Cholongitas E, Senzolo M, Standish R, Marelli L, Quaglia A, Patch D et al. A systematic review of the quality of liver biopsy specimens Am J Clin Pathol 2006; 125(5):710-721.

(155) Poynard T, Munteanu M, Imbert-Bismut F, Charlotte F, Thabut D, Le CS et al. Prospective analysis of discordant results between biochemical markers and biopsy in patients with chronic hepatitis C Clin Chem 2004; 50(8):1344-1355.

(156) Poynard T, Morra R, Halfon P, Castera L, Ratziu V, Imbert-Bismut F et al. Meta-analyses of FibroTest diagnostic value in CLD BMC Gastroenterol 2007; 7:40.

(157) Shaheen AA, Wan AF, Myers RP. FibroTest and FibroScan for the prediction of hepatitis C-related fibrosis: a systematic review of diagnostic test accuracy Am J Gastroenterol 2007; 102(11):2589-2600.

(158) Shaheen AA, Myers RP. Diagnostic accuracy of the aspartate aminotransferase-to-platelet ratio index for the prediction of hepatitis C-related fibrosis: a systematic review Hepatology 2007; 46(3):912-921.

(159) Shaheen AA, Myers RP. Systematic review and meta-analysis of the diagnostic accuracy of fibrosis marker panels in patients with HIV/hepatitis C coinfection HIV Clin Trials 2008; 9(1):43-51.

(160) Castera L, Vergniol J, Foucher J, Le BB, Chanteloup E, Haaser M et al. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C Gastroenterology 2005; 128(2):343-350.

(161) Rockey DC. Noninvasive assessment of liver fibrosis and portal hypertension with transient elastography Gastroenterology 2008; 134(1):8-14.

(162) Roulot D, Czernichow S., Le Clesiau H., Costes J-L., Vergnaud A-C., Beaugrand M. Liver stiffness values in apparently health subjects: Influence of gender and metabolic syndrome. Journal of Hepatology in press. 2008.

(163) Arena U, Vizzutti F, Corti G, Ambu S, Stasi C, Bresci S et al. Acute viral hepatitis increases liver stiffness values measured by transient elastography 2. Hepatology 2008; 47(2):380-384.

(164) Sebastiani G, Vario A, Guido M, Alberti A. Sequential algorithms combining non-invasive markers and biopsy for the assessment of liver fibrosis in chronic hepatitis B World J Gastroenterol 2007; 13(4):525-531.

(165) Callewaert N, Van VH, Van HA, Laroy W, Delanghe J, Contreras R. Noninvasive diagnosis of liver cirrhosis using DNA sequencer-based total serum protein glycomics Nat Med 2004; 10(4):429-434.

(166) Combes B, Emerson SS, Flye NL, Munoz SJ, Luketic VA, Mayo MJ et al. Methotrexate (MTX) plus ursodeoxycholic acid (UDCA) in the treatment of primary biliary cirrhosis Hepatology 2005; 42(5):1184-1193.

(167) Chung RT, Andersen J, Volberding P, Robbins GK, Liu T, Sherman KE et al. Peginterferon Alfa-2a plus ribavirin versus interferon alfa-2a plus ribavirin for chronic hepatitis C in HIV-coinfected persons N Engl J Med 2004; 351(5):451-459.

(168) Halfon P, Bacq Y, De MA, Penaranda G, Bourliere M, Ouzan D et al. Comparison of test performance profile for blood tests of liver fibrosis in chronic hepatitis C J Hepatol 2007; 46(3):395-402.

(169) Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection Hepatology 2006; 43(6):1317-1325.

(170) Kelleher TB, Mehta SH, Bhaskar R, Sulkowski M, Astemborski J, Thomas DL et al. Prediction of hepatic fibrosis in HIV/HCV co-infected patients using serum fibrosis markers: the SHASTA index J Hepatol 2005; 43(1):78-84.

(171) Independent assessment and cost-performance analysis of noninvasive liver fibrosis biomarkers in HIV-HCV coinfected patients: the Fibrovic Study-ANRS HC02. 06 Nov AASLD: 2006.

(172) Nyblom H, Bjornsson E, Simren M, Aldenborg F, Almer S, Olsson R. The AST/ALT ratio as an indicator of cirrhosis in patients with PBC Liver Int 2006; 26(7):840-845.

(173) Corpechot C, Poujol-Robert A, Wendum D, Galotte M, Chretien Y, Poupon RE et al. Biochemical markers of liver fibrosis and lymphocytic piecemeal necrosis in UDCA-treated patients with primary biliary cirrhosis Liver Int 2004; 24(3):187-193.

(174) Plebani M, Giacomini A, Floreani A, Chiaramonte M, Soffiati G, Naccarato R et al. Biochemical markers of hepatic fibrosis in primary biliary cirrhosis Ric Clin Lab 1990; 20(4):269-274.

