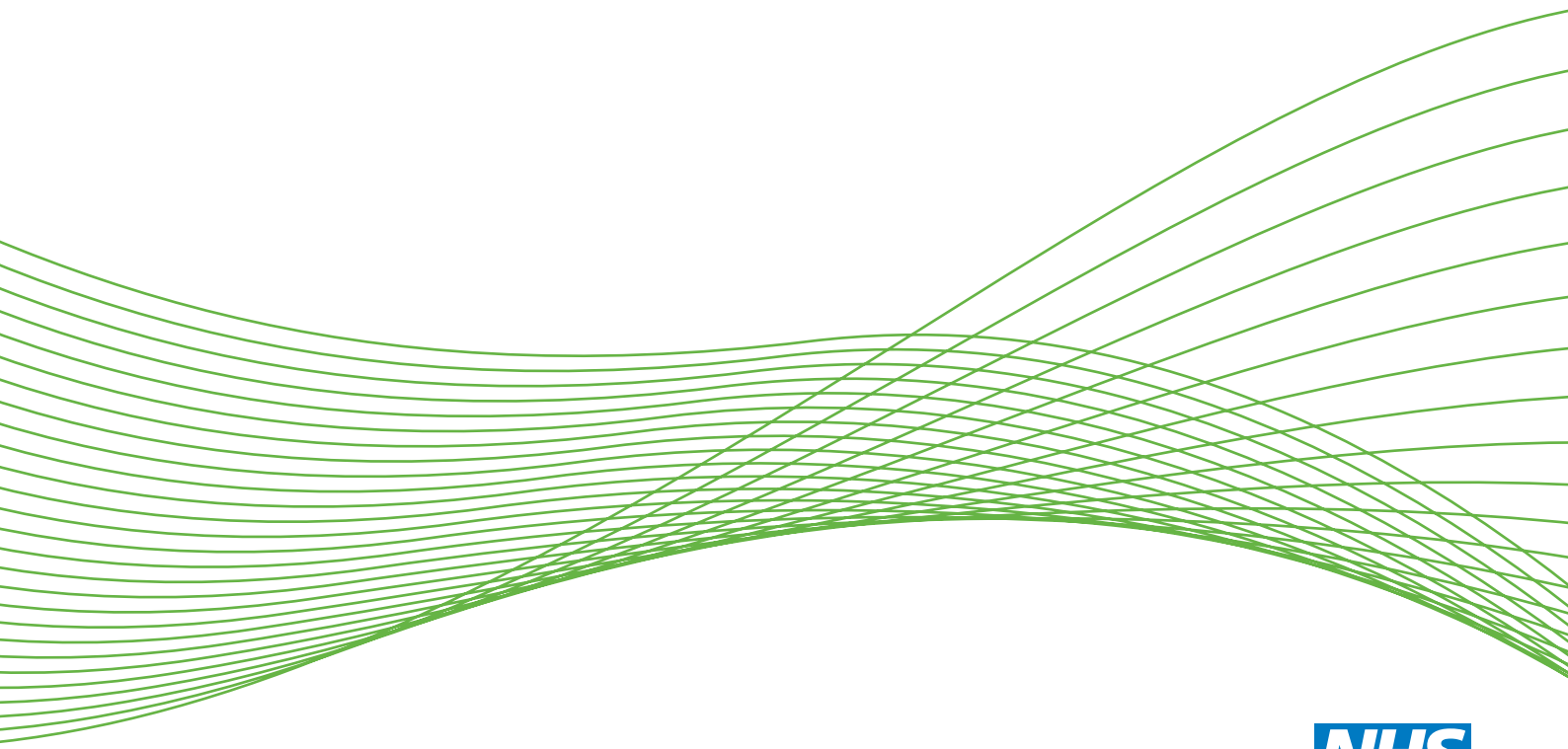
HEALTH TECHNOLOGY ASSESSMENT

VOLUME 17 ISSUE 28 JULY 2013
ISSN 1366-5278

Birmingham and Lambeth Liver Evaluation Testing Strategies (BALLETS): a prospective cohort study

RJ Lilford, L Bentham, A Girling, I Litchfield, R Lancashire, D Armstrong, R Jones, T Marteau, J Neuberger, P Gill, R Cramb, S Olliff, D Arnold, K Khan, MJ Armstrong, DD Houlihan, PN Newsome, PJ Chilton, K Moons and D Altman



NHS
National Institute for Health Research

DOI 10.3310/hta17280

Birmingham and Lambeth Liver Evaluation Testing Strategies (BALLETS): a prospective cohort study

RJ Lilford,[1]* L Bentham,[1] A Girling,[1] I Litchfield,[1] R Lancashire,[1] D Armstrong,[2] R Jones,[2] T Marteau,[3] J Neuberger,[1] P Gill,[1] R Cramb,[4] S Olliff,[4] D Arnold,[5] K Khan,[4] MJ Armstrong,[6] DD Houlihan,[6] PN Newsome,[6] PJ Chilton,[1] K Moons[7] and D Altman[8]

[1]School of Health and Population Sciences, University of Birmingham, Edgbaston, UK
[2]Department of Primary Care and Public Health Sciences, Kings College London, London, UK
[3]Health Psychology Section, Kings College London, London, UK
[4]Queen Elizabeth Hospital Birmingham, Birmingham, UK
[5]School of Medicine, Cardiff University, Cardiff, UK
[6]National Institute for Health Research Biomedical Research Unit, Birmingham, UK
[7]Universitair Medisch Centrum Utrecht, Utrecht, the Netherlands
[8]Centre for Statistics in Medicine, University of Oxford, Oxford, UK

*Corresponding author

Declared competing interests of authors: none

Published July 2013
DOI: 10.3310/hta17280

This report should be referenced as follows:

Lilford RJ, Bentham L, Girling A, Litchfield I, Lancashire R, Armstrong D, *et al*. Birmingham and Lambeth Liver Evaluation Testing Strategies (BALLETS): a prospective cohort study. *Health Technol Assess* 2013;**17**(28).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (*SciSearch*®) and *Current Contents*®/Clinical Medicine.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Five-year impact factor: 5.804

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 03/38/01. The contractual start date was in July 2005. The draft report began editorial review in October 2011 and was accepted for publication in November 2011. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library

NIHR Journals Library Editors

Abstract

Birmingham and Lambeth Liver Evaluation Testing Strategies (BALLETS): a prospective cohort study

RJ Lilford,[1]* L Bentham,[1] A Girling,[1] I Litchfield,[1] R Lancashire,[1] D Armstrong,[2] R Jones,[2] T Marteau,[3] J Neuberger,[1] P Gill,[1] R Cramb,[4] S Olliff,[4] D Arnold,[5] K Khan,[4] MJ Armstrong,[6] DD Houlihan,[6] PN Newsome,[6] PJ Chilton,[1] K Moons[7] and D Altman[8]

[1]School of Health and Population Sciences, University of Birmingham, Edgbaston, UK
[2]Department of Primary Care and Public Health Sciences, Kings College London, London, UK
[3]Health Psychology Section, Kings College London, London, UK
[4]Queen Elizabeth Hospital Birmingham, Birmingham, UK
[5]School of Medicine, Cardiff University, Cardiff, UK
[6]National Institute for Health Research, Biomedical Research Unit, Birmingham, UK
[7]Universitair Medisch Centrum Utrecht, Utrecht, the Netherlands
[8]Centre for Statistics in Medicine, University of Oxford, Oxford, UK

*Corresponding author

Objective: To evaluate mildly abnormal liver function test (LFT) results in general practice among patients who do not have known liver disease.
Design: Prospective cohort study of people with abnormal LFT results identified in primary care. Participants were intensively investigated using a common protocol and followed up for 2 years. Substudies investigated the psychological sequelae of abnormal test results, clinicians' reasons for testing, decision options when LFT results were abnormal and early detection of liver fibrosis.
Setting: Eleven primary-care practices: eight in Birmingham and three in Lambeth.
Participants: Adults with abnormal LFT results who did not have pre-existing or obvious liver disease. Eight analytes were included in the panel of LFTs.
Main outcome measures: Statistical tests were used to identify the interactions between clinical features, the initial pattern of abnormal LFT results and (1) specific viral, genetic and autoimmune diseases, such as viral hepatitis, haemochromatosis and primary biliary cirrhosis; (2) a range of other serious diseases, such as metastatic cancer and hypothyroidism; (3) 'fatty liver' not associated with the above; and (4) the absence of detectable disease.
Results: Fewer than 5% of people with abnormal LFT results had a specific disease of the liver, and many of these were unlikely to need treatment. The diagnostic potential of the LFT panel is largely subsumed into just two analytes: alanine aminotransferase (ALT) and alkaline phosphatase (ALP). Gamma-glutamyltransferase (GGT) offers a small increase in sensitivity at the margin at the cost of a large loss of specificity. Eighty-four per cent of abnormal LFT results remain abnormal on retesting 1 month later. In many cases, carrying out a definitive or specific test will be more efficient than repeating LFTs, with a view to specific testing only if the test remains abnormal. An ultrasound diagnosis of 'fatty liver' was present in nearly 40% of patients with abnormal LFTs and a small amount of weight

loss over 2 years was associated with a reduced incidence of liver fat. There was a J-shaped relationship between alcohol intake and fatty liver in men. An abnormal LFT result causes temporary anxiety, which does not appear to promote sustained behaviour change.

Conclusions: Liver disease is rare among people with abnormal LFT results in primary care. Only two analytes (ALT and ALP) are helpful in identifying the majority of liver disease. GGT adds little information in return for a high false-positive rate but it is sensitive to alcohol intake. LFT results seldom revert from abnormal to normal over a 1-month period, and modelling shows that repeating an abnormal LFT panel, as recommended in the current guidelines, is inefficient. LFTs are often undertaken to meet perceived patient need for a blood test, but as they are neither specific nor indicative of any particular disease they are among the least suitable tests for this purpose. Obesity and raised ALT provide strong evidence for a presumptive diagnosis of 'fatty' liver. Abnormal LFTs and 'fatty' liver provoke only short-term anxiety and neither is associated with sustained weight loss. Even a small amount of weight loss reduces liver fat.

Future work recommendations: (1) the cases of 'fatty liver' and controls should be followed up in the long term to identify features that predict development of hepatosteatosis and then cirrhosis; (2) the acceptability of replacing the traditional six- to eight-analyte LFT panel with a drop down menu including the ALT/ALP combination should be evaluated.

Funding: The National Institute for Health Research Health Technology Assessment programme.

Contents

List of abbreviations

A1AT	alpha-1 antitrypsin
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AMA	anti-mitochondrial antibody
ANOVA	analysis of variance
AST	aspartate aminotransferase
AUC	area under the curve
BALLETS	Birmingham and Lambeth Liver Evaluation Testing Strategies
BMI	body mass index
CI	confidence interval
COREC	Central Office of Research Ethics Committees
df	degrees of freedom
FBC	full blood count
FU1	first follow-up test
FU2	second follow-up test
GGT	gamma-glutamyltransferase
GP	general practitioner
HBV	hepatitis B virus
HCC	hepatocellular carcinoma
HCV	hepatitis C virus
HTA	Health Technology Assessment
ICC	intraclass correlation coefficient
ICER	incremental cost-effectiveness ratio
IQR	interquartile range
LFT	liver function test
NAFLD	non-alcoholic fatty liver disease
NASH	non-alcoholic steatohepatitis
NFS	NAFLD fibrosis score
NIHR	National Institute for Health Research
NPV	negative predictive value
OR	odds ratio
PBC	primary biliary cirrhosis
PCP	primary-care practitioner
PI	Principal Investigator
PPV	positive predictive value
PSC	primary sclerosing cholangitis
Q–Q	quantile–quantile
ROC	receiver operating characteristic
SD	standard deviation
SE	standard error
SMA	smooth muscle antibody
TFT	thyroid function test

ULN	upper limit of normal
USS	ultrasound scan
WHO	World Health Organization

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or in the notes at the end of the table.

Scientific summary

Background

Many millions of liver function tests (LFTs) are performed in England each year. Yet it is not known whether or not it is appropriate to order so many LFTs, what should be done when the LFT result is abnormal, and which of the analytes that might be included in the LFT 'panel' are most useful. These uncertainties all stem from ignorance about what LFT results mean in terms of the probabilities of the various diseases they may portend. This state of affairs has come about because almost all of the 14,000 studies reviewed in the literature describe laboratory findings *given* a disease. The Birmingham and Lambeth Liver Evaluation Testing Strategies (BALLETS) study set out to describe the probability of the various diseases *given* the pattern of abnormal LFTs. BALLETS achieved this objective by generating a large cohort of people with abnormal LFT results in primary care, fully characterising these people on the basis of clinical features and special investigations and then following them up after 2 years.

Objectives

The primary objective was to measure the probabilities of various diseases (or classes of diseases) when LFT results are abnormal and to determine how these probabilities vary according to the type of LFT abnormality and the clinical features of each patient. The secondary objectives were to evaluate the extent to which abnormal LFT results progressed or remitted over a 2-year period; to find out which combinations of clinical features and laboratory tests best predict 'fatty liver'; to determine proportions of 'fatty livers' that progressed, improved or stayed the same; to investigate the effect of ultrasound findings on health behaviour; and to investigate redundancy among LFT analytes. We also set out to:

- measure the psychological effects of positive LFT and ultrasound tests
- explore the effect of these tests on attitudes towards unhealthy behaviours
- document general practitioners motivations for ordering LFTs
- model the efficiency of various options when LFT test results are abnormal
- obtain preliminary information on use of a liver fibrosis scale in primary care.

Methods

We created a cohort of 1290 patients with abnormal LFT results in primary care and characterised them fully by means of their clinical details, an extensive battery of blood tests and ultrasound examination of the upper abdomen. We also followed up the patients after 2 years. Statistical tests were used to identify the interactions between clinical features, the initial pattern of abnormal LFTs and the following categories:

1. specific viral, genetic and autoimmune diseases of the liver, such as viral hepatitis, haemochromatosis and primary biliary cirrhosis
2. a range of other serious diseases affecting the liver, such as metastatic cancer and hypothyroidism
3. 'fatty liver' not associated with the above
4. no disease detected.

These interactions were explored by means of univariate analyses and multivariate analyses carried out with and without imputations for missing data. We also examined the influence of lifestyle and of weight loss on 'fatty liver', and then looked for evidence that the finding of a 'fatty liver' would motivate people to lose weight.

In addition, we studied the psychological effects of receiving an abnormal test result. Patients were sent a validated psychological questionnaire to measure anxiety and self-reported health on entry to the study and again at the 2-year follow-up point. A qualitative study was conducted after 2 years to explore perceptions of the effects of participating in the BALLETS study, and of abnormal test results, on behaviours and attitudes toward health. Clinicians' motivations for ordering LFTs were explored by means of a semistructured interview. We created a decision-analytic model to evaluate strategies that might be pursued in the face of an abnormal LFT result and to identify the most efficient option. Lastly, we conducted a preliminary study of a liver fibrosis score that might identify cases of 'fatty liver' at greatest risk of progression.

Results

1. Fewer than 5% of people with abnormal LFT results had a specific disease affecting the liver and there was a serious liver disease requiring immediate therapy in 1.3% of cases (all 13 cases of viral hepatitis and four cases of homozygous haemochromatosis).
2. The majority of serious or potentially serious diseases can be detected by just two analytes alanine aminotransferase (ALT) and alkaline phosphatase (ALP) from the LFT panel of eight analytes. The ALT enzyme is sensitive for hepatocellular disease, whereas ALP is sensitive for both hepatobiliary diseases and systemic diseases (such as metastatic cancer) affecting the liver.
3. Aspartate aminotransferase (AST) adds little to ALT and is considerably less sensitive (although it is slightly more specific).
4. The gamma-glutamyltransferase (GGT) enzyme was the most frequently abnormal analyte with a very high false-positive rate, but offered only a marginal increase in sensitivity in return. Unlike other analytes, the degree of abnormality is not indicative of the probability of disease. This is consistent with the poor discriminatory characteristics of this test in determining the presence, or absence, of pathology. GGT levels were sensitive, however, to alcohol intake.
5. Protein levels (albumin, globulin and total protein) are the least frequently abnormal analytes and they are typically only very 'mildly' abnormal. Albumin increases with age and comorbidity, but was not strongly related to any disease involving the liver.
6. Viral hepatitis was found in 1% of patients. Nine of the 13 patients with chronic viral hepatitis had more than one abnormal analyte and ALT was the most commonly abnormal analyte, followed closely by AST. The degree of ALT and AST abnormalities was, on average, considerably higher in patients with viral hepatitis than in the remaining patients. Country of origin (not ethnic group) was, by a considerable margin, the strongest predictor of viral hepatitis.
7. Guidelines recommend repeating LFTs in the event of an abnormal result, but 84% of tests remained abnormal on retesting after an average of 1 month, and even at 2 years 75% remained abnormal. Modelling confirmed the intuition that it is frequently more efficient, when confronted by an abnormal LFT, to proceed directly to a specific test rather than repeat the LFT with a view to specific testing only if it remains abnormal.
8. Nearly 4 in 10 patients had a 'fatty liver' on ultrasound, and an abnormal ALT level was the strongest laboratory predictor of this finding. Obesity was more strongly associated with 'fatty liver' than with alcohol use, but one-quarter of patients with 'fatty liver' were neither overweight nor excessive alcohol drinkers.

9. A small amount of weight loss over 2 years (1.3% reduction in body mass index) was associated with a reduced incidence of 'fatty liver'. There was a J-shaped relationship between alcohol intake and 'fatty liver' in men.
10. An abnormal LFT result generated anxiety and this anxiety was non-significantly greater if the liver was 'fatty.' However, anxiety dissipated over 2 years. Recall of an abnormal test result was hazy after 2 years and a tendency towards greater weight loss in patients with 'fatty liver' was not statistically significant.
11. Doctors' motivations for performing LFTs are mixed, and the tests are often carried out to meet perceived patient need for a 'blood test' or as a defensive practice. There was evidence that they were often undertaken as a semiautomatic or 'tick-box' response.
12. Eight per cent of patients with non-alcoholic 'fatty liver' had a fibrosis score that has been shown to be associated with a progressive disease in hospital-based studies.

Conclusions

1. It is unusual for an abnormal LFT result to signify a serious treatable disease of which the doctor was previously unaware.
2. Liver function tests are often carried out for social and psychological, rather than clinical, reasons. Given the high false-positive rate of LFTs and the fact that an abnormal result does not signal any particular disease, we recommend a more selective approach to this particular 'blood test'.
3. Aspartate aminotransferase is less sensitive than ALT for hepatocellular diseases, and GGT is very non-specific. There is a case for omitting these tests from the standard LFT panel and holding them in reserve for patients in whom alcohol abuse is suspected.
4. The standard advice to repeat an abnormal LFT does not gain support from the decision model and was one of the least efficient strategies with respect to diagnosis of viral hepatitis.
5. Country of origin is the strongest predictor of viral hepatitis among people with abnormal LFTs.
6. An abnormal ALT is strongly predictive of a 'fatty liver', as is obesity. If a person is obese and has a high ALT then an ultrasound diagnosis of 'fatty liver' is very probable.
7. There is no good evidence that single abnormal LFTs or ultrasound findings promote healthy behaviour.

Implications for practice

1. Liver function tests should be used sparingly in primary care.
2. The default LFT panel of five to eight analytes is obsolete.
3. When a chronic disease affecting the liver is suspected, a panel of two analytes (ALT and ALP) should be used, supplemented by bilirubin if an acute disease or poisoning is suspected.
4. When the clinician wishes to exclude a non-liver disease or simply reassure the patient, a selection should be made from a 'dropdown' menu of tests, and tests that provide a clear pointer to the next appropriate step should be favoured.
5. All patients who drink too much alcohol or who are obese should be given appropriate advice, irrespective of their LFT result. A single abnormal LFT does not promote healthy behaviour and use of serial LFTs to promote behaviour change is an unproven therapy that might do more harm than good.

Implications for research

1. A pilot study of a 'customised' approach to test ordering should be considered. The clinical value of different tests when patients have vague symptoms, such as tiredness or upper abdominal pain, should be evaluated. Likewise, the need to carry out more blood tests when patients are on treatment for chronic disease, such as hypertension, is unclear. There is a mismatch between the frequency with which blood tests are used to monitor chronic diseases and investigate symptoms, on the one hand, and scientific exploration of this subject, on the other.
2. The BALLETS cohort should be followed up over time to find out whether it is possible to identify the minority of patients with 'fatty liver' who are likely to progress to cirrhosis and to evaluate the fibrosis score in a primary-care setting.
3. A controlled study of the net effects of using serial LFTs (including liver ultrasound) as part of a package to reduce unhealthy behaviours should be seriously considered, especially in light of the rising incidence of obesity.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1

Introduction to report

Background

Introduction

Liver disease represents a major source of morbidity and mortality in the UK.[1] Abnormal liver function tests (LFTs) have been shown to be predictive not only of liver disease mortality, but also of more general causes of mortality.[2] LFTs are a good example of inexpensive tests (modern auto-analysers process large batches of samples using inexpensive reagents) that are frequently ordered as a 'test of exclusion' in patients with non-specific symptoms, such as tiredness or upper abdominal discomfort. The tests are also non-specific in the sense that none of the four to eight analytes included in the LFT panel points directly to a specific diagnosis, and many are not even specific to the liver. A doctor may order a laboratory test because a patient has features of a particular disease; for example, the gradual onset of jaundice in a user of injectable substances points to hepatitis C. The prior risk of hepatitis in such a person would be high: many positives would be true-positives. In most cases, however, LFTs are ordered without such a traceable link between symptoms and a specific diagnosis, for example when patients have vague symptoms or as part of the monitoring of patients with chronic diseases. Such tests are often offered as a type of insurance policy, but the prior risk of disease is low and the predictive value of LFTs is, a priori, likely to be low also. LFTs are interpreted by reference to population norms, rather than explicit calculus of the relative benefits and harms of false-positive and false-negative diagnoses. Many patients have a positive test, but it is not clear what proportion of these are true-positives, especially when the test result is only mildly abnormal. Review of the literature (see *Previous research*) shows that there is little evidence from large cohorts of people with abnormal LFT results to guide clinical actions when LFT results are mildly abnormal. The issue of how, or even whether, to investigate abnormal LFTs under various scenarios is not settled.

It is clear that a very large number of tests are ordered and abnormal results are common. The laboratory at University Hospital Birmingham received 67,182 requests for LFTs in 2003, from 83 general practitioner (GP) practices representing 210 GPs. Of these, 9779 (15%) led to an abnormal result in the sense that at least of one of the analytes on the LFT panel exceeded the reference range. As LFTs are inexpensive and easy to organise as one of the standard 'blood tests' in the GP's repertoire, their use has become widespread without careful study of their meaning in a general practice setting. As the meaning of the various combinations of possible test results and clinical features is unclear, different practitioners respond in different ways to the same test profile – the eclectic nature of practitioners' responses to the same scenarios has been well documented.[3]

Most abnormal LFT results are false-positives. Thus, large numbers of follow-on tests and much anxiety can ensue if a low threshold is used to define abnormality. On the other hand, there are arguments to adopt a low threshold for subsequent evaluation, as LFTs have the ability to detect diseases when they are most treatable, for example by reducing overload in patients with metal storage diseases or by administering antiviral agents in those with chronic viral hepatitis. Furthermore, theory-based interventions designed to modify behaviour that leads to liver damage, while clearly far from a panacea, nevertheless produces worthwhile benefits in that some people adopt healthy lifestyles when they perceive that their health is threatened and that engaging in the recommended behaviour will reduce this threat.[4–6]

The incidence of many liver diseases is rising, for example with migration from places with high rates of chronic hepatotoxic viral infection, and as a result of alcohol and calorie excess. Comorbidity is becoming more common as alcohol misuse and calorie excess unmask other diseases of the liver, such as haemochromatosis.

Thus, three interacting factors create an urgent need to better understand the clinical epidemiology of abnormal LFTs:

1. frequent use of these tests
2. lack of clarity about the meaning of the results
3. increasing treatability and rising incidence of liver diseases.

A number of authors have produced diagnostic algorithms for the investigation of people with abnormal LFTs.[7–12] These provide sensible advice – for example stressing the importance of taking a careful family history or of responding to tests that suggest obstructive biliary disease – but they do not provide a clear probabilistic basis for their reasoning. In particular, there is no scientific rationale for the widespread advice to repeat an abnormal LFT before conducting further tests. Green and Flamm[13] state in their 2002 review of 1400 papers: 'Unfortunately … there are no long term prospective studies to define the natural history of liver disease in patients with abnormal liver chemistries tests.' They call for a substantial prospective study of a well-documented population given a standardised diagnostic work-up in general practice and then followed up for a period of time. It was this gap in the literature that the Birmingham and Lambeth Liver Evaluation Testing Strategies (BALLETS) study was designed to rectify.

Previous research

There is considerable literature on the laboratory measurement of analytes. Dufour *et al.*[14,15] carried out a systematic review of this topic in 2000. This review contains much useful information on biological variability and how it is affected by sex, age, race, use of the oral contraceptive pill (and other medicines), pregnancy, exercise, delay in analysis and time of day. The study also reviews the patterns of abnormality of each analyte given different diseases. A further systematic review that distilled 14,000 references was commissioned by the American Gastroenterology Association Clinical Practice Committee in 2002.[13] Again, most of the references describe probabilities of test results given various diseases, rather than the probabilities of the various diseases given test results. For example, Bonacini[16] describes 'test results in people with cirrhosis due to chronic hepatitis infection'. Only a small proportion of articles report likelihood of disease by test result. Studies in this category tend to be based on hospital patients with serious abnormalities, such as 'notably raised aspartate aminotransferase'[17] or 'requiring liver biopsy'.[18–20] Angulo *et al.*[21] investigated a remarkable 733 patients with non-alcoholic fatty liver disease (NAFLD) confirmed by liver biopsy to determine which features were associated with more serious disease, while Ekstedt *et al.*[22] followed up 129 patients with biopsy-confirmed NAFLD for a mean of 13 years and showed that the subgroup with 'steatohepatitis' had an increased risk of both cardiovascular and liver-related death compared with a reference population.

We updated the above review (*Table 1*) and selected studies that started with the LFT result and then followed the cohort, so as to provide the type of probability needed for decision-making. MEDLINE was interrogated, with limits placed on the overall search with respect to 'humans' and 'publishing date post 1980'. Owing to the variety of nomenclature regarding LFTs a variety of search strings were used for this category. Search strings relating to abnormal LFT results included 'liver function test', 'transaminases', 'alanine aminotransferases', 'aspartate aminotransferases', 'alkaline phosphatase' and 'gamma-glutamyltransferases'. The search was focused by using the limits of blood, analysis and metabolism. Despite the limits, these search

TABLE 1 Search strategy for studies looking prospectively at patients who have received an abnormal LFT result

LFT search strings (limited using the subheadings; blood, analysis and metabolism)	Hepatitis search strings
'liver function test'	'liver diseases' (diagnosis)
'transaminases'	'liver diseases' (epidemiology)
'alanine aminotransferases'	'liver diseases' (enzymology)
'aspartate aminotransferases'	'liver diseases' (virology)
'alkaline phosphatase'	'liver diseases'
'gamma-glutamyltransferases'	
With limits added ('humans' and 'publishing date post 1980')	*With limits added ('humans' and 'publishing date post 1980')*
Papers returned = 35,070	**Papers returned = 8526**

Combining the above sets of papers using the term AND yielded 1448 papers. Abstracts were read for these papers.

strings retrieved over 35,000 references. The term 'hepatitis' was considered too narrow when attempting to find studies that followed up patients for a variety of diseases, so the more general term of 'liver diseases' was included, with limits of diagnosis, enzymology, epidemiology, mortality and virology, which retrieved around 8500 references. When these two search strategies were combined, 1448 papers were returned, the abstracts of which were read.