(175) Nyberg A, Lindqvist U, Engstrom-Laurent A. Serum hyaluronan and aminoterminal propeptide of type III procollagen in primary biliary cirrhosis: relation to clinical symptoms, liver histopathology and outcome J Intern Med 1992; 231(5):485-491.

(176) Remmel T, Remmel H, Salupere V. Aminoterminal propeptide of type III procollagen and hyaluronan in patients with primary biliary cirrhosis: markers of fibrosis in primary biliary cirrhosis J Gastroenterol Hepatol 1996; 11(11):1016-1020.

(177) Gomez-Dominguez E, Mendoza J, Garcia-Buey L, Trapero M, Gisbert JP, Jones EA et al. Transient elastography to assess hepatic fibrosis in primary biliary cirrhosis Aliment Pharmacol Ther 2008; 27(5):441-447.

(178) Corpechot C, El NA, Poujol-Robert A, Ziol M, Wendum D, Chazouilleres O et al. Assessment of biliary fibrosis by transient elastography in patients with PBC and PSC Hepatology 2006; 43(5):1118-1124.

(179) Castera L, Denis J, Babany G, Roudot-Thoraval F. Evolving practices of non-invasive markers of liver fibrosis in patients with chronic hepatitis C in France: time for new guidelines? J Hepatol 2007; 46(3):528-529.

(180) Sackett DL, Haynes RB, Guyatt G, Tugwell P. Making a prognosis. Clinical Epidemiology A basic science for clinical medicine. 2 ed. Little, Brown and Company; 1991. 173-185.

(181) Christensen E. Prognostic models including the Child-Pugh, MELD and Mayo risk scores - where are we and where should we go? Journal of Hepatology 41, 344-350. 2006.

(182) Pugh RNH, Murray-Lyon IM, Dawson JL, Pietroni MC, Williams R. Transection of the Oesophagus for Bleeding Oesophageal Varices. Brit J Surg 1973; 60(8):646-649.

(183) Kamath PS, Wiesner RH, Malinchoc M, Kremers W, Therneau TM, Kosberg CL et al. A Model to Predict Survival in Patients with End-Stage Liver Disease. Hepatology 2001; 33:464-470.

(184) Klion FM, Fabry TL, Palmer M, Schaffner F. Prediction of survival of patients with primary biliary cirrhosis. Examination of the Mayo Clinic model on a group of patients with known endpoint Gastroenterology 1992; 102(1):310-313.

(185) Korner T, Kropf J, Gressner AM. Serum laminin and hyaluronan in liver cirrhosis: markers of progression with high prognostic value. Journal of Hepatology 1996; 25(5):684-688.

(186) Guechot J, Serfaty L, Bonnand AM, Chazouilleres O, Poupon RE, Poupon R. Prognostic value of serum hyaluronan in patients with compensated HCV cirrhosis. Journal of Hepatology 2000; 32(3):447-452.

(187) Ngo Y, Munteanu M, Messous D, Charlotte F, Imbert-Bismut F, Thabut 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.

(188) David Collett. Modelling survival data in medical research. 2 ed. Chapman & Hall; 2003.

(189) Christensen E. Prognostic modeling in Primary Biliary Cirrhosis. In: Neuberger J, editor. Primary Biliary Cirrhosis. West End Studios Ltd; 1999. 93-99.

(190) Poupon RE, Balkau B, Guechot J, Heintzmann F. Predictive factors in ursodeoxycholic acid-treated patients with primary biliary cirrhosis: role of serum markers of connective tissue Hepatology 1994; 19(3):635-640.

(191) Castera L, Vergniol J, Foucher J, Le BB, Chanteloup E, Haaser M et al. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C Gastroenterology 2005; 128(2):343-350.

(192) Altman DG. Systematic reviews of evaluations of prognostic variables. BMJ 2001; 323(28 July):224-228.

(193) Manos MM, Leyden WA, Murphy RC, Terrault NA, Bell BP. Limitations of conventionally derived chronic liver disease mortality rates: Results of a comprehensive assessment Hepatology 2008.

(194) Goldacre MJ. Cause-specific mortality: understanding uncertain tips of the disease iceberg J Epidemiol Community Health 1993; 47(6):491-496.

(195) Myers KA, Farquhar DR. Improving the accuracy of death certification CMAJ 1998; 158(10):1317-1323.

(196) Maxwell JD. Accuracy of Death Certification for Alcoholic Liver Disease. British Journal of Addiction 1986; 81:168-169.

(197) Johansson LA, Westerling R, Rosenberg HM. Methodology of studies evaluating death certificate accuracy were flawed J Clin Epidemiol 2006; 59(2):125-131.

(198) Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation Ann Intern Med 2002; 137(7):598-602.

(199) Goehring C, Perrier A, Morabia A. Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance Stat Med 2004; 23(1):125-135.