Eight studies were found that matched our requirement of following up patients who had experienced an abnormal LFT result. Two additional articles were selected from the references of relevant studies. As a result, to the best of our knowledge, there are only 10 studies for which a cohort of asymptomatic patients with abnormal LFTs was followed up (*Table 2*). However, one article was written in Korean (only the abstract was translated) and was excluded from our analysis.

Two of the remaining nine English-language papers described record linkage studies. One such study was based on the Korean insurance database, which was linked with death certificates.[23] This study reported that increased alanine aminotransferase (ALT), even within the upper end of the normal range, was associated with eventual death from liver disease. A study carried out in Scotland linked general practice and hospital databases.[2,24] However, this was a retrospective study so a full liver screen was not conducted and follow-up was for a median of only 4 years, whereas many diseases, including chronic viral hepatitis, have much longer prodromal periods.[25]

The other seven studies were prospective cohort studies, based on testing asymptomatic members of the general population. The famous Dionysos study,[26] based on three analytes from the LFT panel, is included among these. In this study, an impressive 6917 citizens from two communities in northern Italy were screened. Although the authors tested for viral hepatitis all of those in whom the LFT result was abnormal ($n = 1473$), and among whom they found a prevalence rate of 2.4%, the main aim of their study was to determine the effect of alcohol and diet on LFTs. Testing for viral hepatitis was used as a method of excluding causes of liver damage other than their topic of interest, so in-depth analysis on how viral hepatitis affected the pattern of LFTs was not published. Another Italian study, by Pendino *et al.*,[27] screened 1645 inhabitants from a town in southern Italy, with both a LFT [ALT, aspartate aminotransferase (AST) and gamma-glutamyltransferase (GGT)] and viral screen.[27] The prevalence of viral hepatitis is much higher in this region because of a significant immigrant population, and the authors performed a more extensive analysis of the impact of viral hepatitis on LFTs. Of the 319 (19.4%) individuals in whom LFT results were abnormal, nearly 18% were infected with viral hepatitis. However, the LFT failed to detect 34 (37%) of the 92 cases of viral hepatitis present in the community. Perhaps

the most comprehensive prospective analysis looking at the effect of viral hepatitis on individual analytes was carried out on a population of Japanese office workers.[28] The study used data from compulsory health checks, which included an ALT/AST/GGT panel along with certain additional tests, including a viral screen, which were added for study purposes. The authors found that ALT was the most sensitive of the three analytes used, detecting nearly half of cases of viral hepatitis, while being abnormal in 14% of the cohort (278 abnormal results in 1973 participants). The remaining four prospectively designed studies were carried out in general practice and were therefore closer in population terms to the BALLETS cohort. However, three of these are restricted to patients with persistently abnormal LFT results over a 6-month period,[18,29,30] and one of these did not include a test for viral hepatitis. The final prospective study, by Whitehead *et al.*,[17] was small and based on only one analyte.

After this review of the literature we concluded that no study has fully investigated a cohort of patients in primary care with an abnormal LFT result (from the full LFT panel) and no obvious or known liver disease. BALLETS is thus the first study to test the validity of the various strategies that a GP could use to make a diagnosis in patients with abnormal LFTs. The BALLETS study was based on performing a full LFT panel of investigations to identify diseases such as chronic viral hepatitis and primary biliary cirrhosis (PBC) that could otherwise be identified only by follow-up lasting many decades. The study was therefore designed to look into, and 'concertina', the future. Patients were also followed up over 2 years to detect systemic diseases attacking the liver (e.g. disseminated cancer), to follow the progress of people with excess alcohol consumption and/or 'fatty liver' on ultrasound and to ascertain the rates at which abnormal LFTs reverted to normal according to diagnostic category and type of analyte that was abnormal.

We also identified a relevant study by Kim *et al.*[31] This study prospectively followed a group of 'healthy' Korean factory workers, taking measurements of ALT, AST and GGT on at least two separate occasions. The full article was in Korean so we had access to the abstract only.

Structure of this report

The central idea behind the BALLETS study was to create a well-characterised cohort (as described above) and follow patients for 2 years. A database would thereby be created for statistical analysis. The generation and analysis of this database are referred to as the 'main study'. The objectives of this study are detailed in *Chapter 2*, the methods are described in *Chapter 3* and the results are presented in *Chapter 4*. The report also contains a series of substudies, the objectives of which are spelled out in *Chapter 2*. The methods and results of these substudies are then described in *Chapter 5*, which contains sections dealing with the psychological effects of a positive test (see *Chapter 5 Psychology 1: effects of positive tests*); a qualitative account concerning the effects of testing on behaviour (see *Chapter 5 Psychology 2: effects of results on behaviour*); a qualitative account of clinicians' motivations for testing (see *Chapter 5 Sociology of testing: an exploration of the clinical and non-clinical motives behind the decision to order a liver function test*); a decision analysis covering options following a positive LFT test result (see *Chapter 6*); and a study of markers for fibrosis in a subset of patients with 'fatty liver' from the Birmingham cohort (see *Chapter 6*). In *Chapter 6* we discuss the implications of our study, integrating lessons from the main study and substudies. We approach this task by imagining that all of the scientific information regarding LFTs – including that from the BALLETS study – was available, but that LFTs had not yet come into widespread, routine use. We also make use of the different reasons for testing that emerge from the qualitative substudy of GP reasons for ordering LFTs. This perspective leads to proposals to use different testing strategies according to the different reasons for conducting laboratory investigations. Perhaps provocatively we argue that the idea of a one-size-fits-all panel is obsolete. The original protocol for the study is included as *Appendix 1* (BALLETS study protocol).

TABLE 2 Studies that have followed up patients from a general or general practice population after abnormal LFT result

Author and country	Date	Type of study and population studied	Analytes used	No. of patients enrolled	No. of patients with abnormal LFT results (%)	Prevalence of viral hepatitis in patients with abnormal LFT results (%)	Notes
McLernon *et al.*,[2] Scotland	2009	Record linkage; laboratory database of GP tests, hospital admissions and death certificates	Bilirubin, albumin, ALP, GGT, ALT, AST (transaminases sometimes combined). GP selected	95,977	20,827 (21.7)	2.2	Mean follow-up of 3.7 years. Risk of underascertainment
Pendino *et al.*,[27] Italy	2005	Prospective cohort study; general population	AST, ALT, GGT	1645	319 (19.4)	17.9	High baseline rate of viral hepatitis: 5.6%
Kim *et al.*,[23] Korea	2004	Record linkage: insurance data and death certificates	AST, ALT	142,055	11,193 (7.9)	N/A	Outcome was liver disease mortality
Yano *et al.*,[28] Japan	2001	Prospective cohort study; 'healthy' office workers	AST, ALT, GGT	1973	358 (18.1)	2.7	Assumed that all liver cancer and cirrhosis was a result of viral hepatitis
Daniel *et al.*,[18] USA	1999	Prospective cohort study; primary-care population	ALT, AST raised 50% above normal on at least two occasions across a 6-month period	1124	1124 (100)	N/A	Marker-negative patients only, so infected patients excluded from analysis
Mathiesen *et al.*,[30] Sweden	1999	Prospective cohort study; primary-care population	AST, ALT raised for at least 6 months (ALP had to be normal)	150	150 (100)	15.3	
Whitehead *et al.*,[17] UK	1999	Prospective cohort study; primary-care population	AST markedly raised [10 times (> 400 U/l) above the ULN]	137	137 (100)	2.2	
Bellentani *et al.*,[26] Italy	1994	Prospective cohort; general population	AST, ALT, GGT	6917	1473 (21.3)	2.4	
Hultcrantz *et al.*,[29] Scandinavia	1986	Prospective cohort study; primary-care population	AST, ALT moderately raised for at least 6 months (ALP had to be below twice the ULN)	149	149 (100)	2.7	

ALP, alkaline phosphatase; N/A, not applicable; ULN, upper limit of normal.

Chapter 2

Objectives

Main study

The Health Technology Assessment (HTA) commissioning brief made it clear that the overall objective was to inform general practice decision-making. Thus, the main objective can be framed as follows: 'How does the probability of disease vary by the pattern of abnormal LFTs and the clinical features of a patient?'. 'Pattern' of abnormal LFTs describes which analytes are abnormal (singly or in combination) and the degree (extent, magnitude) of the abnormality. In particular, we set out to ascertain the predictive value of the pattern of LFTs for the specific and often treatable viral, genetic or autoimmune liver diseases in *Table 3*.

Secondary objectives of the main study were:

1. To follow up people who had neither one of the above serious and treatable liver diseases nor another serious disease (such as metastatic cancer) and to evaluate the extent to which abnormal LFTs progressed or remitted over a 2-year period.
2. To determine the proportions where 'fatty liver' progressed, improved or stayed the same and to investigate how clinical, behavioural and biochemical features correlated with progression, resolution or maintenance of the ultrasound finding. This study was not part of the original protocol but was prompted by the high incidence of fatty liver at entry to the study. Repeat ultrasound was funded under an extension to the original grant.
3. To investigate the issue of redundancy among LFT analytes by measuring what would be lost in terms of prognostic accuracy by dropping certain analytes from the full panel of LFT analytes. This is an important issue because the benefit of analytes that offer small marginal gains in detection rates may be outweighed by losses as a result of false-positives.
4. To shed light on the utility of undertaking LFTs in the first place by determining the prevalence of serious disease in the cohort as a whole.

Some of these figures may be underestimates of the incidence of the various pathological entities since we now know that many people may have subclinical disease with such long lead times that they do not present clinically during the person's lifetime. This applies in particular to haemochromatosis and PBC, a point to which we return.

TABLE 3 Serious specific diseases of the liver and prevalence in the British population

Disease	Prevalence (%)	Source
Chronic viral hepatitis C	0.42	Health Protection Agency website, cited 2011[32]
Chronic viral hepatitis B	0.3	Health Protection Agency website, cited 2011[33]
Metal storage disease – iron	0.25	Worwood 1998[34]
PBC	0.024	Metcalf *et al.* 1997[35]
Autoimmune hepatitis	0.001	Autoimmune Hepatitis website, cited 2009[36]
Metal storage disease – copper	< 0.025	Olivarez *et al.* 2001[37]
A1AT deficiency	< 0.025	de Serres 2002[38]

A1AT, alpha-1 antitrypsin; PBC, primary biliary cirrhosis.

Psychological substudy

Abnormal LFTs may have psychological consequences, and this is important given the high proportion of false-positive results that were anticipated. The original protocol thus included a psychological substudy based mainly around the measurement of (any) induced anxiety at various stages following disclosure of a positive result.

We became increasingly aware that knowledge of abnormal LFT results, and performance of some tests prompted by abnormal LFT results, might constitute an intervention in their own right, as news of these results might affect behaviour (see *Sociological substudy*, below). For example, a person with persistently abnormal LFT results and an ultrasound diagnosis of fatty liver may be influenced by these results to modify unhealthy behaviour (excessive calorie and/or alcohol intake). Conversely, a normal result may provide false reassurance. The follow-on study was thus adapted not only to observe any residual anxiety caused by testing, but also to collect data on (any) changes in eating and drinking habits. The additional data collection for this purpose at the 2-year follow-up point was funded by an extension to the HTA grant.

Sociological substudy

A (perhaps predictable) early finding from our study was that LFTs do not offer high diagnostic precision, and that the positive predictive value (PPV) (probability of disease given a positive test) is low. Moreover, the value of LFTs, as of any test, lies in its incremental diagnostic accuracy given what the doctor knows before the result is made available. For example, finding a raised ALT level in a patient with a known alcohol problem would not be a surprise. On the other hand, such a result may buttress the doctor's advice to reduce alcohol consumption. These considerations raise the question of why so many LFTs are ordered in the first place. If GPs (erroneously) thought that LFTs were highly predictive of serious treatable disease then we may expect the BALLETS results to reduce demand for LFTs. If, however, the low predictive value of these tests is not news to GPs then other approaches would be necessary to reduce test ordering (if this was perceived as desirable – see *Decision analysis*, below). We therefore carried out a further study, not included in the original protocol, to find out more about GPs' motivation for ordering LFTs. This substudy included a general review of the literature on GPs' test-ordering behaviour. The protocol for this study is described in this report.

Decision analysis

As stated in *Sociological substudy*, above, it became clear from the literature (and emerging results in this study) that the predictive value of LFTs was rather low. This raises the question of what action (if any) a doctor should take when confronted by a mildly abnormal LFT result. Clearly, if there is an obvious clinical lead then this should be followed; for example, if a person has a history of intravenous drug use then a test for viral hepatitis is indicated. However, the majority of cases are more ambiguous. We therefore decided that it would be helpful to carry out a formal decision analysis to examine the losses and gains associated with various clinical opinions. Conducting a decision analysis for each potential disease and then consolidating them into one composite analysis would be well beyond the scope and resources of this project. We therefore selected one disease class – chronic viral hepatitis – as an exemplar on the basis that:

1. Unlike high alcohol intake and obesity, the clinician can diagnose the condition only by further testing.

2. The disease, if caught early, is highly treatable.
3. It is one of the most common of the specific liver diseases to present clinically.

We were aware of the previous decision analysis in the previous HTA report[2] and our analysis includes a critique of this work.

Biochemistry of ongoing liver disease

It became clear at an early stage that the BALLETS study would generate a sizable cohort of people with fatty liver.

The extensive testing algorithm incorporated in the study did not include all necessary tests for the diagnosis of the enigmatic condition called 'metabolic syndrome'. The literature suggests that a small percentage (5–10%) of people with fatty liver would progress to liver fibrosis, and the BALLETS study provides a platform for the study of novel blood tests that might predict such progression. We therefore performed an add-on study in which a fibrosis score was calculated. In addition, new hypotheses concerning the origin and prognosis of fatty liver may emerge over the next 4 years in this fast-moving field of enquiry. For these reasons, additional funding was sought and granted by the HTA programme to store frozen blood samples from consenting participants.

Chapter 3

Methods: main study

Selection of practices and patients

Practices were selected on the basis of geographic spread and their willingness to join the study. They had to be multiple-partner practices. We deliberately included inner city practices in order to 'enrich' the population to include a higher than average proportion of chronic viral hepatitis. Two city areas were selected: Birmingham and the Lambeth district of London. This was done so that the relationship between LFTs and this disease could be studied. The geographical location and demographic and ethnic features of the eight Birmingham practices and three Lambeth practices that we were able to recruit are described in *Chapter 4* (see *Nature of the population studied*: *Birmingham and Lambeth sites*).

General practitioners from participating practices reviewed all abnormal LFT results arising in their practice to determine eligibility. Patients aged > 18 years were eligible if one or more analyte was abnormal, they did not have known liver disease, they were not deemed to require immediate referral to hospital and they were not pregnant. Seven out of the eight Birmingham practices sent samples to a single laboratory (University Hospitals Birmingham NHS Foundation Trust laboratories), whereas the remaining practice (Wand Medical Centre) sent samples to the laboratory of Russells Hall Hospital. All Lambeth practices used a single laboratory (Guy's and St Thomas' NHS Foundation Trust laboratory). The repertoire of analyses included, prompted by a request for LFTs from the participating practices, was extended over the study period from the usual five analytes in our laboratories to all eight listed in *Table 4*. The idea was to enable redundancy between tests to be detected and to help generalise to centres that included different analytes. The analytes were classified as normal or abnormal according to standard laboratory practice that is compliant with International Quality Control Standards. The classification was based on reference ranges specific to each of the (three) individual laboratories (see *Table 4*).

Eligible patients were contacted to seek verbal consent to participate in the study. The method of contact varied from practice to practice so that it would be compatible with the normal procedures used in the practices. The bespoke protocols to inform patients of their results and the study process are described, for each practice, in *Appendix 1* (*section 10.2a–f*). Once an eligible patient had been identified he or she was contacted and invited to attend the practice for a study session. The practice sent a Patient Information Sheet to all potential patients in advance of their attendance at the study session.

Testing strategy for patients in the Birmingham and Lambeth Liver Evaluation Testing Strategies study

Formal written consent was sought when the patient attended the study session. The following information was collected and recorded:

1. Clinical details (*Table 5*).
2. An alcohol use questionnaire was completed and the patient's weight, height, waist and hip size were measured (*Table 6*).

TABLE 4 Analytes and reference ranges (by centre) for analytes included in the LFT panel in the study

Test	Reference range		
	University Hospitals Birmingham NHS Foundation Trust	Russells Hall Hospital NHS Trust	Guy's and St Thomas' NHS Foundation Trust
ALT	1–41 U/l	1–56 U/l	1–45 U/l M, 1–28 U/l F
AST	1–43 U/l	1–45 U/l	1–49 U/l
Bilirubin	1–22 µmol/l	1–22 µmol/l	1–22 µmol/l
ALP	1–320 U/l age < 40 years M	1–120 U/l	1–129 U/l
	1–330 U/l age ≥ 40 years M		
	1–260 U/l age < 40 years F		
	1–290 U/l age 40–49 years F		
	1–330 U/l age ≥ 50 years F		
GGT	1–40 U/l F, 1–50 U/l M	1–58 U/l	1–65 U/l M, 1–38 U/l F
Albumin	34–51 g/l	35–47 g/l	40–52 g/l
Globulin (derived)	21–37 g/l	21–37 g/l	21–37 g/l
Total protein	60–80 g/l	65–83 g/l	61–79 g/l

ALP, alkaline phosphatase; F, female; M, male.
Low levels were regarded as 'abnormal' for only the protein measurements.

TABLE 5 Clinical data collected

1.	GP name and practice code			
2.	Patient study ID			
3.	Name and address			
4.	Date of birth			
5.	NHS no.			
6.	Gender			
7.	Current and recent medication			
8.	Reason for GP consultation/LFTs ordered?			
9.	Current/past Illnesses			
10.	Recent febrile Illness			
11.	Recent muscle damage			
12.	Substance abuse	Past ❑	Current ❑	Intravenous ❑ Oral ❑
13.	Recent travel history	Over last 6 months?		Where?
14.	Immunisation against HBV			
15.	Transfusion history	No ❑	Yes❑	Date
16.	Length of residence in the UK			
17.	Ethnic group			
18.	Preferred language			
19.	Country of birth			

3. A single blood sample was obtained for detailed analysis. The LFT panel was repeated along with tests for specific (autoimmune, genetic and viral) diseases (*Table 7*).
4. An ultrasound scan (USS) of the liver was obtained using a portable ultrasound machine (TITAN® SonoSite) operated by experienced (10 years minimum) sonographers from the

TABLE 6 Alcohol consumption and weight/height pro forma

Alcohol consumption (units per week over past 6 months?)								
a. How often do you drink?	Annually	Special occasions		Monthly			Fortnightly	
	Weekly/daily	M	T	W	T	F	S	S
b. What is the type or brand?								
c. What size of glass or can do you drink?								
d. Number of each type of drink consumed in a session?								
Measurements								
Height (cm)								
Weight (kg)								
Waist measurement (cm)								
Hip measurement (cm)								

TABLE 7 Specific tests carried out, along with repeat LFT panel and ultrasound, on all consenting patients with abnormal LFT results

Hepatitis B viral markers (HBV surface Ag)
HCV antibody (HCV Ab)
A1AT
Caeruloplasmin
Iron and transferrin
SMA
AMAs

A1AT, alpha-1 antitrypsin; Ab, antibody; AMA, anti-mitochondrial antibody; HBV, hepatitis B virus; HCV, hepatitis C virus; SMA, smooth muscle antibody.

ultrasound department of the University Hospitals Birmingham NHS Foundation Trust, Worcester Acute Hospitals NHS Trust or Guy's and St Thomas' Hospital NHS Trust. The sonographer completed a pro forma (see *Appendix 1*, *section 10.7)* that included a description of liver texture on a four-point scale, indicating normal, mild, moderate and severe echo density. Fatty liver on ultrasound was determined by comparison of brightness/echogenicity in the liver with the right kidney. The sonographer notified the named or on-call GP of any findings of a sinister nature so that they could be acted upon immediately. All scans were recorded on tape and 50 of these were selected at random from the first participating practice for scrutiny by a senior radiologist, as a form of quality control (see *Quality control of ultrasound*).

The research team produced a consolidated report comprising the results of the index LFT and the first follow-up LFT, and all of the information described in *Tables 5–7*, along with the result of the ultrasound examination. The patient participant then attended the GP for a consultation informed by all of these data.

Note that the intention was for each patient to have three LFT panels performed as part of the BALLETS study:

1. the test that confirms eligibility: 'the index test'
2. repeat test on agreeing to enter the study: 'the first follow-up test' (FU1)
3. test at 2-year follow-up: 'the second follow-up test' (FU2).

The GPs were provided with a set of guidelines to assist decision-making when one of the tests in *Table 7* was abnormal or when an abnormality, such as fatty liver, was seen on the USS. The guidelines were produced by members of the study team (JN and RL) and approved by each practice. The guidelines are outlined in *Appendix 1* (*section 10.9*). In addition, clinical members of the research team visited practices to provide proctorship on what to do about abnormal results. The results of follow-up tests were obtained from the laboratories by the research team. In some cases a follow-on test indicated according to the guideline was absent from the laboratory records. In these cases the chief investigator contacted the practice concerned to remind the GP to consider recommending the test to the patient. This issue of missing follow-on blood tests had not been foreseen by the research team and ethical permission was obtained to amend the protocol so that GPs could be contacted.

The 2-year follow-up visit

A second follow-up visit was offered to patients 2 years after the first follow-up visit . The electronic patient records at practices were scrutinised where possible and patients placed in four categories for the purpose of 2-year follow-up:

1. *Deceased* The cause and date of death were ascertained from notes or the practice database.
2. *No longer registered with the practice* The new practice was contacted and the GP asked to invite the patient to attend for a second follow-up LFT for submission to the original laboratory.
3. *Patient under ongoing hospital care* The diagnosis was obtained from study hepatologists in Birmingham or Lambeth.
4. *Remaining patients* The remainder were invited to attend the practice for the second follow-up LFT. The weight and body measurements and alcohol history were repeated at this visit. Extensions to the protocol were obtained from the funder to enable patients at Birmingham to undergo a repeat ultrasound examination and to be asked to consent for an aliquot of blood being preserved for cryogenic storage of cells and serum. These protocol amendments and patient documentation for this enhanced follow-up in Birmingham were approved by the ethics committee.

A summary of the full patient journey is illustrated in *Figure 1*.

Laboratory methods

The biochemical measurements were carried out in the accredited (Clinical Pathology Accreditation UK) laboratories of University Hospitals Birmingham NHS Foundation Trust (Queen Elizabeth and Selly Oak Hospitals, Birmingham), of Guy's and St Thomas' NHS Foundation Trust (St Thomas' Hospital), and of Dudley Group of Hospitals NHS Foundation Trust (Russells Hall Hospital, Dudley). The measurements were performed on serum obtained from blood samples collected into Vacuette tubes (evacuated collection tubes) containing no anticoagulant (Greiner Bio-One GmbH, Kremsmuenster, Austria). Serum was obtained by centrifugation of the samples for 5 minutes at $1200 \times g$ and measurements were performed on a Roche Modular Analytic system using specific reagents supplied by Roche Diagnostics (Roche Diagnostics Ltd, Burgess Hill, UK) in University Hospitals Birmingham NHS Foundation

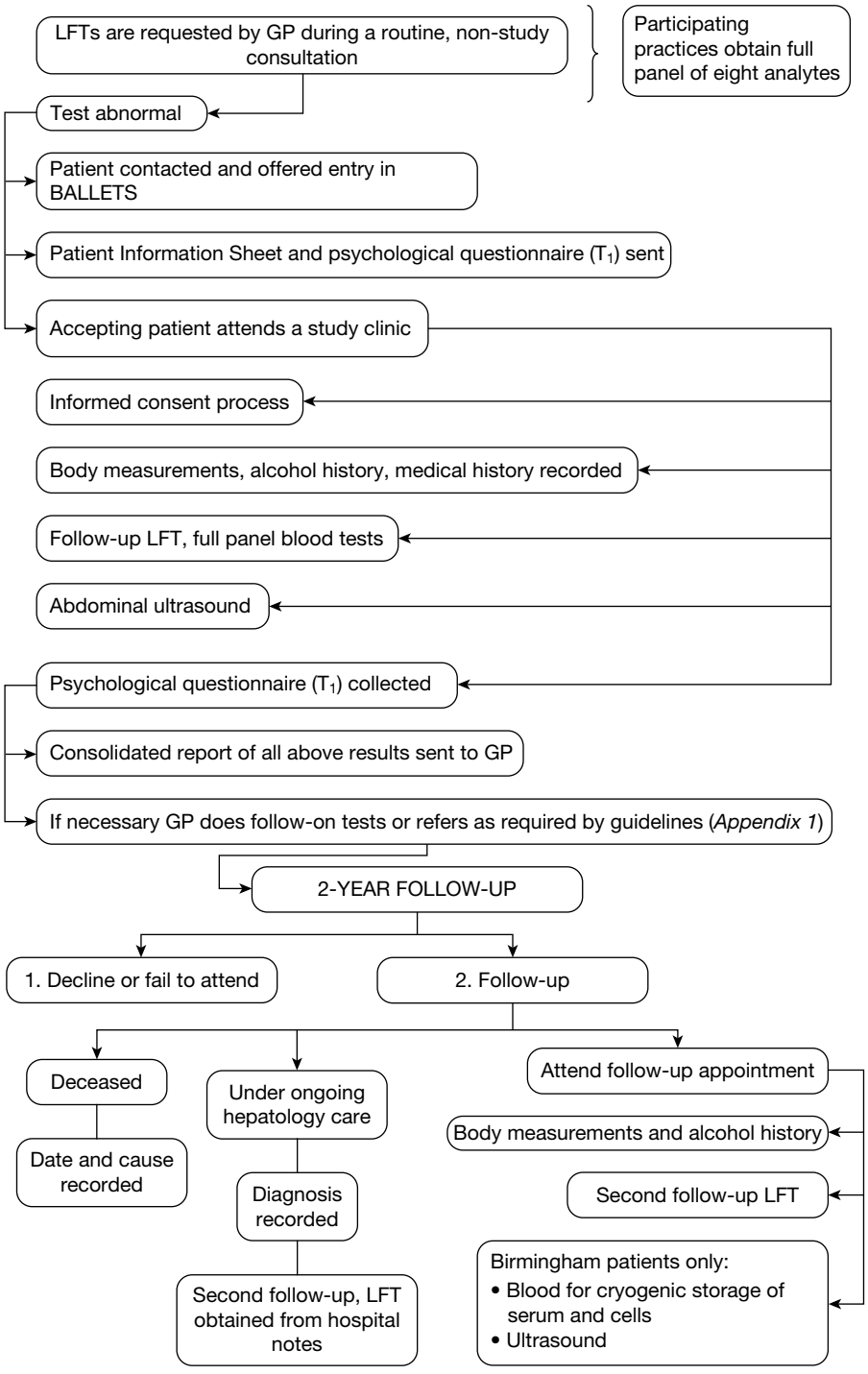


FIGURE 1 Patient journey through the BALLETS study.

Trust (Queen Elizabeth and Selly Oak Hospitals Birmingham) and Guy's and St Thomas' NHS Foundation Trust (St Thomas' Hospital), and on Vitros 5.1 analysers using reagents supplied by Ortho Clinical Diagnostics (Ortho Clinical Diagnostics, Johnson & Johnson, High Wycombe, UK) in the Dudley Group of Hospitals NHS Foundation Trust. ALT, albumin, alkaline phosphatase (ALP), AST, GGT, total bilirubin, total protein, caeruloplasmin and alpha-1 antitrypsin (A1AT) were assayed. Where A1AT concentrations were noted to be < 1.5 g/l, the sample was phenotyped by isoelectric focusing to help in diagnosis and monitoring. The

phenotyping was performed on a Sebia Hydrasys instrument (Sebia UK River Court, Camberley, UK) with specific reagents and isoelectric focusing gels.

Integral pilot

Purpose of integral pilot

Rather than follow convention and collect a full data set before setting out on the analysis it was decided to analyse data from the first practice to complete recruitment – the Hall Green Practice in Birmingham. This practice completed its recruitment at a point in time when recruitment in other practices was nascent or yet to begin. Analysis of the integral pilot was carried out as soon as the FU1 data became available, i.e. the integral pilot does not include the FU2 results.

The purposes of this pilot were threefold:

1. to 'test the system' by detecting incomplete data and exploring systematic failures so that remedial action could be taken where necessary
2. to compare patients entered in the study with those who might have been eligible but who were not entered in the study
3. to conduct a quality control study on the accuracy of ultrasound findings by reviewing a sample of images stored on tape.

Missing data

One hundred and sixty-one patients were entered in the study in the pilot practice. Two patients did not attend for the ultrasound examination and have been excluded from the pilot analysis. The following analyses all relate to the remaining 158 cases. Their age and sex distributions are shown in *Table 8*.

The index panel of LFT analytes was incomplete (i.e. not all of the eight results were available) for 26 out of the 158 patients and complete for 132 (84%) patients. The first follow-up panel of LFT analytes was not available in five cases and the panel was incomplete in 27 cases – thus complete data were available on the follow-up LFT panel for 126 out of the 158 (80%) patients. The full breakdown of the missing data is given in *Table 9*.

The missing data did not follow any anticipated pattern (see *Table 9*). One might have expected that if those cases for which all eight analytes required for study purposes had not been included then the five default analytes for this particular laboratory would have been measured. This would have resulted in bimodal distribution, with high peaks at eight and five analytes. On further enquiry, it transpired that the clerks who receive the request forms and enter the

TABLE 8 Age and sex characteristics of study patients in the integral pilot ($n = 158$)

Age (years)	Male, *n* (%)	Female, *n* (%)	Total, *n* (%)
≤ 44	24 (25.8)	14 (21.5)	38 (24.1)
45–54	20 (21.5)	9 (13.8)	29 (18.4)
55–64	20 (21.5)	19 (29.2)	39 (24.7)
≥ 65	29 (31.2)	23 (35.4)	52 (32.9)
Age (years), mean (SD)	54.7 (15.4)	57.8 (15.0)	56.0 (15.2)
Total	93 (100.0)	65 (100.0)	158 (100.0)

SD, standard deviation.

TABLE 9 Number of integral pilot patients with complete and incomplete enumeration of analytes at the index and follow-up panel of LFTs

No. of tests	Index test	FU1 test
0	0	5
1	0	0
2	0	1
3	0	7
4	13	14
5	6	1
6	6	0
7	1	4
8	132	126
Total	158	158

requests on computer do so with variable fidelity (for study patients as for routine patients). A programme of staff training was therefore put in place to try to reduce this problem. However, we were advised that with large numbers and high turnover of clerical staff in the laboratory, some remaining laboratory omissions were inevitable.

Comparison of patients who were and were not 'recruited'

Some eligible patients declined to participate, but we became aware that many more were not invited to participate by their GPs. Furthermore, some GPs recruited many more patients than others. One possible explanation was a tendency to select patients with the more severely abnormal results for entry in the study. This tendency could have been motivated by a desire to obtain all of the ancillary tests inherent in entry in the BALLETS study while reducing the need for further attendances and testing among those at lower perceived risk. This could lead to bias if, even among cases with equal severity of abnormality, GPs were somehow identifying patients with the worst prognosis for inclusion in the study. This could result in exaggerated estimates of the risks associated with abnormal LFTs.

In order to shed light on this issue, we collected baseline data from all (195) eligible but non-entered patients for two calendar months – May and June 2006 – and compared them with 53 participating patients for those months. This epoch was selected on the grounds that it corresponded to the period of highest recruitment.

The 195 non-entered patients constituted two subgroups: 129 patients had simply not been invited by the GP, despite fulfilling all objective criteria of entry to the study, while the remaining 66 had declined to take part (*Table 10a*). These subgroups are broken down by age and sex in *Table 10b*. The mean age of the invited patients, 58.6 years, is somewhat higher than the mean age of not-invited patients, 54.1 years ($p = 0.028$, two-sided t-test). There was no significant age difference between 'consenters' and 'refusers' within the invited group ($p = 0.766$). Thus, the 53 patients in the study tended to be older than those outside it. To put this in perspective, 68% (40/59) of eligible 65- to 74-year-olds were invited to join compared with 31% (22/72) of eligible patients under 45 years.

By contrast, the sex distribution was stable across all subgroups.

Abnormalities in the index LFTs for these 195 patients are analysed in *Tables 11* and *12*. The proportion of patients with abnormal GGT was higher ($p = 0.011$) among those invited to join the study (73.7%) than among those not invited (58.1%). However, there is no evidence

TABLE 10a Breakdown of eligible patients in the Hall Green practice

Status	*n*	Mean (SD) age (years)
Consented	53	58.2 (13.9)
Refused	66	59.0 (16.5)
Total invited	119	58.6 (15.3)
Not invited	129	54.1 (17.2)
Total	248	55.4 (17.3)

SD, standard deviation.

TABLE 10b Age and sex of eligible patients in the Hall Green practice with index panels taken in May and June 2006

Category	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Age (years)					
≤ 44	9 (17.0)	13 (19.7)	22 (18.5)	40 (31.0)	62 (25.0)
45–54	10 (18.9)	10 (15.2)	20 (16.8)	27 (20.9)	47 (19.0)
55–64	12 (22.6)	13 (19.7)	25 (21.0)	26 (20.2)	51 (20.6)
65+	22 (41.5)	30 (45.5)	52 (43.7)	36 (27.9)	88 (35.5)
Sex					
Male	33 (62.3)	42 (63.6)	75 (63.0)	77 (59.7)	152 (61.3)
Female	20 (37.7)	24 (36.4)	44 (37.0)	52 (40.3)	96 (38.7)

of preferential invitation associated with abnormality on any other analyte, nor with the presence of more than one abnormality in the index panel (see *Table 11*). However, there was an (unexplained) tendency for invited patients with abnormal globulin to decline to participate ($p = 0.002$). Otherwise we found no evidence of recruitment bias.

In the end, we were not able to exclude a degree of selection in patients entered. Some selection effects associated with age and GGT abnormality are suggested. Individual differences in how eligibility criteria are applied are inevitable in a large and busy practice and we cannot exclude a degree of bias owing to hidden confounders. If clinicians selected a group of patients with significantly higher prior risk, then it is possible that the study will somewhat overstate the association between mildly abnormal test results and the various disease end points. All we really know is that some apparently eligible patients were not invited by their GP to participate. This may be because of a purposive decision to exclude or because of some oversight. Such data are difficult to collect because doctors who are unwilling or too hard pressed to select patients for study entry are unlikely to go to the trouble of recording their reasons.

Quality control of ultrasound

In order to quality assure the liver imaging, the first (FU1) ultrasound images and paper reports of 50 randomly selected BALLETS patients were presented to the study radiologist, who was in complete agreement with the sonographer's findings in 34 out of the 50 cases. In the remaining 16 cases there was some 'technical or relatively minor' disagreement but 'no serious clinical disagreement' that might have altered clinical decision-making.

The process was repeated with 50 randomly selected FU2 ultrasound scans. The study radiologist agreed with the written report in 38 instances, and in the remaining 12 cases found that there was 'technical or relatively minor' disagreement, but, once again, no disagreement that would alter the clinical decision-making process.

TABLE 11 Proportions of abnormal analytes among eligible patients in the Hall Green practice from index panels taken in May and June 2006[a,b]

Analytes	Consented, *n* (%)	Refused, *n* (%)	Exact test	Total invited, *n* (%)	Not invited, *n* (%)	Exact test[c]	Total, *n* (%)
Total	53 (100)	66 (100)		119 (100)	129 (100)		248 (100)
ALT[d]	19 (38.0)	18 (27.3)	0.234	37 (31.9)	33 (25.6)	0.322	70 (28.6)
AST[d]	5 (10.2)	3 (4.5)	0.283	8 (7.0)	15 (11.6)	0.274	23 (9.4)
Bilirubin[d]	4 (8.2)	1 (1.5)	0.162	5 (4.3)	10 (7.8)	0.299	15 (6.1)
ALP[d]	2 (3.9)	7 (10.6)	0.295	9 (7.7)	16 (12.4)	0.291	25 (10.2)
GGT[d]	41 (78.9)	46 (69.7)	0.298	87 (73.7)	75 (58.1)	0.011	162 (65.6)
Albumin[d]	1 (1.9)	3 (4.5)	0.628	4 (3.4)	3 (2.3)	0.713	7 (2.8)
Globulin[d]	1 (2.0)	14 (21.2)	0.002	15 (12.9)	23 (17.8)	0.377	38 (15.5)
Total protein[d]	5 (10.0)	18 (27.3)	0.033	23 (19.8)	30 (23.3)	0.538	53 (21.6)
More than one abnormal analyte[e]	17 (37.8)	29 (43.9)	0.560	46 (41.4)	54 (41.9)	1.000	100 (41.7)

a All eligible patients have at least one abnormal analyte.
b Percentages for individual analytes are calculated from those who had a result for that analyte.
c Numbers and percentages computed from complete LFT panels only.
d The *p*-value for comparing proportions between the 'consented' and 'refused' (Fisher's exact test).
e The *p*-value for comparing 'invited' with 'not invited' (Fisher's exact test). Information on the analyte concentrations among abnormal results is assembled in *Table 12*. Among abnormal results there were no significant differences in levels between (1) 'invited' and 'not-invited' patients and (2) 'consenting' and 'refusing patients' for any analyte (Wilcoxon rank-sum tests), although the sample numbers are low for most comparisons.

TABLE 12 Levels of abnormal analytes in the Hall Green practice from index panels taken in May and June 2006[a]

	Consented			Refused			Total invited			Not invited		
Analyte	*n*	Excess mean	Excess median	*n*	Excess mean	Excess median	*n*	Excess mean	Excess median	*n*	Excess mean	Excess median
ALT	19	1.37	1.29	18	1.46	1.32	37	1.42	1.32	33	1.55	1.32
AST	5	1.37	1.16	3	1.52	1.56	8	1.42	1.47	15	1.54	1.26
Bilirubin	4	1.90	1.86	1	1.09	1.09	5	1.74	1.55	10	1.34	1.20
ALP	2	1.20	1.20	7	1.49	1.17	9	1.43	1.17	16	1.85	1.24
GGT	41	1.79	1.30	46	2.16	1.54	87	1.99	1.42	75	2.15	1.42
Albumin	1	1.06	1.06	1	1.02	1.02	2	1.04	1.04	2	1.03	1.03
Globulin	1	1.08	1.08	14	1.15	1.14	15	1.14	1.14	20	1.08	1.05
Total protein	5	1.02	1.02	18	1.05	1.04	23	1.04	1.04	30	1.04	1.02

a Excess means and medians are computed from abnormal (high-side) test results expressed in units of the upper limit of normality.

Production of reference categories (categories of diagnostic groupings)

The analysis plan in the study protocol was to investigate the association between the index LFT panel and clinical outcome in order to address such questions as:

1. Which profiles of index test results suggest higher and lower risk of the presence of serious specific disease, and of the other reference standards?
2. What is the contribution of different test analytes? How does this vary by clinical features?

The idea here was that the information the GP would have at the time of the index test would constitute explanatory variables in an analysis of clinical outcome using polytomous regression

methods to cope with multiple diagnostic categories. The BALLETS study would provide outcome data partly by repeating the LFT (at FU1 and FU2), but, more specifically, by doing exhaustive further testing to 'concertina' the future and reach a diagnosis.

This exercise required that each participant be assigned to an outcome (diagnostic) category. That is to say, we needed a reference standard. However, experience gained from our integral pilot suggests that this is tricky. The problem we encountered might be called 'multiple and overlapping categories'. Briefly, when we came to analyse the data, we found that patients did not fall into a manageable number of discrete categories. For example, the category 'fatty liver' could be divided into 'fatty liver alcohol excess', 'fatty liver overweight', 'fatty liver overweight and alcohol excess' and 'fatty liver and not overweight and no alcohol excess'. However, the job would still not be done – there could then be subcategories for each of the above according to whether the virology was positive or negative, for example. Then there would have to be categories for excess alcohol, overweight, viral diseases, immunological diseases and metal storage diseases – all with and without fatty liver. Our initial discussions with chemical pathologists, liver specialists and GPs suggested that consensus regarding a manageable number of mutually exclusive pathological diagnoses was unlikely to be obtainable. Indeed, even taking a liver biopsy would fall well short of resolving this issue.

It was therefore decided to 'collapse' the reference standards into a small number of broad 'action groups'. These groups are based on the appropriate clinical response, rather than on the precise underlying (and often unknowable) pathophysiological entity.

The following groups were created:

- **Group 1** Specific category of viral, autoimmune or genetic aetiology from *Table 13*:
  - *hepatocellular diseases*: chronic viral hepatitis B and hepatitis C, haemochromatosis, autoimmune hepatitis, Wilson's disease, antitrypsin deficiency, cirrhosis (alcohol or fat induced)
  - *diseases of the intrahepatic bile ducts*: PBC, primary sclerosing cholangitis (PSC).
- **Group 2** Serious liver or other pathology requiring referral. This would include metastatic cancer.
- **Group 3** Non-specific category. This is broken down into:
  - echo-bright (fatty) liver
  - not fatty liver.

The groups are hierarchical in the sense that a person would be assigned to the 'top' category when more than one category might apply. Thus, a person with 'two hits', such as haemochromatosis and 'fatty liver', would be assigned to group 1, not group 3.

The question could be asked as to why we did not include the diagnosis of alcoholic liver damage of a degree less extreme than cirrhosis. The answer is that, had we done so, alcohol use would serve two non-independent functions – as a clinical feature known in advance of testing (in many/most cases) and also as the outcome of testing. That is to say, the results would be subject to incorporation bias (where a variable serves both as an explanatory and as an outcome variable). Formally, the same could be said of alcoholic cirrhosis, but here the diagnosis rests in ultrasound and exclusion of other causes as well as alcohol history.

TABLE 13 Criteria for assigning cases to groups

Reference group	Subgroups and sub-subgroups
Group 1 (serious viral, genetic or autoimmune disease)	*Subgroup A, hepatocellular disease*
	Viral hepatitis B or C
	Haemochromatosis
	Wilson's disease
	Antitrypsin deficiency
	Autoimmune hepatitis
	Cirrhosis (alcohol or fat induced)
	Subgroup B
	PBC
	PSC
Group 2	Metastatic disease
	Paget's disease of bone
	Infectious diseases, such as hepatitis A, glandular fever, leptospirosis
	Thyroid disease
Group 3 (non-specific)	Echo-bright (fatty) liver on ultrasound
	Alcohol excess
	Overweight
	Alcohol + overweight
	Neither alcohol nor overweight
	No fatty liver
	Gilbert syndrome
	Persistence of LFT abnormality at 2 years
	LFT abnormality resolved at 2 years

Statistical methods

This section gives an outline of the methods and approaches used. Fuller details of individual methodologies are presented as appropriate in the results sections.

Variables and data

Demographic and lifestyle information – including body mass index (BMI) and alcohol consumption in units per week – was coded using categorical variables. Six categories each were used for age and alcohol consumption and four for BMI. The details may be read from *Table 14*.

Concentrations of analytes in the LFT panels were recorded (see *Table 4* for units) by the three individual laboratories. Laboratory-specific reference ranges, incorporating adjustments for age and sex (see *Table 4*), were used to categorise values as normal or abnormal. Thus, each LFT result was available in two forms: as a continuous variable (measured concentration) and as a dichotomous variable (normal/abnormal).

Liver fat on ultrasound was recorded on a four-point ordinal scale (normal, mild, moderate and severe). The condition 'fatty liver on ultrasound' was identified with the categories normal, mild and severe, and analysed as a binary variable.

Summaries of categorical variables (with percentages) are presented in tabular form. Summaries of analyte concentrations were expressed in terms of medians and quartiles.

Analysis of liver function test data

Abnormality

The presence of an abnormal analyte in the index panel was a criterion of entry to the study. Redundancy in the test panel was investigated by identifying subsets of analytes (i.e. subpanels) which would have recruited the highest proportions of study patients.

Analyte concentrations

Many of the analytes exhibited positive distributional skewness. Regression analyses of concentrations were conducted on log-transformed data. Differences in the distribution of results between laboratories were examined using quantile–quantile (Q–Q) plots and modelled using multiplicative factors (additive on the log-scale).

Pearson correlation analyses (using logged data standardised within laboratories) were carried out for different analytes in the same panel, and for individual analytes over time.

For each patient in the study, data were available from (up to) three LFT panels, recorded at different times. This gives an opportunity to analyse the development of patient readings over time as well as to relate results to demographic and diagnostic information. However, abnormality on the first panel is a criterion of entry to the study. It was anticipated that this feature would manifest itself in a 'regression to the mean' effect over the course of the study. Such selection effects could compromise the interpretation of any statistical analysis of the measured concentrations. The FU1 panels are the most complete panels in terms of missing data (certainly more complete than the FU2 data) and somewhat less biased by selection effects than the index panels, as they were not used as a criterion of entry to the study.

A time-series modelling approach was used to partition the variation in analyte concentrations between transient (short-term) components and persistent (long-term) components. The latter may be more relevant for the diagnosis of serious conditions. For this analysis, selection bias was handled by conditioning on the index LFTs. The variance explained by the persistent component in the model was compared with that from an analysis based on intrapatient correlations. Full details of the modelling methodology are found in *Appendix 2* (*BALLETS study analysis*) along with the results.

Stepwise regression procedures were used to model the impact of demographic and lifestyle factors on LFT results from the FU1 panel. The explanatory models obtained were used in subsequent analyses.

Diagnostic category and liver function test results

The relationship between diagnostic category and LFT results was investigated in two ways: (1) by adding diagnostic category to the explanatory models already derived and (2) by means of a multiple discriminant analysis. The discriminant analysis was carried out using a set patient-level variables and (logged) analyte concentrations that had been identified from a series of preliminary logistics regression analyses. The preliminary analyses involved separate stepwise logistic regressions designed to find significant predictors of individual disease categories. These predictors were then used in a multiple logistic discriminant (polytomous logistic regression) analysis between the separate diagnostic categories. The performance of the discriminant to distinguish between liver disease and a non-specific diagnosis was assessed using the area under a receiver operating characteristic (ROC) curve.

The stepwise element in the discriminant analysis described here was restricted to patients with a complete panel of LFTs at FU1. The diagnostic groups for serious liver disease were very small compared with the non-specific group, and further depleted in the complete case analysis. In order to make full use of the LFT data that were available from diseased patients, the analysis was repeated using a multiple imputation technique. Further details may be found in *Chapter 4* (see *Analysis of imputed data*).

Fatty liver

Stepwise logistic regression was used to explore the relationship between fatty liver at FU1 and patient characteristics, including (logged) LFT results, BMI and alcohol consumption. In this analysis, linear and quadratic components for age and alcohol consumption were substituted for the categorical variable for ease of interpretation. These components relate to the ordered categories themselves rather than the raw data. Persistence of fatty liver from FU1 to FU2 was investigated by stepwise logistic regression, including fatty liver at FU1 as a predictor for fatty liver at FU2. The consequences for liver fat of a change in BMI within an individual subject were investigated by means of two ordinal regression analyses: (1) liver fat category at FU2 on liver fat category at FU1 with percentage change in BMI as a covariate and (2) numerical difference in liver fat category between FU1 and FU2 on percentage change in BMI. Change in alcohol units (on a square root scale) was incorporated as a covariate in these analyses.

Sample size

A main objective of the study was to investigate the connection between LFTs and serious liver disease. However, there were several diseases under consideration, and no single primary question on which to power the study [see *Production of reference categories* (categories of *diagnostic groupings*)]. In the original protocol, logistic regression methods were proposed to explicate the relationship between diagnostic group and analyte concentrations. Sample size calculations for such problems often focus on 'events per variable' rules, which, in this study, suggest that 5 to 10 positive diagnoses would be needed for each predictor variable. Here there are seven independent LFTs (given that total protein is the aggregate of two other analytes), suggesting that between 35 and 70 positive cases would be needed for an unadjusted analysis. The study was designed for 1500 patients, which gives a satisfactory number of events (60) assuming a prevalence of a positive diagnosis of 4%. In practice, 44 cases of serious liver disease were found. If serious liver disease is considered as a composite outcome, the events per variable approach suggests that a reliable analysis is possible, at least for the five non-protein analytes in the LFT panel together with a small number of covariates.

The 'events per variable' approach focuses purely on technical aspects of logistic regression estimates. In the original protocol, we also considered a novel alternative criterion based on the ability of a logistic discriminant to identify high-risk cases (i.e. patients with risk of disease higher than an acceptable threshold level). For this purpose a baseline level of acceptable risk was taken to be 2% of the average population prevalence. According to this approach, 1500 patients would be sufficient to estimate a logistic discriminant function with a 90% chance of flagging up any patient whose true risk was twice the acceptable baseline level. These calculations posited an average prevalence of 4% – close to that actually observed in the sample. In retrospect, it seems that the degree to which the risk is predictable from the LFT results was underestimated in the original calculations, suggesting that the true performance of the discriminant would exceed expectations. However, there does not seem to be any direct way to verify this, as the true risk profile remains an unknown function of LFT results. In any event, this approach is concerned only with case finding and attaches no penalty to false-positives.

The other principal statistical analysis is concerned with the incidence of fatty liver. For this the event rates are much higher than for serious liver disease, and the sample numbers required for a meaningful analysis are correspondingly less stringent.

Summary of changes to the protocol

The study protocol can be found in *Appendix 1*.

Changes to protocol by section

- *Section 2.5.6 The patient's perspective.* Altered to describe the contents of psychology questionnaires.
- *Section 3.2.1 Practices.* Additional practices were recruited in order to improve and maintain recruitment to the study.
- *Section 3.3.2 Formal enrolment in subsequent testing protocol: defining of the patient population and seeking consent.* Changes to this section reflect alterations to the clinical process at each practice. The original clinical protocol was adapted to routine clinical practice.
- *Section 3.3.4 Collection of patient information.* Altered to indicate that psychology questionnaires would not be translated into languages other than English.
- *Section 3.3.6 Long term follow-up.* The 2-year follow-up phase in Birmingham was more complex than originally planned, including repeat USS, alcohol questionnaire and measurements. This section was altered to reflect changes to the process.
- *Section 3.5.1 Broad aim.* Altered to address the possibility of selection bias, which could occur when suitable patients declined to take part or when suitable patients were not selected by their GP to take part (see *section 3.5*).
- *Section 3.7.1 Psychological pilot study.* This pilot study was implemented to inform the development of psychological questionnaires for use in the main study. The study process was updated to indicate changes to measures and time points used for data collection.
- *Section 5.1 Substudy: Cryogenic storage and later testing* Approval was obtained to collect and store an anonymised blood sample from consenting Birmingham patients attending their 2-year (FU2) clinic appointment.
- *Section 5.2 Substudy: A qualitative investigation into liver function test ordering behaviour of general practitioners involved in the BALLETS study.* A substudy designed to examine the non-clinical motives behind a GP's decision to order an LFT.
- *Section 5.3 Substudy: Follow-up of abnormal test results.* In the course of the study some patients tested positive for some specific liver diseases, but many were not followed up according to the agreed algorithm for referral or further testing. Letters were prepared by the study hepatologist and chief investigator suggesting appropriate follow-up of individual study patients testing positive for particular diseases.
- *Section 5.4 Substudy*: *Qualitative investigation exploring anecdotal and preliminary evidence that events associated with participation in the BALLETS study were motivational to patients with and without fatty liver*. A random selection of study patients were interviewed, in response to anecdotal accounts from patients at 2-year follow-up clinics that they had implemented lifestyle changes following their first BALLETS clinic appointment.

Ethics committee approval for changes to the protocol

The main research ethics committee, St Thomas' Hospital Research Ethics Committee, gave favourable ethical opinion to the BALLETS study in April 2005.

During the recruitment and follow-up phases of the study, the St Thomas' Hospital Research Ethics Committee Modifications Subcommittee approved 10 substantial amendment applications

for alterations to the study protocol and documentation. Detailed summaries of each amendment are provided in the appendices to the main report (*Appendix 3*, *BALLETS study*: *summary of ethics and substantial amendment approval*). All amendments were also approved by South Birmingham and Lambeth local research ethics/research and development committees.

Approval was sought for the recruitment of new study practices in Birmingham, and corresponding patient documentation, for two qualitative substudies, as described in *Chapter 5* (see *Psychology 1*: *effects of positive tests* and *Psychology 2*: *effects of results on behaviour*), and for a more intensive 2-year follow-up phase for Birmingham patients, which included an additional USS and the cryogenic storage and later testing of cells and serum. In addition, approval was obtained for the study team to remind GPs by letter, of the need to follow up patients who tested positive for some specific liver diseases.

Chapter 4

Results: main study

This section begins with a brief description of the practices from which the participants were drawn and the demographics of the patients in the study (see *Nature of the population studied: Birmingham and Lambeth sites* and *Patients and practice characteristics*, below). Numerical summaries of the clinical data are also presented (see *Summary of clinical data*), namely LFT panels, diagnostic categories and ultrasound features. Some observations on the timing and completeness of panels are included here, together with a brief discussion of selection effects.

Analysis of the LFT panels themselves (see *Analysis of the liver function test panels*) is presented in two parts: (1) the inter-relationships between unadjusted LFT results and the utility of laboratory reference ranges for assessing abnormality (see *Analysis of unadjusted liver function test results*) and (2) variation in the concentrations of individual analytes, investigated using regression models to adjust for patient characteristics (see *Impact of patient characteristics on liver function test results*). The contribution of diagnostic grouping and fatty liver status to these models is also considered (see *Impact of diagnostic surgery* and *Impact of fatty liver*, respectively).

The section *Liver-related disease* contains a detailed clinical description of the categories of liver disease in the sample. *Diagnostic value of liver function tests* builds on the regression models (see *Impact of patient characteristics on liver function test results*) to investigate the diagnostic potential of the LFT panel, taking account of individual patient characteristics. The approach is based on stepwise procedures to find the analytes with the greatest discriminatory potential and uses imputation methods to cope with missing values in sparsely populated diagnostic categories.

Fatty liver is investigated in *Fatty liver on ultrasound*. The risk of fatty liver is modelled using logistic regression techniques. Relationships between fatty liver and lifestyle variables over the course of the study are investigated.

Nature of the population: Birmingham and Lambeth sites

In Birmingham, 1197 participants were recruited from eight practices. The location of the practices within the Birmingham conurbation and the socioeconomic and ethnic group characteristics of the surrounding areas are shown in *Figures 2* and *3*. In Lambeth, 147 participants were recruited from three practices. The location of the practices within the Lambeth conurbation and the socioeconomic and ethnic group characteristics of the surrounding areas are shown in *Figures 4* and *5*.