(200) Coggon D, Martyn C, Palmer KT, Evanoff B. Assessing case definitions in the absence of a diagnostic gold standard Int J Epidemiol 2005; 34(4):949-952.

(201) Pepe MS, Longton G. Standardizing diagnostic markers to evaluate and compare their performance Epidemiology 2005; 16(5):598-603.

(202) Biesheuvel C, Irwig L, Bossuyt P. Observed differences in diagnostic test accuracy between patient subgroups: is it real or due to reference standard misclassification? Clin Chem 2007; 53(10):1725-1729.

(203) Hawkins DM, Garrett JA, Stephenson B. Some issues in resolution of diagnostic tests using an imperfect gold standard Stat Med 2001; 20(13):1987-2001.

(204) Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods Health Technol Assess 2007; 11(50):1-72.

(205) Rosenberg WM. Rating fibrosis progression in chronic liver diseases J Hepatol 2003; 38(3):357-360.

(206) Malik AH, Kumar KS, Malet PF, Jain R, Prasad P, Ostapowicz G. Correlation of percutaneous liver biopsy fragmentation with the degree of fibrosis Aliment Pharmacol Ther 2004; 19(5):545-549.

(207) Fabre A, Main J, Thomas H, Foster G, Goldin R. Diagnostic value of liver biopsy in patients with chronic hepatitis C infection. J Hepatol 2002; 36:106.

(208) Neal KR, Ramsay S, Thomson BJ, Irving WL. Excess mortality rates in a cohort of patients infected with the hepatitis C virus: a prospective study GUT 2007; 56(8):1098-1104.

(209) O'Brien MJ, Keating NM, Elderiny S, Cerda S, Keaveny AP, Afdhal NH et al. An assessment of digital image analysis to measure fibrosis in liver biopsy specimens of patients with chronic hepatitis C Am J Clin Pathol 2000; 114(5):712-718.

(210) Wright M, Thursz M, Pullen R, Thomas H, Goldin R. Quantitative versus morphological assessment of liver fibrosis: semi-quantitative scores are more robust than digital image fibrosis area estimation Liver Int 2003; 23(1):28-34.

(211) Ponomarenko Y, Leo MA, Kroll W, Lieber CS. Effects of alcohol consumption on eight circulating markers of liver fibrosis. Alcohol & Alcoholism 2002; 37(3):252-255.

(212) Lindqvist U, Engstrom-Laurent A, Laurent U, Nyberg A, Bjorklund U, Eriksson H et al. The diurnal variation of serum hyaluronan in health and disease Scand J Clin Lab Invest 1988; 48(8):765-770.

(213) Kong SY, Stabler TV, Criscione LG, Elliott AL, Jordan JM, Kraus VB. Diurnal variation of serum and urine biomarkers in patients with radiographic knee osteoarthritis Arthritis Rheum 2006; 54(8):2496-2504.

(214) Fracanzani AL, Burdick L, Raselli S, Pedotti P, Grigore L, Santorelli G et al. Carotid artery intima-media thickness in nonalcoholic fatty liver disease Am J Med 2008; 121(1):72-78.

(215) Honest H, Khan KS. Reporting of measures of accuracy in systematic reviews of diagnostic literature BMC Health Serv Res 2002; 2(1):4.

(216) Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy Am J Med 1998; 104(4):374-380.

(217) Ransohoff DF. Challenges and opportunities in evaluating diagnostic tests 35. J Clin Epidemiol 2002; 55(12):1178-1182.

(218) Iacobellis A, Mangia A, Leandro G, Clemente R, Festa V, Attino V et al. External validation of biochemical indices for noninvasive evaluation of liver fibrosis in HCV chronic hepatitis Am J Gastroenterol 2005; 100(4):868-873.

(219) Lackner C, Struber G, Liegl B, Leibl S, Ofner P, Bankuti C et al. Comparison and validation of simple noninvasive tests for prediction of fibrosis in chronic hepatitis C Hepatology 2005; 41(6):1376-1382.

(220) Bourliere M, Penaranda G, Renou C, Botta-Fridlund D, Tran A, Portal I et al. Validation and comparison of indexes for fibrosis and cirrhosis prediction in chronic hepatitis C patients: proposal for a pragmatic approach classification without liver biopsies J Viral Hepat 2006; 13(10):659-670.

(221) Leroy V, Hilleret MN, Sturm N, Trocme C, Renversez JC, Faure P et al. Prospective comparison of six non-invasive scores for the diagnosis of liver fibrosis in chronic hepatitis C. J Hepatol 2007; 46(5):775-782.

(222) Parise ER, Oliveira AC, Figueiredo-Mendes C, Lanzoni V, Martins J, Nader H et al. Noninvasive serum markers in the diagnosis of structural liver damage in chronic hepatitis C virus infection. Liver Int 2006; 26(9):1095-1099.