Patients and practice characteristics

Forty-six patients of the total 1344 patients were excluded because none of the original LFTs was, in fact, abnormal. Eight patients were excluded because the second blood test result was completely missing such that neither the FU1 LFT nor any of the tests for specific diseases was available (*Figure 6*). The analyses use data from the remaining 1290 patients, whose individual characteristics are summarised in *Table 14*. This includes basic demographic information

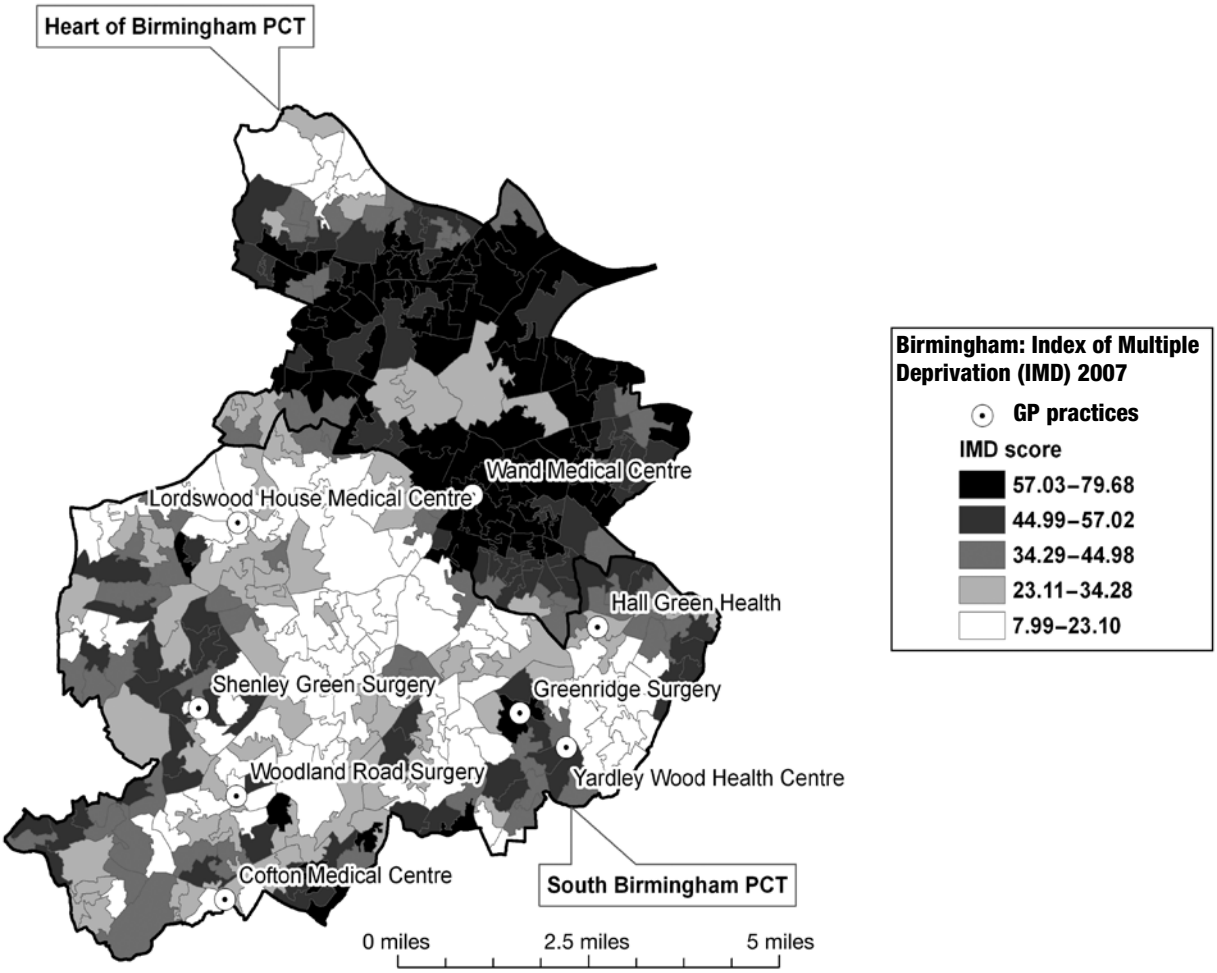


FIGURE 2 Location of participating practices in relation to socioeconomic group (Townsend Deprivation score quintile by lower-level super output area). PCT, primary care trust. Adapted from 2001 Census, Output Area Boundaries.[39] ©Crown Copyright 2003. Census output is Crown Copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

(recorded at FU1), together with BMI measurements (taken at FU1 and again at FU2) and results from the alcohol questionnaire (administered at FU1 and FU2). Results are given for all subjects (at FU1) and for the subsample who contributed to the FU2 LFT data. The reasons for ordering the index LFTs are summarised in *Table 15*.

Eleven general practices contributed to the study and their participation is summarised in *Table 16*. The first eight practices in the table are situated in Birmingham, and the remaining three in London.

Summary of clinical data

Diagnostic categories

Patients were categorised into three broad diagnostic groups, described more fully in *Chapter 3* (*Methods*: *main study*). Categories 1 and 2 are the two broad 'serious liver disease' categories, which may be further subdivided into: category 1a (hepatitis B and C + other hepatocellular diseases). Category 1b (hepatic bile duct disease) and category 2 [other diseases (such as metastatic cancer)] affecting the liver. The remaining cases (category 3) are rather non-specific.

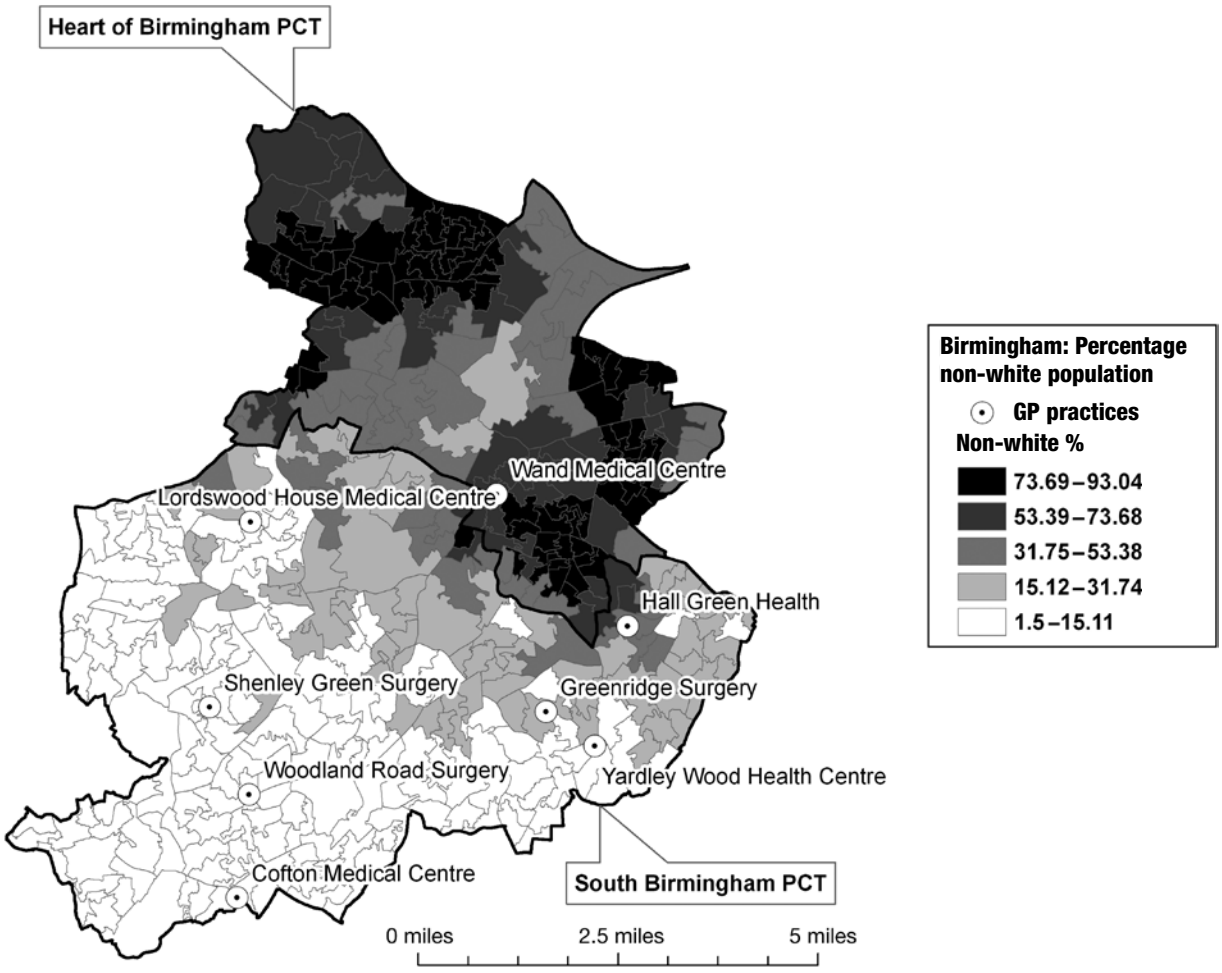


FIGURE 3 2001 Census location of participating practices in relation to ethnic mix (proportion of black and ethnic minority population by lower-level super output area). PCT, primary care trust. Adapted from 2001 Census, Output Area Boundaries.[39] ©Crown Copyright 2003. Census output is Crown Copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

Ultrasound features

Sonography reports for the liver were obtained at FU1 for 1277 patients and at FU2 for 658 out of 1152 patients from Birmingham practices. Second ultrasound examinations were not performed in Lambeth (see *Chapter 3*, *The 2-year follow-up visit*). A four-point ordinal scale (normal, mild, moderate and severe) was used to describe liver fat on each occasion (see *Chapter 3*, *Testing strategy for patients in the BALLETS study*), with results summarised in *Table 17*.

The subsequent version of the table (*Table 18*) shows the persistence of the ultrasound diagnosis of fatty liver between the two epochs.

Liver function test panels

Timing and completeness of panels

The LFT panel was extended from the usual five analytes to eight analytes for study purposes – that is to say the intention was to report on eight analytes on each occasion – the index result that triggered entry in the study and the FU1 and FU2 tests used as part of the comprehensive testing strategy. The number of analytes actually reported on each occasion is shown in *Table 19*. Complete reporting (all eight analytes) occurred for 915 (70.9%) on index testing, and 1168 (90.5%) at FU1 and 642 (81.3%) at FU2. A complete panel was recorded on the first

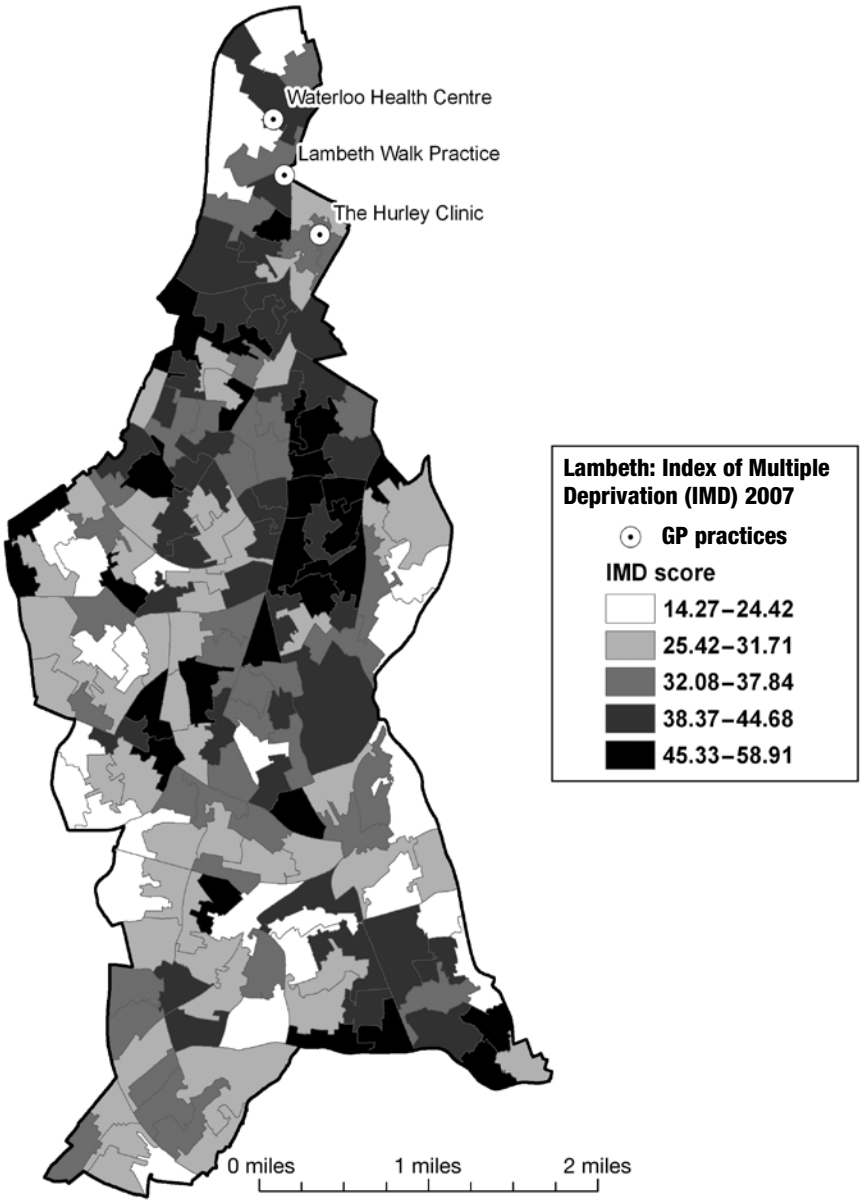


FIGURE 4 Location of participating Lambeth practices in relation to socioeconomic group (*English Indices of Deprivation* 2004). Adapted from 2001 Census, Output Area Boundaries.[39] ©Crown Copyright 2003. Census output is Crown Copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

two occasions (index and FU1) for 844 (65.4%) participants. Compared with the integral pilot data (see *Chapter 3*, *Laboratory methods*), completion improved for the follow-up visit but deteriorated for the index visit. A bimodal pattern of testing can be observed, with modes at eight analytes (as required for study purposes) and five analytes (the default situation) (see *Chapter 3*, *Missing data*).

At baseline, 1290 (100%) patients provided at least one LFT result. At FU1 this number fell to 1275 (98.8%) and at FU2 to 790 (61.2%). However, as described in *Table 19*, not all LFT panels were complete. *Table 20* shows the number of times each individual *analyte* was recorded as a percentage of the number of panels available.

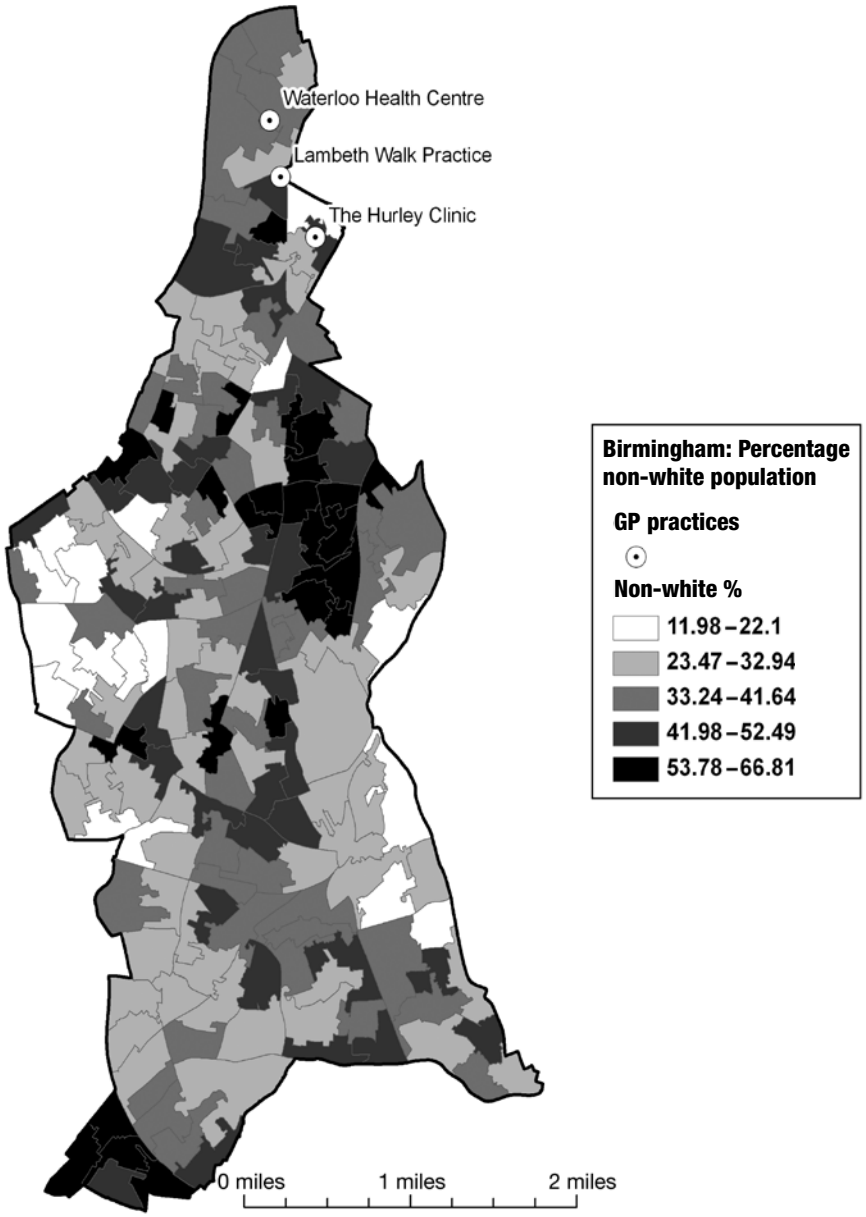


FIGURE 5 Location of participating Lambeth practices in relation to ethnic mix (proportion of black and ethnic minority population). Adapted from 2001 Census, Output Area Boundaries.[39] ©Crown Copyright 2003. Census output is Crown Copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

It was intended that FU1 occur as soon after the index LFT panel as might occur in practice, and that FU2 would occur after a further 2 years. The actual times that elapsed between the index and follow-up tests are summarised in *Table 21*.

The intended interval between FU1 and FU2 was 2 years. In the event, the median interval was almost exactly 2 years (23.9 months) with an interquartile range (IQR) of 21–27 months.

Abnormalities in the index liver function tests

The presence of some abnormality in the index panel was a main criterion for entry to the study. The number of analytes that were abnormal in the index panel is shown in *Table 22*.

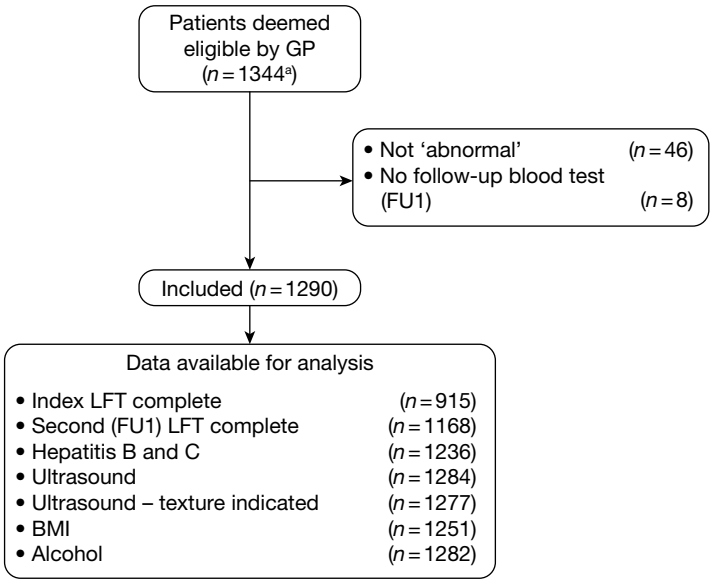


FIGURE 6 Flow diagram showing exclusions and data completeness. a, For an analysis of information on patients with abnormal LFTs but not recruited to the study, see *Chapter 3*, *Comparison of patients who were and were not 'recruited'*.

The extent to which analytes were abnormal, when abnormal, is summarised in *Table 23* in terms of average values expressed in units of the threshold of abnormality. Thus, for example, the median of the abnormally high ALT values is 1.37 times the upper limit of normal (ULN) (as defined by the laboratory concerned). It can be seen that the degree of abnormality is low in most cases. The exception is GGT, for which the corresponding median is 1.68.

Summary of liver function test data

The blood samples were processed by three different laboratories (see *Chapter 3*, *Selection of practices and patients*), labelled 1–3. Each laboratory operates its own reference ranges for the detection of abnormality. Mostly the differences between laboratories were slight, but for one analyte (ALP) the reported results for laboratory 1 followed a markedly different distribution from the other two laboratories (see *Appendix 2*, *Liver function test results by laboratory*). Given the potential for differences in laboratory practice, the summary statistics in *Table 24* have been computed separately for each laboratory (medians and quartiles).

The distribution of analyte concentrations is represented by the histograms in *Figures 7* and *8*. It is clear that the non-protein analytes exhibit substantial positive skewness. Accordingly, much of the analysis reported here deals with log-transformed LFT values, as discussed in *Appendix 2* (see *Liver function test results by laboratory*). One advantage of this approach is that a multiplicative laboratory effect can be readily incorporated as an additive term in any linear model for LFTs. As necessary (e.g. in the stepwise analyses of *Diagnostic value of liver function tests* and *Fatty liver on ultrasound*, below) the log-transformed data have been explicitly standardised to zero mean and unit standard deviation (SD) within laboratories.

Selection effects

Considered as a sample from a natural population, the index results are subject to selection bias, as an abnormal LFT was a criterion of entry to the study. It might be anticipated that the selection effects would be attenuated over time as the impact of abnormal results arising from short-term or chance effects dies away, at least in the group with no specific serious liver diagnosis. This phenomenon is investigated in *Table 25*. The first five 'non-protein' analytes (ALT, AST, bilirubin,

TABLE 14 Characteristics of patient participation in the BALLETS study

	All subjects ($n=1290$)		Subjects with 2-year follow-up LFTs ($n=790$)	
Characteristic	***n***	**%**	***n***	**%**
Sex				
Male	724	56.12	453	57.34
Female	566	43.88	337	42.66
Age (years)				
≤ 34	106	8.22	33	4.18
35–44	165	12.79	91	11.52
45–54	240	18.60	149	18.86
55–64	325	25.19	243	30.76
65–74	273	21.16	187	23.67
75+	181	14.03	87	11.01
Ethnic group				
White	1056	81.86	663	83.92
Asian	89	6.90	56	7.09
Black	66	5.12	33	4.18
Other	40	3.10	18	2.28
Not known	39	3.02	20	2.53
BMI (kg/m^2)[a]				
< 20	49	3.80	13	1.65
20–24.99	250	19.38	149	18.86
25–29.99	454	35.19	248	31.39
30+	498	38.60	294	37.22
Not known	39	3.02	86	10.89
Alcohol consumption (units per week)[a]				
0	547	42.40	282	35.70
1–14	352	27.79	251	31.77
15–29	153	11.86	90	11.39
30–49	122	9.46	53	6.71
50–99	84	6.51	39	4.94
100+	24	1.86	4	0.51
Not known	8	0.62	71	8.99

a Data for BMI and units of alcohol were collected initially (at FU1) and also after 2 years (at FU2). Thus, the 2-year data incorporate changes in BMI and drinking habits that may have occurred during the study.

ALP, GGT) all show a general reduction over time, which may be interpreted as 'regression to the mean'. Although the 2-year follow-up data are most likely to be free of LFT-related selection effects, it is comparatively incomplete (only 61.2% of patients) and more vulnerable to systematic dropout. For example, middle-aged patients are over-represented in the FU2 data. By contrast, FU1 was the primary focus of the data collection exercise, and is essentially complete, with 98.8% of patients represented.

The absence from the study of any patients with an index panel of normal LFTs will compromise assessments of the diagnostic value of LFTs. When considering diagnostic criteria based on conventional limits of abnormality, the number of 'negatives' (both true-negatives and false-negatives) will be underestimated. This means that many conventional measures of diagnostic performance will be biased. The direction of bias for some common measures is given in *Table 26*.

Notice that PPVs can be estimated without bias because they do not depend on negative results. It is plausible that NPVs would be underestimated in the study, but this cannot be established without further assumptions.

TABLE 15 Reasons for LFT ordering by category

Reasons for LFT ordering	Total
Investigations	
Abdominal symptoms or signs[a]	70
General symptoms or signs	318
Suspected alcohol abuse	18
Reviews	
CVD	53
Cholesterol	57
Hypertension	151
Diabetes	220
Medication	95
Other	308
Total	1290

CVD, cardiovascular disease.
a Excludes liver-specific symptoms, such as jaundice, right upper quadrant pain (or tenderness) and ascites.

TABLE 16 Patient participation by practice

	All subjects ($n=1290$)		Subjects with 2-year follow-up LFTs ($n=790$)		
Practice	*n*	%	*n*	%	Processing laboratory
Hall Green	161	12.48	117	14.81	a
Lordswood	213	16.51	134	16.96	a
Greenridge	195	15.12	103	13.04	a
Yardley Wood	144	11.16	100	12.66	a
Woodland Road	149	11.55	97	12.28	a
Cofton	126	9.77	76	9.62	a
Shenley Green	75	5.81	42	5.32	a
Wand	89	6.90	58	7.34	b
Lambeth Walk	71	5.50	31	3.92	c
Waterloo Health	48	3.72	26	3.29	c
The Hurley Clinic	19	1.47	6	0.76	c
Total	1290		790		

a University Hospital, Birmingham.
b Russells Hall Hospital, Dudley.
c Guy's and St Thomas' Hospital, London.

TABLE 17 Sonography results

First follow-up	Second follow-up						
	Normal	Mild	Moderate	Severe	Not known	DNA[a]	Total
Normal	324	44	10	1	1	413	793
Mild	62	61	28	0	1	111	263
Moderate	23	37	36	6	1	74	177
Severe	4	4	8	2	0	26	44
Not known	0	1	1	0	1	5	8
DNA	2	0	0	0	0	3	5
Total	415	147	83	9	4	632	1290

DNA, did not attend.
a Includes all Lambeth patients.

TABLE 18 Diagnosis of fatty liver on ultrasound[a]

	Two-year sonography		
Initial sonography	Normal (%)	Abnormal (%)	Total (%)
Normal	324 (85.49)	55 (14.51)	379 (100.00)
Abnormal	89 (32.84)	182 (67.16)	271 (100.00)
Total	413 (63.54)	237 (36.46)	650 (100.00)

a The entries are derived from the previous table by collapsing 'mild', 'moderate' and 'severe' into a single 'abnormal' category.

TABLE 19 Number of tests done at index visit, FU1 and FU2, with proportions (%) of patients participating at each stage

No. of tests	Index	FU1	FU2
0	0 (0.0)	15 (1.2)	–
1	3 (0.2)	3 (0.2)	1 (0.1)
2	6 (0.5)	3 (0.2)	1 (0.1)
3	6 (0.5)	23 (1.8)	26 (3.3)
4	103 (8.0)	16 (1.2)	42 (5.3)
5	134 (10.4)	21 (1.6)	29 (3.7)
6	99 (7.7)	17 (1.3)	25 (3.2)
7	24 (1.9)	24 (1.9)	24 (3.0)
8	915 (70.9)	1168 (90.5)	642 (81.3)
Total	1290	1290	790[a]

–, not applicable.
a All patients who attended FU2 and had a blood test.

TABLE 20 Analytes present in index and follow-up panels (as a percentage of numbers of available panels)

Analyte	Index panel (*n* = 1290)	FU1 panel (*n* = 1275)	FU2 panel (*n* = 790)
ALT	86.4	96.8	89.5
AST	89.8	95.1	91.6
Bilirubin	98.1	96.7	96.3
ALP	98.6	96.9	95.9
GGT	89.3	97.5	90.6
Albumin	99.1	98.4	98.5
Globulin	75.7	95.2	88.2
Total protein	76.1	96.9	89.5

TABLE 21 Elapsed time between LFT testing (months)

LFT testing	*n*	Minimum	Q1	Median	Q3	Maximum
Index to FU1	1288	0.1	0.7	1.0	1.7	9.0
FU1 to FU2	790	3.2	21.1	23.9	27.1	41.9
Index to FU2	790	4.1	22.3	25.3	28.4	43.4

Q, quartile.

TABLE 22 Number of analytes that were abnormal at index visit

No. of abnormal analytes	Total (%)
1	750 (58.1)
2	342 (26.5)
3	152 (11.8)
4	41 (3.2)
5	5 (0.4)
Total	1290 (100.0)

TABLE 23 Normal and abnormal analytes and extent of the abnormality expressed as a proportion of the upper (or lower) limit of normal for the laboratory concerned

		Normal	Below normal			Above normal		
Analyte	Total	*n*	*n*	Mean	Median	*n*	Mean	Median
ALT	1114	676	–	–	–	438	1.62	1.37
AST	1158	903	–	–	–	255	1.52	1.26
Bilirubin	1265	1117	–	–	–	148	1.44	1.26
ALP	1272	1083	–	–	–	189	1.30	1.16
GGT	1152	285	–	–	–	867	2.41	1.68
Albumin	1278	1248	9	0.87	0.91	21	1.04	1.04
Globulin	977	922	10	0.90	0.93	45	1.10	1.05
Total protein	981	884	4	0.85	0.91	93	1.04	1.03

TABLE 24a Index LFTs

	Laboratory 1				Laboratory 2				Laboratory 3			
Analyte	***n***	**Q1**	**Median**	**Q3**	***n***	**Q1**	**Median**	**Q3**	***n***	**Q1**	**Median**	**Q3**
ALT	898	22	32.5	50	79	31	48	65	137	22	35	55
AST	1049	23	28	38	79	29	40	53	30	27	33.5	44
Bilirubin	1049	6	9	13	79	6	8	11	137	9	12	20
ALP	1053	166	203	264	81	71	94	124	138	65	78	99
GGT	934	46	64	106	81	39	67	89	137	29	72	99
Albumin	1059	43	45	47	81	43	45	48	138	45	47	49
Globulin	863	27	29	32	71	28	31	34	43	26	29	32
Total protein	864	71	74	77	74	73	76	79	43	73	76	78

Q, quartile.

TABLE 24b First follow-up LFTs (FU1)

	Laboratory 1				Laboratory 2				Laboratory 3			
Analyte	***n***	**Q1**	**Median**	**Q3**	***n***	**Q1**	**Median**	**Q3**	***n***	**Q1**	**Median**	**Q3**
ALT	1021	22	30	46	84	29	45	60.5	129	20	31	44
AST	1018	22	27	36	82	28	35.5	47	112	25	32	39.5
Bilirubin	1018	6	9	13	84	6	8.5	11	131	7	10	16
ALP	1021	163	200	251	84	75	94.5	121	131	64	77	91
GGT	1027	38	59	99	88	34	58	92.5	128	27.5	52.5	92.5
Albumin	1039	44	46	48	84	43.5	46	48	131	45	47	49
Globulin	1015	27	30	33	84	28	31	33	115	26	29	31
Total protein	1027	73	76	79	88	73	77	80	120	73	76	78.5

Q, quartile.

TABLE 24c Two-year follow-up LFTs (FU2)

	Laboratory 1				Laboratory 2				Laboratory 3			
Analyte	***n***	**Q1**	**Median**	**Q3**	***n***	**Q1**	**Median**	**Q3**	***n***	**Q1**	**Median**	**Q3**
ALT	588	19	27	39	56	30.5	39	51	63	22	32	55
AST	640	21	26	32	28	24.5	34	39	56	26	34.5	52.5
Bilirubin	640	6	8	11	58	5	8	11	63	7	9	14
ALP	640	161	193	240	55	76	85	117	63	60	71	97
GGT	613	34	54	90	40	29	49.5	68.5	63	27	44	96
Albumin	659	44	46	48	56	44	45	47	63	46	47	49
Globulin	613	25	28	31	33	27	31	35	51	27	29	32
Total protein	623	71	74	77	33	73	77	81	51	74	76	78

Q, quartile.

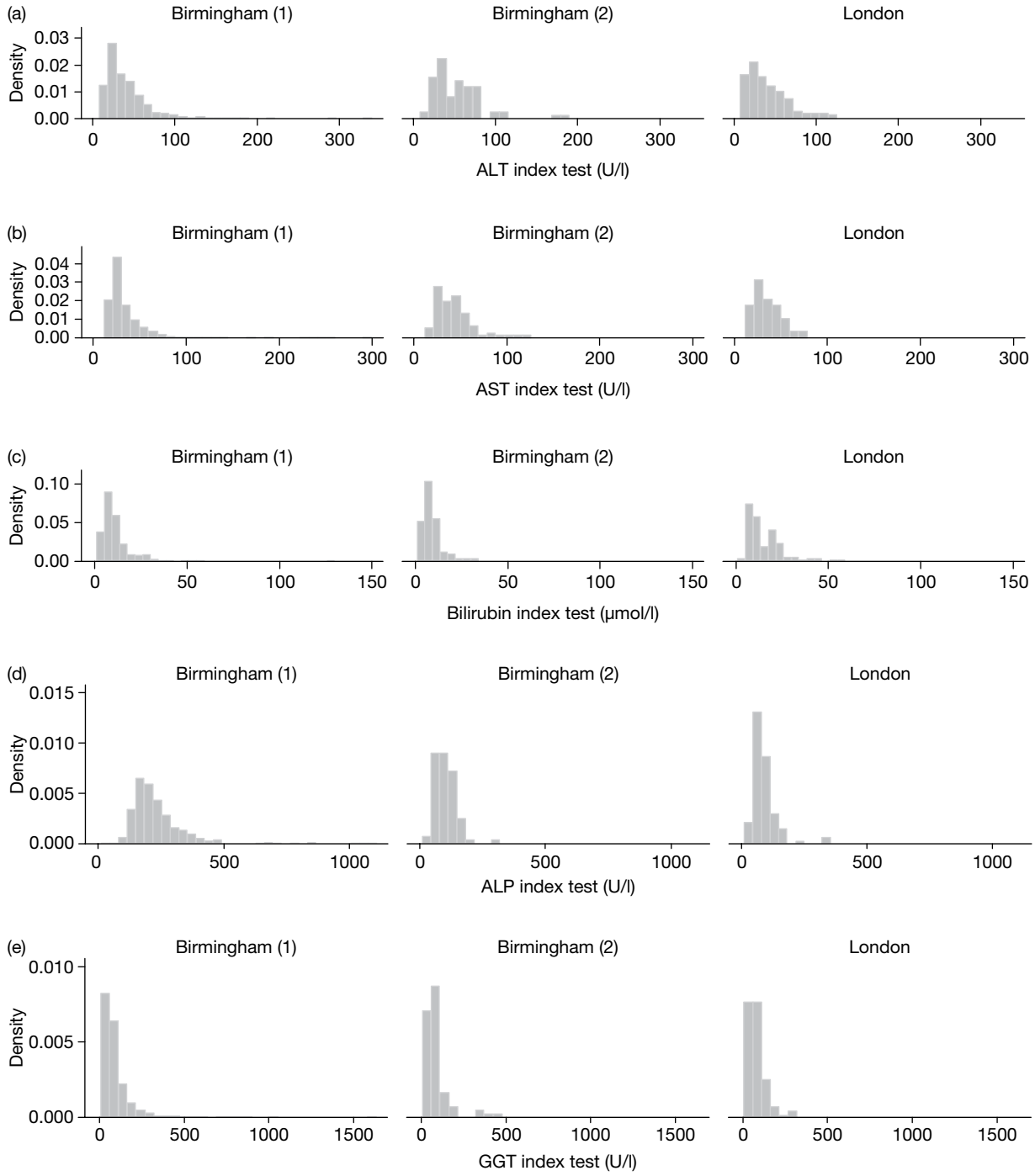


FIGURE 7 Distribution of enzyme analytes from index LFTs (by laboratory).

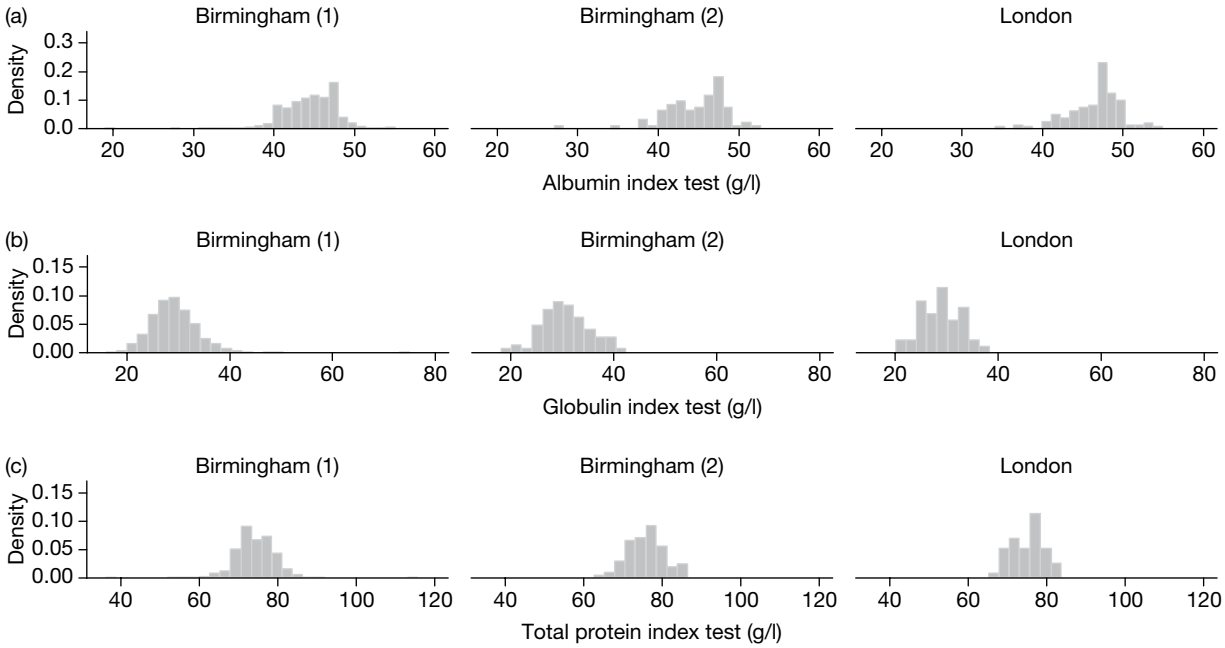


FIGURE 8 Distribution of protein analytes from index LFTs (by laboratory).

TABLE 25 Summary statistics for LFT values over time (adjusted for multiplicative laboratory effects and standardised to laboratory 1)[a]

	All subjects					
Analyte	***n***	**Minimum**	**Q1**	**Median**	**Q3**	**Maximum**
ALT						
Index	1114	8	22	33	50	329
FU1	1234	6	21	31	45	534
FU2	707	6	19	27	40	170
AST						
Index	1158	12	23	28	38	299
FU1	1212	11	22	27	36	248
FU2	724	10	21	26	33	152
Bilirubin						
Index	1265	1	6	9	13	130
FU1	1233	1	6	9	13	62
FU2	761	1	6	8	11	49
ALP						
Index	1272	11	165	204	264	1075
FU1	1236	46	164	201	252	1105
FU2	758	16	161	192	241	1340
GGT						
Index	1152	7	45	65	109	1582
FU1	1243	8	37	60	101	2890
FU2	716	7	33	54	92	683
Albumin						
Index	1278	19	43	45	47	55
FU1	1254	19	44	46	47	56
FU2	778	28	44	46	48	145
Globulin						
Index	977	16	27	29	32	75
FU1	1214	4	27	30	33	112
FU2	697	16	25	28	31	47
Total protein						
Index	981	36	71	74	77	113
FU1	1235	39	73	76	79	162
FU2	707	54	71	74	77	90

Q, quartile.

a The non-protein analytes exhibit a general reduction over time.

TABLE 26 Biases of diagnostic performance measures for LFTs in the BALLETS study

Measure	Definition	Direction of bias in BALLETS study
TPR = sensitivity	$\frac{TP}{TP+FN}$	Overestimate
FPR	$\frac{FP}{FP+TN}$	Overestimate
FNR	$\frac{FN}{TP+FN}$	Underestimate
TNR = specificity	$\frac{TN}{FP+TN}$	Underestimate
PPV	$\frac{TP}{TP+FP}$	Unbiased
NPV	$\frac{TN}{TN+FN}$	Unclear

FNR, false-negative rate; FPR, false-positive rate; NPV, negative predictive value; TNR, true-negative rate; TPR, true-positive rate.

Analysis of liver function test panels

Analysis of unadjusted liver function test results

Patterns of abnormality within the index panel

One of our aims was to evaluate test redundancy and, more generally, investigate which tests tend to group together when more than one test is abnormal. In *Table 27* we analyse abnormalities between index tests.

The entries in *Table 27* estimate the sensitivity (in the usual diagnostic sense) of using the analyte defined by the row, to detect abnormalities in the analyte defined by the column. From this point of view, ALT alone is superior to AST alone because it misses relatively few cases of AST abnormality, whereas AST would miss more than half of the cases of abnormality in ALT. Where ALT is abnormal, there is a high chance that GGT will also be abnormal. However, GGT is frequently abnormal when ALT is not. There is relatively little overlap between an abnormal ALT level and abnormal bilirubin, ALP or protein levels. Interestingly, an abnormal bilirubin was associated with abnormal GGT (although the reverse was not true). A raised bilirubin was *not* strongly associated with abnormal ALP and vice versa.

It is clear that GGT is the best candidate if a single analyte is to be chosen to detect abnormality in other analytes. Not only does it provide the greatest individual rate of abnormality in this group, but also it finds a substantial proportion of abnormalities in other analytes too.

Nevertheless, GGT misses nearly one-quarter of abnormal cases. Other analytes must be considered if this shortfall is to be addressed. The effects of removing analytes from the standard panel are investigated in *Table 28*. The best subsets of analytes (i.e. those that flag up the greatest number of cases) of given size have been obtained using the 915 complete index panels of LFTs.

In our sample, it appears that > 90% of abnormal cases can be identified by recourse to a panel of only three analytes. The table also suggests a hierarchy of analytes in

TABLE 27 Patterns of abnormality in the index LFTs[a]

		Analyte							
Analyte	% abnormal	ALT	AST	Bilirubin	ALP	GGT	Albumin	Globulin	Total protein
		39.3	22.0	11.7	14.9	75.3	2.3	5.6	9.9
ALT	39.3	1.00	0.88	0.22	0.30	0.37	0.23	0.18	0.31
AST	22.0	0.44	1.00	0.15	0.22	0.18	0.08	0.11	0.15
Bilirubin	11.7	0.06	0.06	1.00	0.04	0.05	0.17	0.06	0.05
ALP	14.9	0.09	0.15	0.05	1.00	0.10	0.13	0.13	0.08
GGT	75.3	0.71	0.72	0.33	0.64	1.00	0.48	0.49	0.54
Albumin	2.3	0.01	0.01	0.03	0.02	0.01	1.00	0.07	0.08
Globulin	5.6	0.02	0.03	0.04	0.06	0.03	0.24	1.00	0.37
Total protein	9.9	0.08	0.08	0.06	0.07	0.07	0.47	0.65	1.00

a The entries are the proportions of patients with abnormality in the row-analyte given that the column-analyte is abnormal. For example, the proportion of abnormal ALTs among patients whose AST is abnormal is 0.88; the proportion of abnormal ASTs among those with abnormal ALTs is 0.44.

TABLE 28 Cases detected by subsets of the index LFT panel with the greatest sensitivity[a]

No. of analytes	Best subset (with next best alternative)	Cases detected by best (and next best) subset	Sensitivity (%) of best (and next best) subset
1	GGT (ALT)	687 (352)	75.1 (38.5)
2	GGT + ALT (AST)	795 (738)	86.9 (80.7)
3	GGT + ALT + bilirubin (total protein)	843 (832)	92.1 (90.9)
4	GGT + ALT + bilirubin + total protein (ALP)	877 (875)	95.8 (95.6)
5	GGT + ALT + bilirubin + total protein (globulin) + ALP	905 (895)	98.9 (97.8)
6	GGT + ALT + bilirubin + total protein + ALP + globulin	909	99.3
	GGT + ALT + bilirubin + total protein + ALP + AST	909	99.3
7	GGT + ALT + bilirubin + total protein + ALP + globulin + AST	913 (911)	99.8
8	GGT + ALT + bilirubin + total protein + ALP + globulin + AST + albumin	915	100.0

a For each fixed number of analytes, the subset that identifies the largest number of patients as abnormal has been obtained. The next best alternative is also indicated (in parentheses). For example, for five analytes the best subset is GGT, ALT, bilirubin, total protein and ALP. The next best is obtained on replacing total protein with globulin. The analysis is restricted to the 915 patients with a complete index panel.

decreasing order of importance, as the best subsets obtained from an increasing sequence (GGT > ALT > bilirubin > total protein ≥ ALP). Between them, these five account for nearly 99% of abnormal cases in our sample. If formal 'abnormality' – as defined within standard laboratory practice – is the only criterion then the remaining analytes – AST, globulin and albumin – may be seen as redundant. However, some caution is indicated here, given that no allowance has been made for sampling variability. Moreover, the detection of analyte abnormality may be a relatively minor concern in practice; the diagnostic value of individual analytes for predicting liver disease is of much greater importance when considering which tests might be dropped from the panel.

Patterns of abnormality and disease classes

The predictive value of index abnormality for individual analytes, and pairs of individual analytes, is investigated in *Table 29*. The table contains the percentage of patients in each liver disease category among those who have registered an abnormality in the analyte concerned, or on at least one member of a pair of analytes. It is interesting to compare these predictive values

with prevalence of liver disease among the subsample of patients with a complete index panel of eight analytes. This is given in the first row of the table, and represents the PPV of the LFT panel.

One striking feature of the table is the poor predictive performance of GGT. It is outperformed by ALP in all disease categories and by ALT and AST for category 1a.

Positive predictive value may not be the most important criterion of diagnostic performance because a high value can be achieved despite missing a large fraction of cases present. However, it is the one measure that can be properly estimated from this study as our data relate to a comprehensive set of patients' abnormal LFTs.

True-positive rates can be estimated, but these can be applied only to a restricted population with an abnormal LFT panel. Of course, there may be cases of liver disease whose LFT panel happens to be normal. These estimates are shown in *Table 30*.

From *Table 30* it appears that the combination of GGT and ALT gives the best overall disease coverage, identifying nearly 96% of all cases of serious liver disease.

The limits of abnormality used to define a positive diagnosis in *Tables 29* and *30* are based on conventional reference ranges. These are determined so as to reflect the limits of "normal" physiology, but without regard to the impact of specific diseases. Because the BALLETS index panels include all patients with a positive LFT result (as conventionally determined), they can be used to investigate the effect of increasing the upper thresholds of abnormality on the numbers of positive diagnoses. If this approach is taken then the number of positives will inevitably fall, as some cases will be missed by the more stringent criterion. For an analyte which carries diagnostic information, it is to be expected that the ratio of true-positives to false-positives (the PPV) will increase as the threshold rises above the normal limit. This is investigated in *Table 31* for thresholds set at twice the normal limit. A range of thresholds is considered in *Figure 9*: here the curves for ALT, AST and ALP lie consistently above the diagonal line, confirming that the ratio of true-positives to false-positives does in fact increase with increasing thresholds. Indeed the slopes of these curves rise sharply as they approach the origin, suggesting that a very high analyte concentration has very high predictive value for liver disease. For GGT the situation is less clear-cut. The ratio of true-positives to false-positives remains effectively constant as the threshold increases even to twice the conventional limit, rising only as it approaches a threefold increase. This observation may cast doubt on the value of GGT as a marker of liver disease in an unselected population, or at least suggest that its full diagnostic contribution is not captured by conventional reference ranges.

Diagnostic performance of alternative liver function test panels

As already noted (see *Selection effects*), the selected nature of the sample – confined, as it is, to subjects with at least one abnormal analyte – precludes direct estimation of absolute sensitivities and specificities of the LFT tests as markers of liver disease. However, the relative diagnostic performance of alternative LFT panels can be assessed by comparing the numbers of true-positives and false-positives in the index sample. These quantities are plotted in *Figure 10* for each of the 255 ($= 2^8 - 1$) possible LFT panels that can be constructed using eight analytes. In this analysis, 'liver-related disease' is defined broadly to include all group 1 and 2 diseases. The data are restricted to the 915 subjects with a complete set of index LFTs.

One panel can be said to dominate another if it generates both more true-positives and fewer false-positives than its competitor. On this basis, the best candidates are those that are not dominated by any other panel. This set of candidates is well approximated by the panels on the frontier in *Figure 10*, which involve three analytes only: ALT, ALP and GGT. In theory, the overall

TABLE 29 Positive predictive values for specific disease categories (i.e. percentage of patients with abnormalities who have disease) for individual analytes (and pairs of analytes) in the index panel

Analyte combination	Liver disease 1a	Liver disease 1b	Liver disease 2	Any liver disease: 1a + 1b + 2	No. abnormal
Complete panel	2.5	0.8	1.5	4.7	891
Single analytes					
ALT	4.3	0.7	1.0	6.0	418
AST	5.6	1.2	0.8	7.7	248
Bilirubin	3.7	0.7	0.0	4.4	135
ALP	2.7	4.9	2.2	9.8	184
GGT	2.2	1.0	1.7	4.8	833
Albumin	3.6	0.0	0.0	3.6	28
Globulin	3.6	0.0	0.0	3.6	55
Total protein	4.3	0.0	0.0	4.3	94
Pairs of analytes					
ALT *or* AST	4.6	1.0	0.8	6.3	505
ALT *or* bilirubin	4.2	0.8	0.8	5.7	527
ALT *or* ALP	4.1	1.8	1.2	7.1	564
ALT *or* GGT	2.7	0.8	1.5	5.0	963
ALT *or* albumin	4.3	0.7	0.9	5.9	442
ALT *or* globulin	4.3	0.6	0.9	5.8	464
ALT *or* total protein	3.9	0.6	0.8	5.3	486
AST *or* bilirubin	4.4	1.1	0.5	6.0	367
AST *or* ALP	4.1	2.3	1.3	7.6	394
AST *or* GGT	2.4	1.1	1.5	5.0	942
AST *or* albumin	5.5	1.1	0.7	7.3	274
AST *or* globulin	5.4	1.0	0.7	7.1	297
AST *or* total protein	4.6	0.9	0.6	6.1	328
Bilirubin *or* ALP	2.9	2.9	1.3	7.1	312
Bilirubin *or* GGT	2.3	1.0	1.5	4.7	930
Bilirubin *or* albumin	3.2	0.6	0.0	3.8	158
Bilirubin *or* globulin	3.2	0.5	0.0	3.7	187
Bilirubin *or* total protein	3.6	0.4	0.0	4.0	224
ALP *or* GGT	2.1	1.3	1.6	5.0	933
ALP *or* albumin	2.9	4.3	1.9	9.1	208
ALP *or* globulin	3.0	3.9	1.7	8.6	232
ALP *or* total protein	3.3	3.3	1.5	8.1	270
GGT *or* albumin	2.2	0.9	1.6	4.8	851
GGT *or* globulin	2.3	0.9	1.6	4.9	864
GGT *or* total protein	2.3	0.9	1.6	4.8	880
Albumin *or* globulin	2.5	0.0	0.0	2.5	79
Albumin *or* total protein	4.3	0.0	0.0	4.3	115
Globulin *or* total protein	4.4	0.0	0.0	4.4	113

TABLE 30 'True-positive rates', i.e. percentage of abnormal results within disease groups[a]

Analyte combination	Disease 1a (*n* = 32)	Disease 1b (*n* = 12)	Disease 2 (*n* = 15)	1a + 1b + 2 (*n* = 59)	No. of patients
Single analytes					
ALT	66.7	37.5	28.6	51	1064
AST	48.3	25	14.3	34.5	1131
Bilirubin	16.1	8.3	0	10.5	1213
ALP	16.1	75	28.6	31.6	1220
GGT	64.3	100	100	80	1101
Albumin	3.1	0	0	1.7	1226
Globulin	8.7	0	0	4.4	951
Total protein	17.4	0	0	8.9	955
Pairs of analytes					
ALT *or* AST	79.2	37.5	30.8	57.8	966
ALT *or* bilirubin	76.9	37.5	30.8	57.4	1054
ALT *or* ALP	76.9	75	46.2	68.1	1057
ALT *or* GGT	92.3	100	100	95.8	1051
ALT *or* albumin	70.4	37.5	28.6	53.1	1062
ALT *or* globulin	87	28.6	28.6	59.1	916
ALT *or* total protein	82.6	28.6	28.6	56.8	917
AST *or* bilirubin	51.7	33.3	14.3	38.2	1125
AST *or* ALP	55.2	75	35.7	54.5	1125
AST *or* GGT	76.9	100	100	87.2	1005
AST *or* albumin	51.7	25	14.3	36.4	1125
AST *or* globulin	54.5	12.5	15.4	34.9	927
AST *or* total protein	50	12.5	15.4	32.6	930
Bilirubin *or* ALP	29	75	28.6	38.6	1212
Bilirubin *or* GGT	70.4	100	100	83.3	1080
Bilirubin *or* albumin	16.1	8.3	0	10.5	1212
Bilirubin *or* globulin	18.2	12.5	0	11.6	939
Bilirubin *or* total protein	27.3	12.5	0	16.3	939

a Denominators here are the observed numbers in disease groups, but these are incomplete because the data relate only to abnormal LFT panels.

TABLE 31 Effect on PPV of doubling the upper threshold of normality[a]

	Standard thresholds		Twice standard thresholds	
Analyte	True-positives (*n*)	PPV (%)	True-positives (*n*)	PPV (%)
ALT	25	6.0	9	11.7
AST	19	7.7	4	14.3
Bilirubin	6	4.4	0	–
ALP	18	9.8	3	20.0
GGT	40	4.8	15	4.7

a Positives encompass any liver disease (1a + 1b + 2).

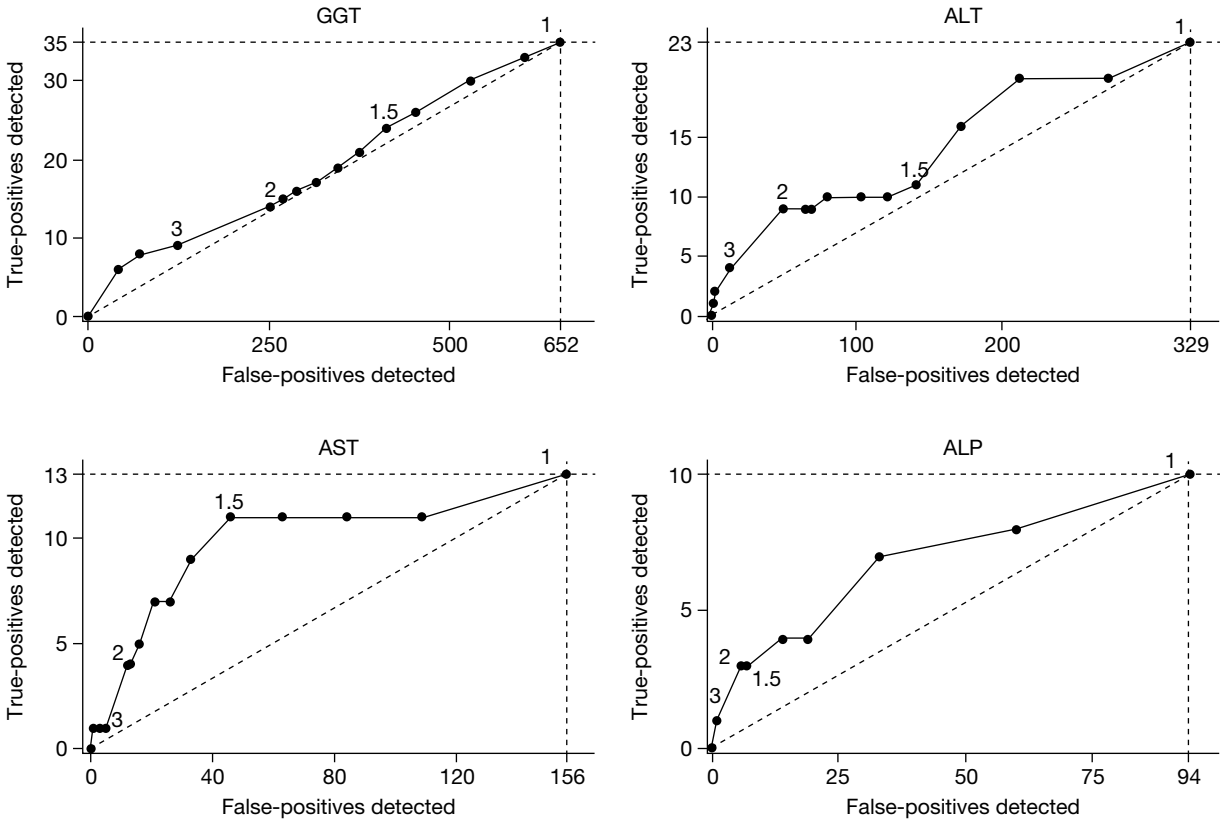


FIGURE 9 The effect of increasing the threshold of abnormality for four analytes. Numbers of true-positives (i.e. patients in category 1 or category 2 with analyte concentration above the threshold) are plotted against numbers of false-positives, using thresholds set at fixed multiples of the current laboratory reference limit. Points are plotted at intervals of 0.1 up to twice the reference limit, and at 3, 4 and 5 times the limit. Points at 1, 1.5, 2 and 3 times the limit are labelled accordingly.

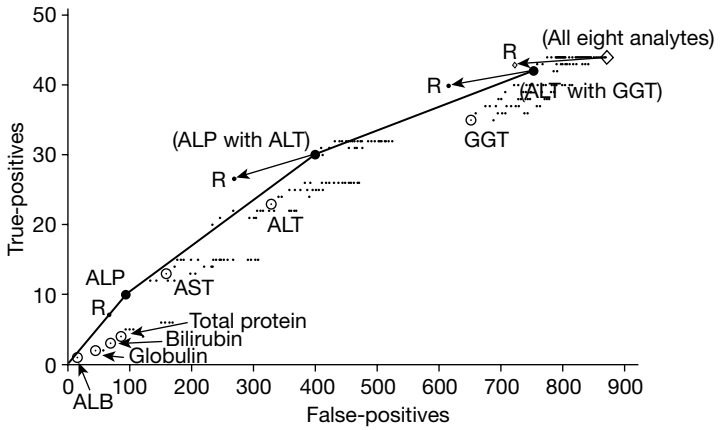


FIGURE 10 Positive diagnoses from different LFT panels, split between the non-specific group (false-positives) and the pooled disease groups (true-positives). All 255 possible panels from the eight analytes are shown. Single analyte panels (open circles) and the complete panel of eight analytes (diamond) are identified. The frontier (solid circles joined by line segments) shows the best diagnostic performance that can be attained using the analytes ALP, ALP and GGT. Results from repeating a panel at follow-up if it is positive initially are indicated by the letter 'R' joined by an arrow to the initial panel.

'best' panel would follow on specifying the trade-off between the value of a true-positive and the cost of a false-positive, and comparing this ratio with the slope of the line segments that go to make up the frontier in the figure. Using the three 'frontier' analytes only, the best panel would be determined as follows:

- *Value–cost ratio < 10* It is best not to use routine LFTs for these patients.
- *Value–cost ratio between 10 and 15* The best LFT panel is ALP alone.
- *Value–cost ratio between 15 and 30* The best panel is ALP with ALT.
- *Value–cost ratio > 30* The best panel is ALT with GGT.

The panels mentioned here delineate a plausible range of options, although, in practice, it may not be straightforward to specify an appropriate value–cost ratio.

The PPV of a panel depends on the ratio of true-positives to false-positives. For the 915 complete index panels that contribute to *Figure 10* it ranges from 9.6% for ALP alone, through 7.0% (ALP and ALT) to 5.3% (ALT and GGT). The PPV of the eight-analyte panel is 4.8%.

The results above suggest that the functions of a routine LFT panel can be largely subsumed into just three analytes: ALT, ALP and GGT. Nevertheless, an LFT panel – however carefully chosen – will never function as a definitive diagnostic procedure for any particular pathology; symptomatic patients will continue to be monitored in a primary-care setting regardless of a 'negative' test result. In this context, a panel with high PPV is perhaps most useful to the clinician, particularly if a plausible biological interpretation of positive results can suggest a route towards a definitive diagnosis. For these reasons, we incline towards the recommendation that routine testing should be confined to ALP and ALT alone. Addition of GGT will certainly increase the number of positive results but will not necessarily help specify a future clinical pathway for the patient. As remarked above (see *Figure 9*) the PPV of an abnormal GGT result does not increase when a higher threshold of abnormality is used. This is not suggestive of a strong clinical effect.

Repeat testing

There is a natural impulse to repeat a positive test to see if it is confirmed, particularly if the test has low specificity. It appears that repeating the full eight-analyte panel is of little value, as it achieves results similar to a single administration of the two-analyte panel (ALT + GGT) (see *Figure 1*). In our sample, repeating a positive ALP result does not affect the PPV, although the number of positive results that survive is reduced by 30%. For both ALP + ALT and ALT + GGT, repeat testing improves the PPV and may be recommended.

In summary, the analysis in this section points to a reduction in the number of analytes in the routine panel, to ALT + ALP, leaving open the possibility of repeat testing should the initial result be abnormal.

Correlation analysis of liver function tests

Table 32 shows the correlations between all pairs of LFTs over all three panels (index, FU1 and FU2). These were obtained from the log-transformed data, standardised within laboratories and panels to have zero mean and unit standard deviation.

Each correlation coefficient in *Table 32* derives from more than 2500 pairs of observations. Even although these originate from only 1290 patients (implying a lack of independence between pairs), statistical significance ($p < 0.05$) can be claimed for any coefficient > 0.05 in absolute magnitude. On this basis, most of the coefficients in *Table 32* (22 out of 28) represent

significant relationships between analytes. However, most of these correlations are too small to be of practical interest. Disregarding the structural correlations involving total protein and its constituent components (albumin and globulin), the only analytes with mutual correlation above 0.3 (i.e. sufficient to account for around 10% of each other's variation) are ALT, AST and GGT. ALT and AST are the most highly correlated, explaining about 60% of one another's variation.

The proportion of the variation in each analyte that can be explained by the others (i.e. an R^2 value from a regression model) is presented in *Table 33*, using FU1, which is the most comprehensive data set. The values have been obtained by regression of the log-transformed LFTs on all other (log-transformed) analytes, after adjustment for additive laboratory effects. This is equivalent to fitting multiplicative laboratory effects to the raw data, as discussed above (see *Summary of LFT data*) and also *Appendix 2*, *Liver function test results by laboratory*). Total protein has been excluded from these analyses.

From *Table 33* it appears that around 65% of the variation in ALT and AST can be explained by regression on the other analytes, which still leaves 35% unexplained. For all other analytes the explained variation is < 25% of the total. This analysis has some bearing on the question of test redundancy in that no LFT can be dropped from the panel without substantial loss of information. We shall return later to the question of whether or not the information that would be lost in dropping a particular LFT from the panel is relevant to any significant clinical decision.

TABLE 32 Correlations between LFTs from the same panel[a]

Analyte	ALT	AST	Bilirubin	ALP	GGT	Albumin	Globulin	Total protein
ALT	1.000							
AST	0.773	1.000						
Bilirubin	0.068	0.150	1.000					
ALP	–0.047	0.001	–0.204	1.000				
GGT	0.363	0.379	–0.103	0.199	1.000			
Albumin	0.150	0.058	0.162	–0.206	–0.015	1.000		
Globulin	0.032	0.099	–0.067	0.152	0.072	–0.146	1.000	
Total protein	0.122	0.120	0.049	0.014	0.064	0.485	0.778	1.000

a Results pooled over index and both follow-up panels.

TABLE 33 Variation in individual LFT values at FU1 explained by other analytes in the panel, adjusted for laboratory effects[a]

Analyte	R^2 from best linear predictor (%)	Analyte	R^2 from best single predictor (%)	Analytes	R^2 from best pair of predictors (%)
ALT	65.4	AST	62.0	AST, albumin	63.5
AST	66.4	ALT	62.0	ALT, GGT	66.4
Bilirubin	9.5	ALP	3.6	ALP, AST	5.6
ALP	13.5	Bilirubin	3.6	Bilirubin, GGT	7.5
GGT	24.5	AST	16.9	AST, ALP	19.6
Albumin	8.3	ALP	3.6	ALP, ALT	5.3
Globulin	5.8	ALP	2.8	ALP, AST	4.4

a The first column was obtained by multiple regression on all other LFTs; the predictors in columns 2 and 3 were identified using a forward selection procedure.

Associations over time

Persistence of abnormality from index to first follow-up

We analysed the proportion of cases where the FU1 panel of LFTs was abnormal given that an abnormal index test was the entry criterion for the study. This analysis is restricted to the 844 participants for whom a full panel of eight analytes was available on both occasions (index and FU1). Only 138 (16%) had a normal repeat (FU1) test result and the breakdown by test is given in *Table 34*. The proportion of normal repeat panels at FU1, about 1 month after the index test, is < 10% if three tests are abnormal, bilirubin is abnormal or if two tests are abnormal and one of the two is GGT. An isolated abnormal ALT has a rather high chance of reverting to normal (43%), whereas the repeat panel is seldom normal if GGT is raised (16%).

There were 52 patients with serious liver disease and a complete panel of repeat LFTs at FU1 (28 in category 1a; 10 in category 1b; 14 in category 2). Of these, only two had reverted to normal (one each in categories 1b and 2).

Persistence of abnormality at 2 years

Of the 1168 complete panels available at FU1, 176 (15.1%) had reverted to normal. At FU2, only 577 complete panels were available, of which 176 (30.5%) were normal. The pattern of abnormality for each analyte over time is summarised in *Table 35*.

Correlations between epochs

Correlations between measurements of the same analyte within individual patients are presented in *Table 36*.

The index and FU1 panels were taken close together in time (median interval 1 month), and it is therefore not surprising to see some high correlations in the first column of *Table 36*. The median interval between the index test and FU2 was 25.3 months. It is notable that the temporal correlation remains relatively high for all analytes over such a time interval. In practice, there was considerable variation in the timing of the FU1 and FU2 panels (see *Table 21*) for different patients. This feature is exploited in the analysis of *Appendix 2* (see *Temporal modelling of liver function tests*), which investigates the persistence of LFT results over time and seeks to identify the proportion of the variance in LFTs that is due to genuine differences between patients.

Variation between and within patients

For any particular patient, the measured concentration of an analyte from the LFT panel will be subject to temporal fluctuations. Thus, only part of the variation in the recorded LFT panels can be attributed to genuine differences between patients. A simple estimate of the proportion of the

TABLE 34 Comparison of abnormality at initial and subsequent tests[a]

Analyte(s)	Abnormal index test, *n*	Same abnormality at FU1, *n* (%)	Different abnormality at FU1, *n* (%)	Normal panel at FU1, *n* (%)
ALT alone	50	17 (34.0)	12 (24.0)	21 (42.0)
Bilirubin alone	39	22 (56.4)	5 (12.8)	12 (30.8)
GGT alone	328	213 (64.9)	65 (19.8)	50 (15.2)
ALT + GGT	103	42 (40.8)	52 (50.5)	9 (8.7)
ALT + AST	35	10 (28.6)	16 (45.7)	9 (25.7)
ALT + AST + GGT	75	21 (28.0)	49 (65.3)	5 (6.7)
ALP + GGT	32	17 (53.1)	12 (37.5)	3 (9.4)

a This analysis is restricted to those who had all eight tests at index and subsequent (FU1) visit. The most frequently encountered abnormalities or combinations are itemised individually ($n = 844$).

TABLE 35 Persistence of abnormality over time for individual analytes[a]

	Baseline/first follow-up, *n* (%)			
2-year follow-up	Normal/normal	Normal/abnormal	Abnormal/normal	Abnormal/abnormal
ALT				
Normal	310 (92.8)	14 (73.7)	60 (73.2)	71 (47.7)
Abnormal	24 (7.2)	5 (26.3)	22 (26.8)	78 (52.3)
AST				
Normal	447 (94.5)	21 (84.0)	50 (80.6)	47 (69.1)
Abnormal	26 (5.5)	4 (16.0)	12 (19.4)	21 (30.9)
Bilirubin				
Normal	603 (98.2)	10 (90.9)	17 (68.0)	16 (34.8)
Abnormal	11 (1.8)	1 (9.1)	8 (32.0)	30 (65.2)
ALP				
Normal	587 (97.2)	9 (69.2)	24 (80.0)	26 (46.4)
Abnormal	17 (2.8)	4 (30.8)	6 (20)	30 (53.6)
GGT				
Normal	120 (90.9)	5 (62.5)	58 (69.9)	79 (20.4)
Abnormal	12 (9.1)	3 (37.5)	25 (30.1)	308 (79.6)
Albumin				
Normal	692 (98.3)	12 (80.0)	13 (100.0)	2 (50.0)
Abnormal	12 (1.7)	3 (20.0)	0 (0.0)	2 (50.0)
Globulin				
Normal	455 (94.2)	8 (80.0)	11 (84.6)	9 (64.3)
Abnormal	28 (5.8)	2 (20.0)	2 (15.4)	5 (35.7)
Total protein				
Normal	399 (91.5)	43 (76.8)	16 (84.2)	11 (37.9)
Abnormal	37 (8.5)	13 (23.2)	3 (15.8)	18 (62.1)

a Entries are numbers of patients with percentages for two-year results.

TABLE 36 Correlations between measurements of the same LFT at different times, based on data standardised within laboratories

	Correlations over time between		
Analyte	Index and FU1	Index and FU2	FU1 and FU2
ALT	0.792	0.585	0.596
AST	0.736	0.470	0.511
Bilirubin	0.760	0.701	0.717
ALP	0.865	0.651	0.708
GGT	0.891	0.720	0.756
Albumin	0.726	0.575	0.617
Globulin	0.676	0.491	0.492
Total protein	0.659	0.544	0.529

variance that can be explained in this way may be obtained by means of an intraclass correlation coefficient (ICC) computed across the three epochs (index, FU1, FU2) at which the LFT panels were recorded (see *Table 37*, column 1).

TABLE 37 Estimates of the proportion of LFT variance attributable to long-term differences between patients

Analyte	Intrapatient correlation (%)	Estimate from temporal analysis		
		%	95% confidence limits	
ALT	68.8	52.5	47.7	57.4
AST	61.2	38.3	32.9	43.8
Bilirubin	72.5	69.6	64.9	74.3
ALP	76.0	66.7	62.2	71.2
GGT	80.9	76.5	72.8	80.4
Albumin	55.8	46.7	25.6	67.8
Globulin	55.2	51.8	42.7	60.8
Total protein	56.6	62.9	58.9	67.0

However, an analysis based on intraclass correlation necessarily disregards the selection effects that arise because abnormal LFTs were used as a criterion of entry to the study (see *Selection effects*). Moreover, it takes no account of the actual time intervals between the three test epochs for individual patients. Hence, the method does not adjust for ephemeral or 'transient' variation in levels attributable to medium-term (perhaps seasonal) fluctuations in the patient's environment or behaviour, which are of limited interest in the context of long-term clinical conditions. These issues are addressed more fully by the temporal analysis in *Appendix 2* (see *Temporal modelling of liver function tests*). Some of the results of this analysis are reproduced in the final columns of *Table 37*, which show the estimated proportion of the variance in each analyte that is attributable to long-term patient differences according to the methods described in *Appendix 2* (see *Temporal modelling of liver function test*).

It appears that the intraclass correlation tends to overestimate the variance attributable to differences between patients – particularly striking in the case of AST. Nevertheless, the impact of patient differences is substantial for all analytes.

The usefulness of LFTs as discriminatory tool for long-term clinical prognosis may be limited by the magnitude of the patient-level component of variance. In this respect it is possible that ALT has more potential than its close competitor AST. Although these are highly correlated with one another (see *Diagnostic performance of alternative liver function test panels*), the level of ALT may be more reflective of long-term differences between patients.

Impact of patient characteristics on liver function test results

Patient characteristics selected

Patient characteristics selected here are sex, age, ethnic group, BMI and alcohol consumption.

Univariate analysis

Results of one-way analysis of variable on log-LFT data with adjustment for laboratory effects are shown in detail in *Appendix 2* (see *Univariate analyses*). The results may be summarised as follows.

Sex

Most of the analytes exhibit a sex effect. ALP levels tend to be higher in females than in males. For all other non-protein LFTs, and for albumin, average levels for females are lower than those for males.

Age

All analytes except globulin show significant relationships with age. These are strongest for ALT, there being some evidence that levels are lowest in the youngest and oldest patients, and for albumin, which declines steadily with increasing age.

Ethnic group

Here significant effects are largely confined to the protein measures globulin and total protein, with black and Asian subjects exhibiting higher average levels than white subjects.

BMI

Separate BMI measurements were not available at baseline, so the index analysis of LFTs utilises FU1 BMI categories. Both ALT and GGT levels were raised among patients in the higher BMI categories. Globulin also shows some increase, although this effect is not evident at the 2-year follow-up.

Alcohol

Alcohol intake was not requested at baseline, so the index analysis of LFTs utilises FU1 alcohol categories. All the non-protein analytes are significantly related to alcohol intake, particularly at the beginning of the study (index and FU1). For ALT, AST and GGT, the association is in the positive direction, with high alcohol accompanied by raised LFTs. For ALP, the direction of effect is reversed. For bilirubin, the association is less strong and its character unclear.

Multivariate analysis

It is possible that some of the effects identified in the univariate analyses can be attributed to confounding between different patient characteristics. To address this issue, stepwise analyses were carried out using a backwards elimination method, with the aim of finding the most convincing models for the influence of patient-level covariates on LFT results.

These analyses were carried out using FU1 data only. The model-fitting strategy is described below.

1. For each analyte, perform an analysis of variance (ANOVA) on the log-transformed LFT values, including the following terms:
   i. main effect of laboratory
   ii. main effects and interaction for sex and age categories
   iii. main effects of ethnic group, BMI and alcohol intake (categorised as usual)
   iv. all two-way interactions of sex category with items in (iii)
   v. all two-way interactions of age category with items in (iii).
2. Eliminate non-significant ($p > 0.05$) interaction terms under (iv) and (v) using backwards elimination.
3. Drop non-significant main effects under (iii) using backwards elimination, unless the retention of a main effect was necessary to support the interpretation of a significant interaction term.

Where the final model contained a two-way interaction between age or sex and one of BMI, alcohol intake or ethnic group, this interaction was extended to a three-way interaction in which both age and sex were included. In no case was the three-way interaction statistically significant.

In addition to the laboratory effects, all of the 'final models' obtained by this strategy necessarily include age, sex and their interaction effect, whether or not these achieved formal statistical significance.

The outcome of the model-fitting strategy is briefly summarised in *Table 38*, which contains the *p*-values for the terms labelled (iii), (iv) or (v) from the list above that were retained in the final models. All models include age, sex and age × sex, as well as an adjustment for laboratory effects. The final column in *Table 38* refers to the proportion of the variance explained by the model (i.e. an adjusted R^2-statistic), computed net of laboratory effects. Parameter estimates representing the detailed impact of the covariates in the final fitted models are given in *Appendix 2* (see *Multivariate analyses*).

It is striking that ethnic group impacts only on protein measures, and that alcohol impacts (to some extent, at least) on all non-protein measures. The influence of BMI is pervasive, affecting everything except ALP, but its effects vary with age in most cases. The weak effect on albumin of the alcohol × sex interaction may be accidental given that the main effect of alcohol is not significant for this analyte. For some purposes [e.g. in the temporal correlation analysis of *Appendix 2* (see *Summary of analyses of liver function test results*)], these terms have been omitted together with the weak ethnicity × age interaction in the total protein model.

The final column in the table quantifies the extent to which an LFT is predictable from the general characteristics of the patient in our data set, without taking direct account of any pathological condition.

Impact of diagnostic category

Table 39 shows the impact of 'serious liver disease' on the models developed above. The results were obtained by adding a four-level diagnostic category (see *Diagnostic categories* and *Liver-related disease*) to the multivariate models (see *Multivariate analysis*). Entries in this table are estimates and confidence intervals (CIs) for the multiplicative factors that represent the impact of diseases on the LFT result with the 'non-specific' (i.e. category 3) as base.

From this analysis, it appears that both ALT and AST are raised in category 1a diseases (including viral hepatitis). ALP is raised in category 1b and also in category 2. GGT is raised in categories 1a and 1b, and globulin in 1b. These are the only findings to achieve formal statistical significance, although the small size of the disease categories will certainly compromise the power here.

Impact of fatty liver

A similar analysis is presented in *Table 40* for the impact of fatty liver (as reported on ultrasound at FU1) on the panel of LFTs. Entries in the table represent multiplicative factors that apply to the average LFT level under the condition described at the head of each column, with the

TABLE 38 Summary of final LFT models

Analyte	BMI	Alcohol	Ethnicity	BMI × age	Other terms	Variance explained (%)
ALT	$p=0.0000$[a]	$p=0.0031$		$p=0.0012$		19.7
AST	$p=0.3367$[a]	$p=0.0000$		$p=0.0055$		6.7
Bilirubin	$p=0.0021$	$p=0.0321$				9.8
ALP		$p=0.0301$				9.3
GGT	$p=0.0002$[a]	$p=0.0000$		$p=0.0068$		13.8
Albumin	$p=0.0005$[a]	$p=0.4734$[a]	$p=0.0192$	$p=0.0081$	Alcohol × sex; $p=0.0281$	11.9
Globulin	$p=0.0000$[a]		$p=0.0000$	$p=0.0083$		8.7
Total protein	$p=0.2224$[a]		$p=0.0000$[a]	$p=0.0035$	Ethnicity × age; $p=0.0384$	8.5

a Denotes the *p*-value for the main effect in a simplified model with the associated interaction terms removed.

TABLE 39 Impact of disease on LFT after adjustment for patient characteristics

	Hepatitis B or C ($n=13$)		Hepatocellular disease (1a), excluding hepatitis ($n=19$)		Hepatic bile duct disease (1b) ($n=12$)		Other diseases affecting liver (2) ($n=15$)	
Analyte	**Ratio**	**95% CI**	**Ratio**	**95% CI**	**Ratio**	**95% CI**	**Ratio**	**95% CI**
ALT	2.25	1.65 to 3.07	1.61	1.27 to 2.04	1.07	0.79 to 1.46	1.07	0.82 to 1.39
AST	1.69	1.35 to 2.13	1.56	1.30 to 1.88	1.20	0.94 to 1.51	1.00	0.82 to 1.24
Bilirubin	0.93	0.68 to 1.28	1.27	0.99 to 1.64	0.83	0.60 to 1.14	1.02	0.77 to 1.35
ALP	0.94	0.77 to 1.15	0.98	0.84 to 1.15	1.64	1.35 to 1.98	1.27	1.07 to 1.50
GGT	1.17	0.76 to 1.79	1.51	1.08 to 2.13	1.68	1.06 to 2.68	1.19	0.82 to 1.74
Albumin	1.02	0.97 to 1.06	1.00	0.97 to 1.04	0.97	0.93 to 1.01	0.96	0.93 to 1.00
Globulin	1.05	0.95 to 1.15	0.97	0.90 to 1.05	1.13	1.03 to 1.24	1.02	0.94 to 1.10
Total protein	1.02	0.98 to 1.07	0.99	0.96 to 1.03	1.03	0.99 to 1.07	0.99	0.96 to 1.02

TABLE 40 Impact of fatty liver on LFT after adjustment for patient characteristics

	Fatty liver present ($n=484$)		Mild ($n=263$)		Moderate ($n=177$)		Severe ($n=44$)	
Analyte	**Ratio**	**95% CI**	**Ratio**	**95% CI**	**Ratio**	**95% CI**	**Ratio**	**95% CI**
ALT	1.35	1.27 to 1.44	1.29	1.19 to 1.39	1.40	1.29 to 1.53	1.58	1.35 to 1.84
AST	1.20	1.14 to 1.26	1.16	1.09 to 1.23	1.22	1.14 to 1.31	1.34	1.18 to 1.53
Bilirubin	1.02	0.95 to 1.09	1.05	0.97 to 1.14	0.97	0.89 to 1.07	1.03	0.87 to 1.22
ALP	0.96	0.92 to 1.00	0.95	0.91 to 1.00	0.98	0.92 to 1.04	0.95	0.85 to 1.05
GGT	1.13	1.03 to 1.25	1.06	0.95 to 1.19	1.23	1.08 to 1.40	1.28	1.02 to 1.62
Albumin	1.02	1.01 to 1.03	1.02	1.01 to 1.03	1.02	1.01 to 1.03	1.02	1.00 to 1.04
Globulin	0.99	0.97 to 1.01	0.97	0.95 to 0.99	1.02	0.99 to 1.05	0.99	0.94 to 1.04
Total protein	1.01	1.00 to 1.02	1.00	0.99 to 1.01	1.02	1.01 to 1.03	1.00	0.98 to 1.02

'Normal' category (i.e. no fatty liver) as base. They derive from two analyses: for the column 'fatty liver present' a two-level categorical variable was added to the patient characteristic model; the remaining columns derive from an analysis in which fatty liver was represented by a four-level categorical variable.

The contribution of fatty liver was highly significant ($p<0.0005$) in both analyses for ALT, AST and albumin with an impact that worsens with the severity of the condition. For GGT, a similar qualitative effect was observed, but with less extreme p-values ($p=0.0088$ and 0.0075 for the contribution in the two analyses). Bilirubin and ALP exhibited no significant effects ($p=0.5333$ and 0.4606 for bilirubin, $p=0.0593$ and 0.2357 for ALP). The results for globulin and total protein are anomalous. Although there is no significant evidence for the impact of the presence of fatty liver ($p=0.3011$ for globulin; $p=0.1434$ for total protein), the impact of severity is formally significant in both cases ($p=0.0076$, $p=0.0187$). Inspection of the ratio estimates suggests a pattern in which the impact rises with moderate fattiness and falls back in severe cases. This is scarcely a plausible finding but it could arise as a chance result for globulin, with a knock-on effect in the aggregate protein measure total protein.

Laboratory and practice effects

In the data set, differences between laboratories are confounded with practice effects; indeed, they may be regarded as a component of the differences between practices. Thus, it is possible that some 'laboratory' effects (see *Appendix 2*, *Liver function test results by laboratory*) have little to do with differences in laboratory procedure.

Practice effects could arise from several different sources. For instance:

- as a proxy for geographical effects
- as a reflection of different testing policies in different practices
- as a reflection of different testing policies between different GPs within practices.

Practice and laboratory effects were assessed using the models developed above for the first follow-up LFTs. *Table 41* shows the results of testing for laboratory and practice effects in a nested ANOVA. Patient-level covariates identified in *Table 38* above are included for each analyte.

The evidence for differences between laboratories is overwhelming only in the case of ALP, where it was already known to be present. For six of the other analytes (i.e. except for GGT) there is some modest evidence for a laboratory effect. The protein measures (albumin, globulin and total protein) seem not susceptible to variation at the practice level. However, the evidence for such variation is moderately strong for the other LFTs (especially AST, GGT and ALP), with only bilirubin yielding a result that is not formally significant.

The impact of practice effects on the overall fitted models is comparatively small and does not appear to distort the impact of patient characteristics on LFT results. In particular, none of the terms would have been dropped (as non-significant) from any of the models in *Table 38* following inclusion of fixed practice effects.

Liver-related disease

Disease categories

Over the course of the study, 59 patients were diagnosed with a serious liver-related condition. Of these, 32 had category 1a diseases (including 13 with hepatitis B or C), 12 had category 1b diseases, and 15 had category 2 diseases. The breakdown of these conditions is shown in *Table 42*. The remaining 1231 patients were classified as having no specific liver diagnosis. However, 53 of the non-specific group were not tested for at least one of hepatitis B and C – the most common of the serious diseases. These 53 patients were excluded from analyses that involve the diagnostic

TABLE 41 Laboratory and practice effects[a]

	Between laboratories (assessed with respect to variation between practices)		Differences between practices within laboratories	
Analyte	*F* (2 and 8 df)	*p*-value	*F* (8 and 1100+ df)[b]	*p*-value
ALT	4.19	0.0570	2.50	0.0107
AST	4.99	0.0393	3.71	0.0003
Bilirubin	6.26	0.0231	1.85	0.0646
ALP	168.44	0.0000	2.87	0.0036
GGT	0.17	0.8461	4.18	0.0001
Albumin	8.35	0.0110	1.04	0.4056
Globulin	4.12	0.0588	1.04	0.4066
Total protein	4.50	0.0490	0.37	0.9388

df, degrees of freedom.

a The *F*-statistics in the second column represent the variance attributable to differences between laboratories compared with what would be expected if variation between laboratories was entirely due to differences at the practice level. The *F*-statistics in the third column describe the observed variation between practices compared with what would be expected in the absence of systematic practice effects.

b The df in the denominator varies between analytes (because of missing values), but is always >1100.

grouping, leaving 1178 patients in the non-specific group. A combination of non-specific and diagnostic categories yields 1237 patients.

Chronic viral hepatitis

Thirteen out of 1236 patients for whom results of both test were available (hepatitis B *or* hepatitis C test results were missing for 54 of 1290 patients; hepatitis B *and* hepatitis C results were missing in 38 cases) had chronic viral hepatitis: nine patients had hepatitis B and four had hepatitis C. The breakdown of the type of analyte that was abnormal in the index panel is shown in *Table 43*. In 10 out of these 13 cases, more than one analyte was abnormal. In eight cases, the ALT was abnormal, and there was no case in which AST was high but ALT was normal. In one case, only

TABLE 42 Patients with serious liver-related diseases (categories 1 and 2)

Category	No. of patients	Subtotal	Total
1. Specific liver diseases			44
1a. Hepatocellular diseases		32	
Viral hepatitis	13		
Haemochromatosis	10		
A1AT deficiency	3		
Alcoholic cirrhosis	5		
Hepatocellular cancer	1		
1b. Intrahepatobiliary duct disease		12	
PBC	10		
PSC	2		
2. Systemic diseases involving the liver			15
Liver metastases	4		
Cancer pancreas/bile duct	4		
Lyme disease	1		
Hypothyroidism[40]	4		
Amoebic liver abscess	1		
Chronic pancreatitis	1		

TABLE 43 Results of index LFT among viral hepatitis cases

Type of hepatitis	Case	ALT	AST	Bilirubin	ALP	GGT	Albumin	Globulin	Total protein	Fatty[a]
Hepatitis B	1	High	High	Normal	Normal	High	Normal	Normal	High	Yes
	2	Normal	Normal	High	Normal	Normal	High	Low	Normal	No
	3	High	Normal	Normal	Normal	High	Normal	Normal	Normal	Yes
	4	High	High	High	Normal	High	Normal	Normal	High	Yes
	5	High	High	Normal	Normal	Normal	Normal	Normal	Normal	Yes
	6	High	–	–	–	Normal	Normal	Normal	Normal	No
	7	Normal	–	High	Normal	–	Normal	–	–	No
	8	–	Normal	Normal	Normal	Normal	Normal	High	High	No
	9	–	High	Normal	Normal	–	Normal	–	–	No
Hepatitis C	1	High	Normal	Normal	Normal	High	Normal	Normal	Normal	No
	2	High	High	Normal	Normal	Normal	Normal	Normal	High	Yes
	3	Normal	–	Normal	Normal	High	Normal	–	–	Yes
	4	High	High	Normal	Normal	Normal	Normal	Normal	Normal	No

–, test missing.
a Features of fatty liver were detected on ultrasound.

protein levels were abnormal and all the enzyme tests (ALT, AST, GGT and ALP) were normal. Fatty liver was present in six cases, a proportion that is similar to the overall rate of fatty liver in the complete data set (see *Ultrasound features*). One patient with hepatitis B has subsequently progressed to cirrhosis.

When ALT or AST levels were abnormal, the values tended to be more extreme in patients with viral hepatitis than in patients who did not have this disease (*Table 44*). The same trend was observed of all patients as a whole, not only those in whom the analyte was abnormal (data not shown). In nine participants either ALT or AST was abnormal. Country of origin was recorded in 1208 out of the 1236 patients in whom both viral hepatitis tests were undertaken: 107 were born in a medium- or high-risk country for hepatitis B or hepatitis C, according to World Health Organization (WHO) definitions, but 11 out of the 13 patients with chronic hepatitis originated from a medium- or high-risk country. None of the 13 patients admitted to use of intravenous drugs at any time.

Primary biliary cirrhosis and primary sclerosing cholangitis

Primary biliary cirrhosis was defined as a cholestatic blood picture with positive serology for anti-mitochondrial antibody (AMA). AMAs were positive in 13 BALLETS cases. Three were weakly positive, leaving 10 positive cases included as category 1b cases in the statistical analysis. In retrospect, one of those cases (case 7) may have been misclassified (*Table 45*). Nine patients had a diagnosis of PBC confirmed by liver specialist follow-up, with a strong predominance for female sex (8/9) and white race (9/9). The mean age was 69.1 years, with two-thirds being aged > 65 years. Other risk factors for liver disease, namely alcohol excess and obesity, were unremarkable in this PBC cohort with a mean BMI of 27.7 kg/m^2, of whom 100% consumed ≤ 6 units of alcohol per week on average. ALP was abnormal on index blood testing in all the identified female patients with PBC. The only exception was the male patient with a solitary GGT abnormality on index testing. The ALP remained abnormal on repeat blood testing at FU1 in seven individuals. The course of the disease is usually benign in patients detected by LFTs rather than features of cirrhosis. However, two cases in our sample had features of early cirrhosis on ultrasound.

In summary, if a GP identifies an incidental raised ALP (± GGT) in a white woman aged > 65 years, having excluded obesity and/or alcohol excess as causes of liver disease, it is likely to be cost-effective and clinically more intuitive to proceed straight to an AMA test rather than proceed to repeat tests or a full liver screen. An AMA test costs £8.01 (University Hospitals Birmingham NHS Foundation Trust, 2011) and the result should be available within 7 days, thus not delaying further clinical investigation if indicated. Two cases of PSC were detected in the study (*Table 46*).

Autoimmune hepatitis

Smooth muscle antibodies (SMAs) were positive in 47 cases (weakly positive in five of these). In two cases (*Table 47*) either the ALT or AST exceeded twice the ULN. These cases had not been

TABLE 44 Comparison of abnormal ALT and AST results from the index panel in HBV and HCV cases compared with non-hepatitis cases

		HBV or HCV			Non-hepatitis		
Analyte	Upper limit	*n*	Mean	Median	*n*	Mean	Median
ALT	41	8	98.0	89.5	426	65.4	56.0
AST	43	6	94.5	69.5	254	64.5	53.5

HBV, hepatitis B virus; HCV, hepatitis C virus.

TABLE 45 Abnormal mitochondrial antibodies: diagnosis of PBC

Case no.	Age (years)	Sex	Titre	M2 Subtype	ALP	Proportional increase of ALP above ULN	Abdominal ultrasound	Confirmed diagnosis by liver specialist
1	73	M	1 : 40	Weak positive	278	Normal	Mild fatty liver	PBC
2[a]	58	M	1 : 40	Negative	170	Normal	Normal	PBC
3	75	F	1 : 100	Strong positive	456	1.38	Mildly irregular liver surface	PBC
4	86	F	1 : 100	Strong positive	462	1.4	1-cm simple cyst head of pancreas and pancreatic duct at head at ULN (at 3 mm); common bile duct dilated at 0.9 cm	PBC
5	87	F	1 : 100	Positive	362	1.1	Normal	PBC
6	69	F	1 : 100	Strong positive	346	1.05	Normal	PBC
7[b]	45	F	1 : 25	Negative	206	Normal	Mild fatty liver; mid-portion bile duct minimally dilated at 72 cm	Not referred
8	48	F	1 : 100	Strong positive	364	1.26	Moderate fatty liver; liver enlarged at 18 cm; bright, coarse texture – possibly cirrhotic change	PBC
9	52	F	1 : 40	Positive	407	1.23	Gallbladder multiple small calculi – largest 5 mm	PBC
10	81	F	1 : 100	Strong positive	633	Normal	Gallbladder multiple small calculi; small aortic aneurysm	PBC

F, female; M, male.
a Termed pre-symptomatic by the hepatologist.
b The diagnosis of PBC could be questioned in this case.

TABLE 46 Diagnosis of PSC

Case no.	Age (years)	Sex	ALP	Co-existing conditions	Abdominal ultrasound	Confirmed diagnosis by liver specialist
1	40	M	349	Ulcerative colitis	Mild intrahepatic duct dilatation; mildly heterogeneous liver texture	PSC (MRI)
2	29	M	1075	Ulcerative colitis	Common bile duct at ULN (7 mm)	PSC (liver biopsy)

M, male.

TABLE 47 Smooth muscle antibodies with index ALT or AST levels exceeding twice the ULN

Case no.	Age (years)	Smooth muscle	ALT and AST	Abdominal ultrasound	Action
1	74	Positive	ALT 29 (normal); AST 29 (normal)	Stone present in fundus of gallbladder; enlarged spleen at 15.7 cm	Referred. Reviewed by two liver specialists. Outcome = alcohol excess, unlikely autoimmune hepatitis
2	70	Positive	ALT missing; AST 249 (raised)	Abnormal parenchyma; small calcified speck in right lobe	Not referred

followed up at the time of the report and the study hepatologist is reticent about making a firm diagnosis of this very rare disease in elderly patients. Moreover, such a diagnosis would use test results as both the topic of investigation and a diagnostic criterion thereby risking inclusion bias.

Haemochromatosis

Iron saturation exceeded 50% in 39 cases, and in eight of these it exceeded 80%. We obtained a haemochromatosis genotype for 27 cases with iron saturation > 50% during the 2-year follow-up.

These cases are summarised in *Table 48*, in which it can be seen that there are six cases of homozygous haemochromatosis (*C282Y* or *H63D*) and four compound heterozygote (*C282Y* + *H63D*) who may be classed as having haemochromatosis (category 1a disease). There were also two carrier heterozygotes (*C282Y*). Five of the six homozygous patients had iron saturations above 80%. In none of the compound heterozygote patients did iron saturation exceed this level. Three patients were deceased and one patient was no longer registered at this practice. Five patients did not attend follow-up clinic appointments. Three patients attended hospital liver clinic appointments but did not have *HFE* genotype results. Four out of the six homozygous cases had ferritin levels of > 1000 mg/dl and receive frequent venesection. Their families have been screened.

Wilson's disease

Four patients had abnormal caeruloplasmin levels (*Table 49*). Wilson's disease was excluded by 24-hour urine test in three patients. We reminded the GP of the remaining patient of the possible desirability of referral, but this does not seemed to have occurred.

TABLE 48 Features of patients with raised iron saturation (> 50%)

Case no.	Sex	Age (years)	Fe saturation (%)	*HFE* genotype	Ultrasound abnormalities
1	M	43	58.0	HHCC normal	None
2	F	60	51.9	HHCC normal	Mildly fatty liver
3	F	75	53.3	HHCC normal	Mildly fatty liver
4	M	56	51.7	Compound heterozygote	Abnormal parenchyma and moderately fatty liver
5	M	60	55.7	HHCC normal	Abnormal parenchyma and single, solid focal lesion 1.4 × 1 cm; possibly haemangioma
6	F	72	61.9	Haemochromatosis homozygote (*H63D*)	Three polyps in gallbladder – largest = 4.3 mm
7	F	45	85.9	Alcohol dependent; did not return to clinic	Enlarged liver and abnormal parenchyma and moderately fatty liver
8	M	51	87.8	Deceased; septicaemia/septic arthritis	Enlarged liver and abnormal parenchyma and marked fatty liver; Spleen enlarged at 14 cm
9	M	75	94.6	Haemochromatosis homozygote (*C282Y*)	Abnormal parenchyma and small highly echogenic focus in posterior aspect of right lobe – 1.1 cm
10	M	64	64.0	HHCC normal	None
11	F	53	56.7	HHCC normal	Abnormal parenchyma and mildly fatty liver; multiple angiomyolipomas in right kidney
12	M	54	79.2	Compound heterozygote	Abnormal parenchyma and moderately fatty liver
13	F	67	85.1	Haemochromatosis homozygote (*C282Y*)	Abnormal parenchyma and moderately fatty liver; two small calculi in gallbladder
14	M	41	61.4	Did not return to clinic	Left renal calculus – 8 mm

continued

TABLE 48 Features of patients with raised iron saturation (>50%) (*continued*)

Case no.	Sex	Age (years)	Fe saturation (%)	*HFE* genotype	Ultrasound abnormalities
15	M	61	64.8	Deceased; metastatic cholangiocarcinoma	Dilated tubular structure extending into right lobe, numerous areas of calcification
16	M	74	62.2	HHCC normal	Abnormal parenchyma and a solid focal lesion on liver adjacent to gallbladder; moderately fatty liver
17[a]	M	73	70.5	No result; alcoholic cirrhosis	Abnormal parenchyma and liver has irregular surface and abnormal coarse texture; cirrhotic changes; enlarged spleen at 15 cm
18	M	32	51.2	No longer registered	None
19	M	58	59.5	Did not return to clinic	None
20	M	54	55.2	HHCC normal	Ascites around liver and in pelvis
21	M	50	50.8	Did not return to clinic	None
22	M	45	81.5	Haemochromatosis; homozygote (*C282Y*)	Moderately fatty liver; 1.9 cm × 1.3 cm solid hypoechoic area adjacent to gallbladder
23	F	64	60.5	HHCC normal	None
24	M	78	65.4	HHCC normal	None
25	M	54	64.8	Compound heterozygote	None
26	M	51	91.1	Carrier, heterozygote	Mildly fatty liver and abnormal parenchyma
27	F	50	94.1	Haemochromatosis; homozygote (*C282Y*)	None
28	F	75	50.1	HHCC normal	Abnormal parenchyma and moderately fatty liver; enlarged liver
29	F	72	71.5	HHCC normal	None
30	M	62	58.8	Did not return to clinic	Abnormal parenchyma and moderately fatty liver; dilation of intrahepatic common bile duct at 8 mm
31	F	54	72.4	Hepatology referral; no result	None
32[a]	F	67	57.0	Deceased; liver failure; alcohol dependence	Liver appears coarse – increased reflectivity in gallbladder – no shadowing
33	M	39	53.7	HHCC normal	None
34	F	75	94.3	Haemochromatosis; homozygote (*C282Y*)	None
35	M	54	50.9	HHCC normal	None
36	M	49	53.2	Compound heterozygote	Abnormal parenchyma and moderately fatty liver; area of calcification in head of pancreas
37	M	36	55.5	Hepatology referral; no result	None
38	F	33	79.7	HHCC normal	Abnormal parenchyma and mildly fatty liver
39	M	62	54.2	Carrier, heterozygote	None

F, female; *HFE*, haemochromatosis gene; HHCC, a code for wild-type for HFE genotyping (i.e.negative for C282Y and H63D); M, male.
a These two cases are likely to have suffered from alcoholic cirrhosis, possibly aggravated by haemochromatosis.

TABLE 49 Caeruloplasmin: abnormal results

Case no.	Age (years)	Caeruloplasmin (mg/dL)	Abdominal ultrasound	GP action
1	66	0.11	Two renal calculi	Referral to liver clinic; Wilson's disease excluded by 24-hour urine test
2	53	0.14	Mildly fatty liver, smaller left lobe	Wilson's disease excluded by 24-hour urine test
3	36	0.09	None	Wilson's disease excluded by 24-hour urine test
4	57	0.14	Mildly fatty liver, abnormal parenchyma	Not referred

Alpha-1 antitrypsin deficiency

Low A1AT levels were found in 47 patients. Thirty-seven have had phenotype testing and these were abnormal in three cases (*Table 50*). Cases 1 and 2 are under the care of a specialist. Case 3 has a lower risk phenotype (Pi MZ).

Alcoholic cirrhosis

There were five cases in which the hepatologist agreed (on the basis of the ultrasound picture and history) that the patient had alcoholic cirrhosis (with some overlap with haemochromatosis). A further case of hepatocellular cancer was detected in a patient who did not have hepatitis B, but biopsy confirmed diagnosis of non-alcoholic steatohepatitis (NASH).

Other diseases (category 2) involving the liver

Metastatic cancers in the liver, cancer of the pancreas/bile duct and hypothyroidism are the common diseases. We did not include incidental cancers (e.g. cancer of kidney) or gallstones confined to the gallbladder in this category.

Diagnostic value of liver function tests

Objective

In this section the possibility of using LFTs to predict diagnostic group is explored by means of a discriminant analysis incorporating appropriate adjustment for patient-level demographic and clinical variables. These comprise age, sex, BMI, ethnicity and country of birth. Alcoholic consumption is excluded; although it may be implicated in the onset of certain conditions, it is clearly a lifestyle variable and, moreover, is the one variable in the data set which has not been objectively determined. The aim here is to summarise clinical information in such a way as to inform diagnosis. The LFT panels used here were obtained at FU1.

The analysis concerns (and is restricted to) the patients diagnosed with category 1a and 1b liver disease, together with those in the non-specific diagnostic category. These groups comprise 1222 patients in all. For 14% of these patients some of the clinical and demographic data were missing. In order to make full use of the information available the complete case analysis was supplemented with an analysis using an imputation method to cope with missing observations.

Method (complete case analysis)

Adjustment for laboratory effects

For the current analysis, each LFT was corrected for a multiplicative laboratory effect obtained from fitting the explanatory model derived in *Chapter 4* (see *Multivariate analysis*) to the patients in the 'non-specific' diagnostic group. Inclusion of all patients – and incorporating diagnostic group in the model – might have weakened the ability to identify diagnostic groups whose prevalence varied across the groups of practices associated with different laboratories. In practice, there is little difference between these two approaches, as is evident from *Table 51*.

TABLE 50 Alpha-1 antitrypsin: abnormal results

Case no.	Age (years)	A1AT	Phenotype	Follow-up
1	61	A1AT	Pi SZ	Respiratory
2	52	A1AT	Pi SS	Hepatology
3	60	A1AT	Pi MZ	

TABLE 51 Laboratory adjustment factors for the complete-case analysis

Analyte	Adjustment factors used (estimated from 'non-specific' diagnostic group)			Factors estimated from all diagnostic groups		
	Birmingham (1): laboratory 1	Birmingham (2): laboratory 2	London: laboratory 3	Birmingham (1): laboratory ab 1	Birmingham (2): laboratory 2	London: laboratory 3
ALT	1	1.37	1.05	1	1.32	0.99
AST	1	1.32	1.20	1	1.31	1.17
Bilirubin	1	0.87	1.24	1	0.87	1.22
ALP	1	0.47	0.40	1	0.47	0.39
GGT	1	1.12	0.98	1	1.10	0.98
Albumin	1	1.00	1.03	1	1.00	1.03
Globulin	1	0.97	0.96	1	0.97	0.96
Total protein	1	0.98	1.00	1	0.99	1.00

Birmingham (1) is the base laboratory, with adjustment factor = 1 for all analytes. Results from other laboratories were divided by the adjustment factors in the analysis.
Note: The factors are estimated from the covariate models in *Analysis of liver function test panels*, using only patients from the non-specific diagnostic group. The second set of factors are those derived from the full set of patients in *Laboratory and practice effects*, and are included here solely for purposes of comparison.

Analysis strategy

There are three disease categories of primary interest:

1. liver disease (1a), including viral hepatitis ($n = 32$)
2. hepatitis B and C ($n = 13$)
3. liver disease (1b) ($n = 12$).

In addition, a separate analysis was conducted for liver disease (1a) – excluding viral hepatitis ($n = 19$).

In each case we attempt to discriminate between the 'non-specific' group ($n = 1178$) and the diagnostic group of interest using a linear predictor within a logistic discrimination set-up. In each analysis, subjects not falling into one of these two groups were ignored.

Forwards and backwards stepwise procedures were applied to the following list of variables:

- log-LFTs (eight analytes)
- sex
- age group (six categories)
- BMI (four categories)
- ethnic group (four categories)
- country of birth (three categories).

Interaction terms were not considered, mainly because the diagnostic groups were too sparsely populated in the data to avoid problems associated with complete outcome specification within some subgroups.

Initially, the predictors were identified using backwards elimination, with a $p = 0.01$ threshold for exclusion from the model. Comparison was made with a forwards selection procedure (with a

$p = 0.01$ threshold for inclusion) and with more general stepwise selection procedures involving both forward and backward steps. In no case did the more general procedures identify a set of predictors that had not been obtained already using either the backwards or forwards methods, which are the only ones reported here.

Demographic variables and LFTs with independent discriminatory power (identified from the stepwise procedures) were included in a multiple logistic discriminant analysis of the non-specific group with liver disease categories 1a and 1b.

Results (complete case analysis)

Stepwise procedures

A summary of results for the four discrimination problems is presented in *Tables 52–55*. Pseudo R^2 is a likelihood-based measure of fit with properties similar to a conventional R^2-statistic, and the area under the curve (AUC) denotes the area under the ROC curve, a *c*-statistic that measures diagnostic potential. For both of these quantities, values close to '1' indicate excellent predictive ability.

TABLE 52 Stepwise discrimination results for disease category 1a

Stepwise procedure	Variables retained	Pseudo R^2	AUC
Backwards elimination	ALT, BMI, ethnic group	0.219	0.861
Forwards selection	AST, country of birth	0.160	0.771

AUC, area under the curve.

TABLE 53 Stepwise discrimination results for hepatitis B and C (amalgamated)

Stepwise procedure	Variables retained	Pseudo R^2	AUC
Backwards elimination	ALT, BMI, ethnic group	0.443	0.960
Forwards selection	AST, country of birth	0.304	0.860

AUC, area under the curve.

TABLE 54 Stepwise discrimination results for disease category 1b

Procedure	Variables retained	Pseudo R^2	AUC
Backwards elimination	ALP	0.164	0.842
Forwards selection	ALP	0.164	0.842

AUC, area under the curve.

TABLE 55 Stepwise discrimination results for results for liver disease category 1a: excluding hepatitis

Procedure	Variables retained	Pseudo R^2	AUC
Backwards elimination	ALT	0.099	0.776
Forwards selection	AST	0.105	0.749

AUC, area under the curve.

Discrimination between diagnostic groups

From the results above, it appears that ALT, AST and ALP are the only LFT analytes that contribute independent discriminatory power to the problem of diagnosis. Accordingly, these were included in the discriminant analysis, alongside age, sex, BMI, ethnic group and country of birth. Four groups were retained in the analysis: non-specific; 1a, excluding hepatitis; hepatitis; and 1b. The potential of the method to distinguish between liver disease (categories 1a and 1b) and a non-specific diagnosis is summarised in *Figure 11*. This shows a ROC plot of the true-positive rate against the false-positive rate when the estimated probability of disease (i.e. probability of lying in either category 1a or category 1b) from a logistic discrimination analysis is used as a marker of disease.

According to this plot, when the threshold for a positive diagnosis is set so that the true-positive rate (sensitivity) is 90%, then more than half of the non-specific group will be misclassified as having liver disease (false-positive rate = 53%). This gives a pragmatic indication of discriminatory capability. In practice, the performance is likely to be worse than that depicted because the same data have been used both for estimation and assessment of the discriminant function.

Analysis of imputed data

Method

The stepwise method described above uses only complete cases, with the result that the final equation is computed only from patients with a complete LFT panel, and also ignores individuals with unrecorded BMI, ethnicity and country of birth. In all, 176 (14.4%) of the patients eligible for analysis had incomplete data, including 9 of the 44 patients with diagnosed liver disease in category 1. The result is that a substantial fraction of the information in the diseased groups has not contributed to the analysis presented above. To make better use of the available information, missing values were imputed from the rest of the data, augmented by age group (six categories), sex, alcohol consumption (six categories), liver fat (five categories from the sonographer's report, including non-visualised liver as a separate category) and diagnosis (five categories: non-specific; category 1a, excluding hepatitis; hepatitis; category 1b; and category 2). Imputation was carried

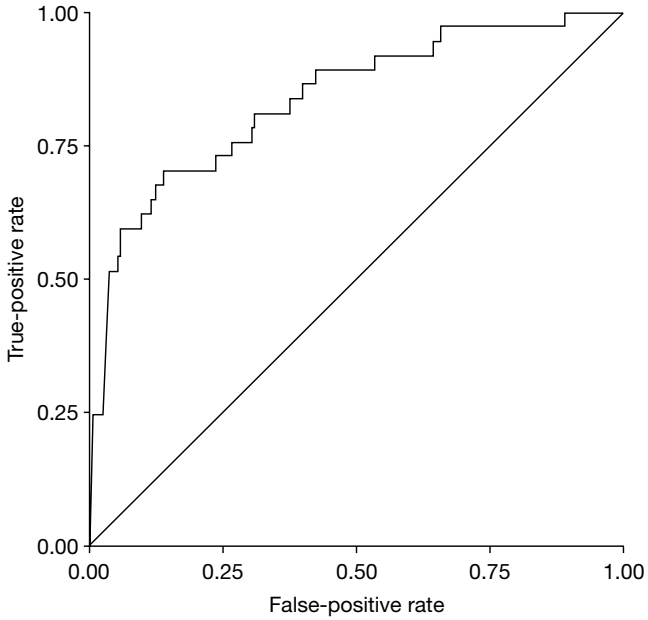


FIGURE 11 Error rates when the presence of liver disease (hepatitis, categories 1a, 1b) is assessed from a logistic discriminant analysis. The AUC = 0.839 [standard error (SE) 0.037].

out using the chained equation method of van Buuren *et al.*,[41] as implemented in Stata version 11 (StataCorp LP, College Station, TX, USA).[42] Twenty imputed samples were simulated. Within the stepwise procedures, the regression analyses were repeated for each imputed sample and the results combined using Rubin's rules.

The laboratory effects were re-estimated using the imputed samples, and are displayed in *Table 56*.

Results (imputed data)

Similar stepwise procedures were followed as for the complete case analyses. The results are displayed in *Tables 57–60*. These may be compared with *Tables 52–55*. For disease category

TABLE 56 Laboratory adjustment factors from imputed samples

	Adjustment factors from imputed samples		
Analyte	**Birmingham (1): laboratory 1**	**Birmingham (2): laboratory 2**	**London: laboratory 3**
ALT	1	1.30	1.06
AST	1	1.31	1.14
Bilirubin	1	0.89	1.21
ALP	1	0.48	0.42
GGT	1	1.12	0.98
Albumin	1	1.00	1.03
Globulin	1	0.97	0.96
Total protein	1	0.98	1.00

TABLE 57 Stepwise discrimination results for disease category 1a

Stepwise procedure	Variables retained	Pseudo R^2	AUC
Backwards elimination	ALT, BMI	0.144	0.800
Forwards selection	AST, country of birth	0.135	0.755

TABLE 58 Stepwise discrimination results for hepatitis B and C (amalgamated)

Stepwise procedure	Variables retained	Pseudo R^2	AUC
Backwards elimination	ALT, country of birth	0.308	0.924
Forwards selection	AST, country of birth	0.302	0.894

TABLE 59 Stepwise discrimination results for disease category 1b

Procedure	Variables retained	Pseudo R^2	AUC
Backwards elimination	ALP	0.189	0.835
Forwards selection	ALP	0.189	0.835

TABLE 60 Stepwise discrimination results for results for liver disease category 1a: excluding hepatitis

Procedure	Variables retained	Pseudo R^2	AUC
Backwards elimination	ALT	0.075	0.758
Forwards selection	AST	0.096	0.756

1b there are eight complete cases, augmented to 12 in the imputation analysis. Nevertheless, the same variable (ALP) emerges as the only significant predictor in both sets of analyses. The situation in category 1a is less clear cut. Here there are 27 complete cases and 32 cases in the imputation analyses. ALT and AST emerge as the only LFTs that figure in any of the selected models, although never together in the same model. For the non-hepatitis category, category 1a (19 cases, 16 complete), BMI features in the imputation model, enhancing the discrimination achieved by ALT in the complete case analysis. In the imputation analysis, country of birth (in combination with ALT or AST) has superseded ethnicity and BMI (also with ALT or AST) in predicting hepatitis (13 cases, 11 complete).

The discriminant analysis described above for complete cases was repeated for the imputed samples, using ALT, AST and ALP alongside age, sex, BMI and country of birth. In contrast to the complete case analysis, ethnic group was omitted from further analysis as it does not feature in the imputed stepwise results. The implied diagnostic performance does not change substantially, although the trade-off between false-positive and true-positive rates (as measured by the AUC statistic) is marginally less favourable (*Figure 12*).

Discussion

The general conclusion appears to be that automatic diagnosis of serious conditions has little chance of success given the high false-positive rates. The diagnosis of viral hepatitis is the most promising possibility, with an estimated AUC of > 0.90. The power of the method relies heavily on the country of origin of the patient.

Among the LFT panel only three analytes make a significant contribution: ALP is the only useful predictor of disease category 1b; ALT and AST are both implicated in the diagnosis of category 1a diseases, though they appear to substitute for one another. It remains unclear which should be preferred if a choice has to be made.

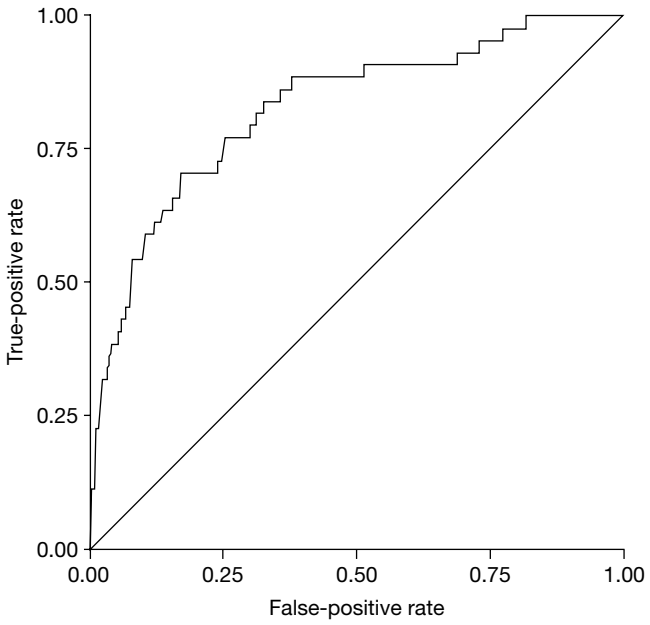


FIGURE 12 Error rates when the presence of liver disease (categories 1a, 1b, hepatitis) is assessed from a logistic discriminant analysis using multiple imputation to handle incomplete cases. The AUC = 0.828. [Estimated AUC from combination of imputed samples using Rubin's method = 0.822, with standard error (SE) = 0.036.]

Fatty liver on ultrasound

Of the 1277 participants in whom the texture of the liver could be discerned, 484 (38%) (see *Figure 6*) had an ultrasound diagnosis of fatty liver at FU1, and this was classified as moderate or marked in 221 cases (46%).

Fatty liver in patients with abnormal liver function tests

The presence of an abnormal ALT or AST was associated with an increased likelihood of fatty liver, but the prevalence of fatty liver among patients with abnormal bilirubin or ALP was reduced. Results for GGT were of marginal significance and the protein analytes exhibited no clear effects (*Table 61*).

A detailed breakdown of the severity of fatty liver when different analytes (and pairs of analytes) are abnormal is given in *Table 62*. The PPVs refer to the performance of the LFTs in determining that the liver is in the mild, moderate or severe category on ultrasound. The aminotransferase enzymes (ALT and AST) have the highest PPVs.

The proportions of fatty livers according to BMI and alcohol consumption are shown in *Figure 13* and *Table 63*. Among people with abnormal LFTs, the probability of fatty liver is over 64.6% if they are obese (BMI $\geq 30\,kg/m^2$) drinkers, rising to 73.8% if the abnormality includes an abnormal ALT. However, there is a sizeable probability (31%) of fatty liver in participants who were neither (moderate to heavy) drinkers nor obese.

Patient characteristics associated with ultrasound diagnosis of fatty liver

The present study provides an opportunity for a detailed investigation of the impact of patient characteristics on fatty liver – especially the lifestyle factors of weight and alcohol consumption. LFT results may also be considered to see how well they discriminate between fatty and non-fatty livers without reference to ultrasound. The analysis was done using the FU1 results.

A logistic regression model was constructed for the presence of fatty liver on ultrasound using a backwards elimination approach. The explanatory variables considered were:

- sex
- age group (six categories)
- ethnic group (four categories)
- BMI (four categories)
- alcohol consumption (six categories).

TABLE 61 Breakdown of abnormal and normal index LFT in people with a diagnosis of fatty liver (*p*-values from exact test)

Analyte	Tests, *n*	Abnormal, *n*	Fatty liver, *n* (%)	Normal, *n*	Fatty liver, *n* (%)	*p*-value
ALT	1102	434	230 (53.0)	668	178 (26.6)	<0.001
AST	1147	255	137 (53.7)	892	307 (34.4)	<0.001
Bilirubin	1252	148	36 (24.3)	1104	437 (39.6)	<0.001
ALP	1259	186	41 (22.0)	1073	432 (40.3)	<0.001
GGT	1139	858	342 (39.9)	281	90 (32.0)	0.019
Albumin	1265	29	7 (24.1)	1236	470 (38.0)	0.174
Globulin	966	55	16 (29.1)	911	348 (38.2)	0.199
Total protein	970	96	34 (35.4)	874	332 (38.0)	0.659

TABLE 62 Positive predictive value of LFT abnormality (on index test) for fatty liver on ultrasound[a]

	Fatty liver status when LFT abnormal				Per cent fatty liver	
Analyte	Normal, *n* (%)	Mild, *n* (%)	Moderate, *n* (%)	Severe, *n* (%)	PPV (%)	CI
Complete panel	524 (62.5)	170 (20.3)	119 (14.2)	26 (3.1)	37.5	34.3 to 40.9
Single analytes						
ALT	181 (46.5)	102 (26.2)	82 (21.1)	24 (6.2)	53.5	48.4 to 58.5
AST	106 (46.3)	63 (27.5)	42 (18.3)	18 (7.9)	53.7	47.0 to 60.3
Bilirubin	96 (74.4)	25 (19.4)	7 (5.4)	1 (0.8)	25.6	18.3 to 34.0
ALP	126 (77.3)	19 (11.7)	15 (9.2)	3 (1.8)	22.7	16.5 to 29.9
GGT	473 (60.3)	151 (19.3)	129 (16.5)	31 (4.0)	39.7	36.2 to 43.2
Albumin	19 (73.1)	6 (23.1)	1 (3.8)	0 (0.0)	26.9	11.6 to 47.8
Globulin	37 (69.8)	9 (17.0)	7 (13.2)	0 (0.0)	30.2	18.3 to 44.3
Total protein	58 (65.2)	20 (22.5)	10 (11.2)	1 (1.1)	34.8	25.0 to 45.7
Pairs of analytes						
ALT *or* AST	174 (48.3)	94 (26.1)	72 (20.0)	20 (5.6)	51.7	46.4 to 56.9
ALT *or* bilirubin	248 (53.1)	111 (23.8)	84 (18.0)	24 (5.1)	46.9	42.3 to 51.5
ALT *or* ALP	253 (53.7)	105 (22.3)	88 (18.7)	25 (5.3)	46.3	41.7 to 50.9
ALT *or* GGT	499 (58.6)	179 (21.0)	139 (16.3)	34 (4.0)	41.4	38.0 to 44.8
ALT *or* albumin	194 (48.1)	103 (25.6)	82 (20.3)	24 (6.0)	51.9	46.9 to 56.8
ALT *or* globulin	181 (49.9)	95 (26.2)	70 (19.3)	17 (4.7)	50.1	44.9 to 55.4
ALT *or* total protein	196 (51.0)	100 (26.0)	71 (18.5)	17 (4.4)	49.0	43.9 to 54.1
AST *or* bilirubin	169 (54.9)	77 (25.0)	43 (14.0)	19 (6.2)	45.1	39.5 to 50.9
AST *or* ALP	197 (57.4)	75 (21.9)	51 (14.9)	20 (5.8)	42.6	37.3 to 48.0
AST *or* GGT	462 (59.9)	157 (20.4)	122 (15.8)	30 (3.9)	40.1	36.6 to 43.6
AST *or* albumin	119 (48.6)	67 (27.3)	41 (16.7)	18 (7.3)	51.4	45.0 to 57.8
AST *or* globulin	113 (53.3)	51 (24.1)	34 (16.0)	14 (6.6)	46.7	39.8 to 53.7
AST *or* total protein	135 (55.3)	59 (24.2)	36 (14.8)	14 (5.7)	44.7	38.3 to 51.1
Bilirubin *or* ALP	216 (76.3)	43 (15.2)	20 (7.1)	4 (1.4)	23.7	18.8 to 29.1
Bilirubin *or* GGT	520 (62.1)	160 (19.1)	126 (15.0)	32 (3.8)	37.9	34.6 to 41.3
Bilirubin *or* albumin	112 (74.2)	30 (19.9)	8 (5.3)	1 (0.7)	25.8	19.1 to 33.6
Bilirubin *or* globulin	90 (72.0)	24 (19.2)	11 (8.8)	0 (0.0)	28.0	20.3 to 36.7
Bilirubin *or* total protein	111 (69.8)	33 (20.8)	14 (8.8)	1 (0.6)	30.2	23.2 to 38.0
ALP *or* GGT	505 (62.0)	150 (18.4)	127 (15.6)	33 (4.0)	38.0	34.7 to 41.5
ALP *or* albumin	141 (76.6)	24 (13.0)	16 (8.7)	3 (1.6)	23.4	17.5 to 30.2
ALP *or* globulin	115 (77.2)	17 (11.4)	15 (10.1)	2 (1.3)	22.8	16.3 to 30.4
ALP *or* total protein	134 (73.2)	28 (15.3)	18 (9.8)	3 (1.6)	26.8	20.5 to 33.8
GGT *or* albumin	476 (60.7)	151 (19.3)	126 (16.1)	31 (4.0)	39.3	35.8 to 42.8
GGT *or* globulin	418 (62.0)	131 (19.4)	102 (15.1)	23 (3.4)	38.0	34.3 to 41.8
GGT *or* total protein	431 (62.3)	136 (19.7)	102 (14.7)	23 (3.3)	37.7	34.1 to 41.4
Albumin *or* globulin	47 (73.4)	10 (15.6)	7 (10.9)	0 (0.0)	26.6	16.3 to 39.1
Albumin *or* total protein	65 (67.7)	20 (20.8)	10 (10.4)	1 (1.0)	32.3	23.1 to 42.6
Globulin *or* total protein	72 (67.3)	22 (20.6)	12 (11.2)	1 (0.9)	32.7	24.0 to 42.5

a The data are restricted to the non-specific diagnostic group, i.e. category 3.

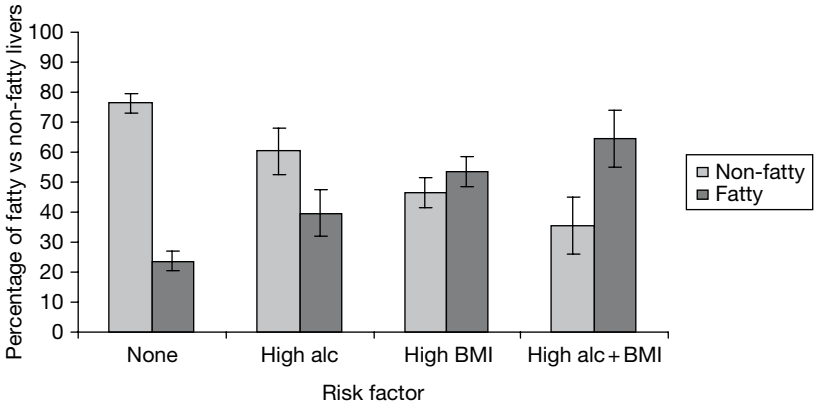


FIGURE 13 Relationship between obesity, alcohol use and fatty liver. High alc, high alcohol consumption.

TABLE 63 Relationship between obesity, alcohol use and fatty liver

	Liver		
Risk	**Normal, *n* (%)**	**Fatty, *n* (%)**	**Total, *n* (%)**
No risk	450 (76.3)	140 (23.7)	590 (100.0)
High alcohol	93 (60.4)	61 (39.6)	154 (100.0)
High BMI	183 (46.6)	210 (53.4)	393 (100.0)
High alcohol + BMI	34 (35.4)	62 (64.6)	96 (100.0)
Not known	33 (75.0)	11 (25.0)	44 (100.0)

High alcohol consumption was classed as ≥ 21 units per week for females and > 28 units per week for males.
High BMI was classed as a BMI score ≥ 30 kg/m^2.

These variables were entered into the model, together with all two-way interactions involving age or sex. Then all non-significant interaction terms ($p > 0.05$) were sequentially removed. At this stage, the intention had been to remove also non-significant main effects for variables not included in any surviving interaction (in practice, this eventuality did not arise). The method was repeated for an analysis of the ordinal category ('severity') for fatty liver using ordinal logistic regression.

The model was supplemented by LFTs (log transformed) obtained from a stepwise procedure involving both forwards and (potentially) backwards steps.

The backwards elimination and stepwise analyses were performed both for presence/absence of fatty liver (using ordinary logistic regression) and for severity (using ordinal logistic regression). The results, in terms of variables selected, were identical. Although several analytes are individually correlated with the presence of fatty liver, these correlations are subsumed into two analytes only – ALT and albumin. Once these two are entered into the model, no other analyte achieves predictive significance for the presence or severity of fatty liver.

Table 64 summarises the patient characteristic model from the logistic regression analysis. This model achieved a pseudo R^2 of 15.3% and an AUC measure of discrimination of 0.75. The ordinal regression results (not shown) are similar. It can be seen that BMI is by far the most important predictor here.

The model described above includes main effects for sex, age group (six categories), ethnic group (four categories), BMI (four categories) and alcohol consumption (six categories) together with the sex × alcohol interaction. The model may be simplified using linear and quadratic components for these terms as appropriate, leading to a more readily interpretable analysis.

Table 65 shows the effect on the deviance of sequentially replacing the categorical variables alcohol, age group and BMI by their linear and quadratic components. It appears from the *p*-values in the table that the fit of the model is not compromised by this process since the change in deviance is less than the change in degrees of freedom (df) in every case. Finally, it turns out that the interaction of sex with the quadratic component of alcohol (1 df) is not formally significant ($p = 0.0697$). It is convenient to omit this term from the final model, especially as its interpretation is not straightforward in any case.

The final (simplified) model (numbered '5' in the table) predicts that the probability of having a fatty liver:

- has an inverted U-shaped relationship with age, reaching a maximum at around age 55 years
- increases with BMI
- increases with alcohol intake above 30 units per week, though with some variation between the sexes
- is less for patients of Asian origin (compared with white patients).

These features are shown graphically in *Figures 14* and *15*, which show the relationship between fatty liver, age and alcohol intake separately for males and females of normal BMI and BMI > 30 kg/m^2. It is apparent that fatty liver is more responsive to alcohol intake in females than in males. In males there is perhaps even a suggestion that alcohol may have a protective effect at low doses, a finding corroborated in the literature (see *Discussion*).

TABLE 64 Terms retained in the patient characteristic logistic model from backwards elimination procedure, incorporating results of Wald tests for individual factors

Terms in model	Degrees of freedom	Chi-squared	*p*-value
Age group	5	37.05	0.0000
BMI	3	102.79	0.0000
Ethnic group	3	10.72	0.0134
Alcohol × sex interaction	5	13.33	0.0205
Alcohol[a]	5	16.07	0.0067
Sex[a]	1	2.97	0.0850

a From reduced model with interaction term removed.

TABLE 65 Analysis of deviance for the fatty liver logistic model

			Likelihood ratio (chi-squared) test		
Source	Deviance (negative)	df	Change in deviance	df	*p*-value
1. Full model	244.95	22	–	–	–
2. Linear and quadratic alcohol	239.34	16	5.61	6	0.4683
3. Linear and quadratic age group and alcohol	237.57	13	1.77	3	0.6215
4. Linear BMI category	237.13	11	0.44	2	0.8025
5. After removal of sex × quadratic alcohol	233.84	10	3.29	1	0.0697

–, not applicable.

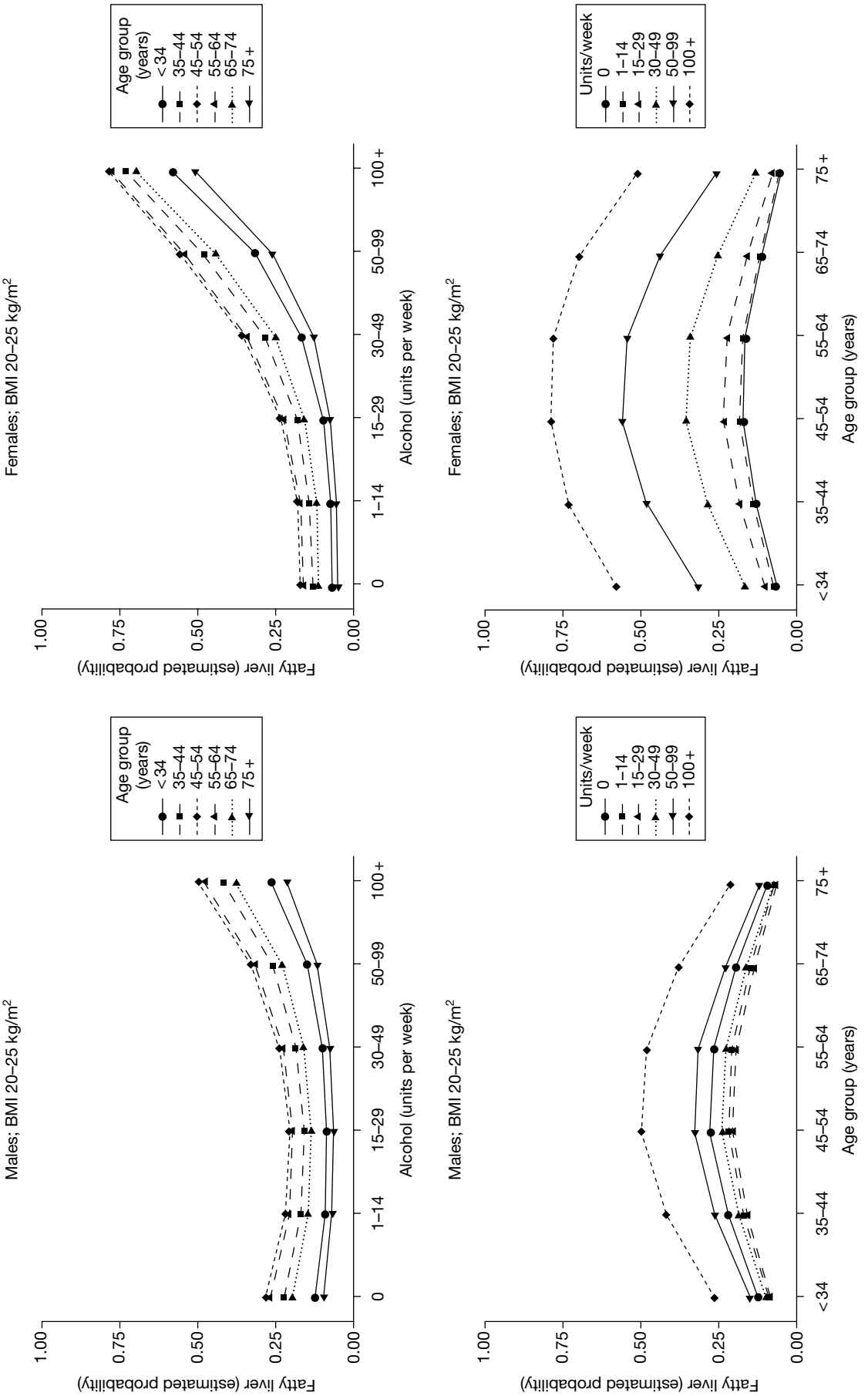


FIGURE 14 Fatty liver, alcohol and age of patients with normal BMI (20–25 kg/m^2).

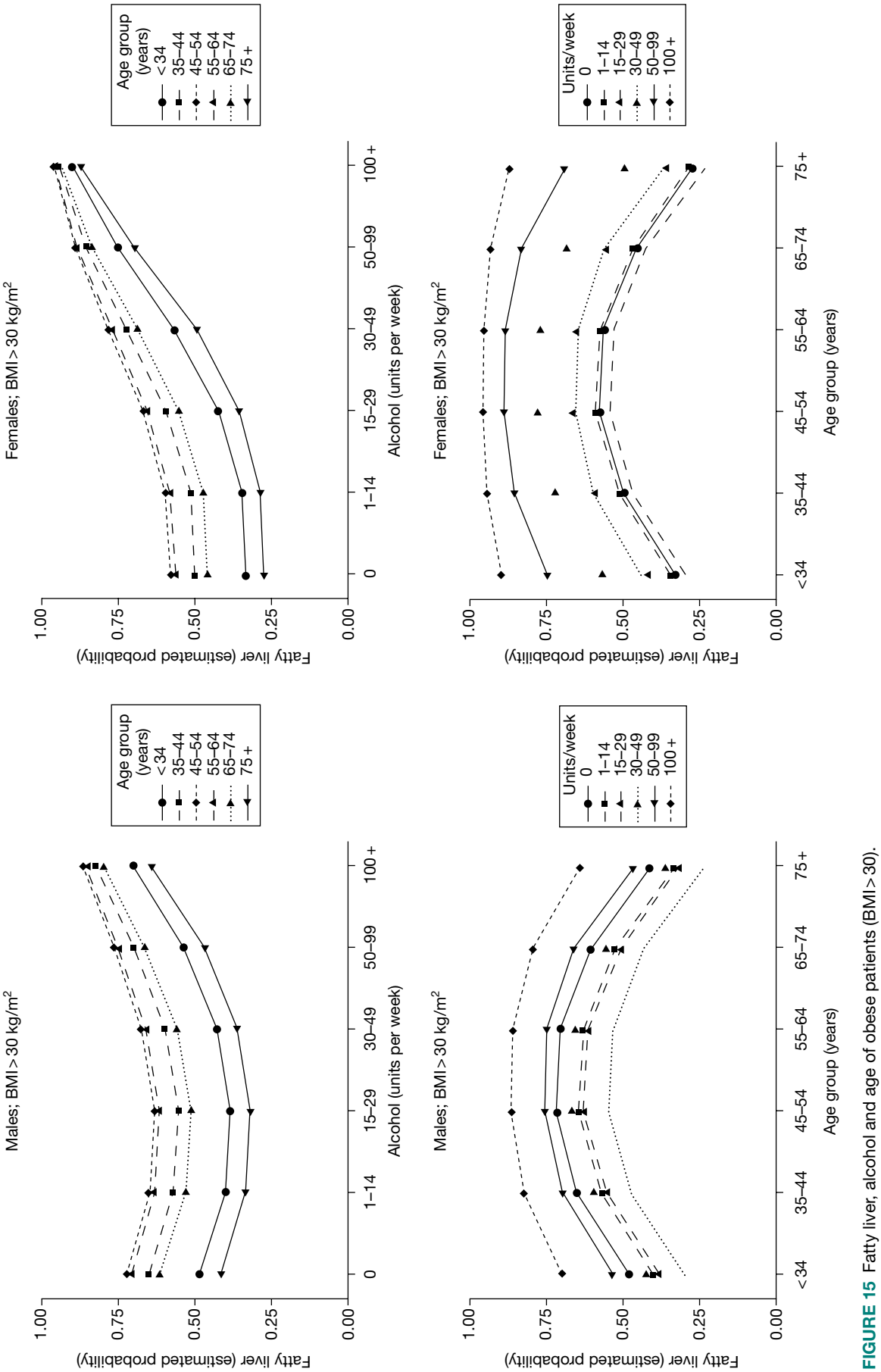


FIGURE 15 Fatty liver, alcohol and age of obese patients (BMI > 30).

Liver function tests were added to the model of *Table 65* using a stepwise procedure. Only two LFTs (ALT and albumin) were retained and the results are summarised in *Table 66*. LFT is an important predictor here, second only to BMI. The pseudo $R^2 = 22.1\%$ and AUC = 0.802. These figures do not suggest that LFTs could furnish a reliable substitute for ultrasound for the determination of fatty liver.

Persistence of fatty liver from first to second follow-up

Thirteen patients were excluded from this analysis, as belonging to the serious disease categories (categories 1 and 2). This left 628 cases for analysis. LFTs, BMI and alcohol intake were all taken from the FU2 data. Logistic regression analyses were performed with FU2 fatty liver as the outcome variable. Fatty liver at FU1 was included as a predictor in the analyses.

Sparseness of data for some of the covariate combinations impeded the fitting of the above models to the second follow-up sonography data. (For example, there are no instances of fatty liver at FU2 among the eight subjects with BMI of $< 20\,kg/m^2$.) To simplify the model fitting, age group (six categories) was replaced by linear and quadratic effects (2 df) and the category for BMI $< 20\,20\,kg/m^2$ was amalgamated with the base category ($25\,kg/m^2 \leq BMI < 30\,kg/m^2$). For ease of interpretation, units of alcohol were represented by the linear component of the six-level categorical variable.

The results (without LFTs) are summarised in *Table 67* (pseudo $R^2 = 29.8\%$ and AUC = 0.845). As might be expected, the presence of fatty liver at FU1 is the most important predictor of fatty liver at FU2.

The results of adding ALT and albumin to the model are summarised in *Table 68* (pseudo $R^2 = 28.7\%$ and AUC = 0.841). Fatty liver at FU1, BMI and ALT are the only variables that contribute independently to the chance of fatty liver at FU2. AST could stand in for ALT, but results in a marginally inferior fit. Albumin no longer features as a significant predictor either instead of ALT or in addition to it.

The effect of changes in body mass index and alcohol intake on fatty liver

It is clear from earlier sections that raised BMI is the most important risk factor for fatty liver. As BMI is partly determined by lifestyle and voluntary behaviour, it is of some interest to determine whether or not a change in BMI over the period of the study is associated with a concomitant change in fatty liver status for the individual patient. The analysis of this question is confined

TABLE 66 Terms retained in the logistic model incorporating LFTs, including results of Wald tests

Terms in model	df	Chi-squared	*p*-value
Age group	5	19.36	0.0016
BMI	3	95.98	0.0000
Ethnic group	3	7.52	0.0571
Alcohol × sex interaction	5	12.35	0.0303
ALT	1	66.80	0.0000
Albumin	1	18.24	0.0000
Alcohol[a]	5	9.78	0.0816
Sex[a]	1	0.01	0.9192

a These calculations have been made using all of the available data. Arguably restriction to the 'non-specific' group would be preferable, and makes some impact on the ORs, although there is no evidence that diagnostic group is independently predictive of fatty liver ($\chi^2 = 1.15$, 3 df, $p = 0.7655$). When this is done, the general pattern of results is unaffected and ALT and albumin once more emerge as the most important predictors of fatty liver.

TABLE 67 Model summary

Terms in model	df	Chi-squared	*p*-value
Fatty liver at FU1	1	97.92	0.0000
Age group	2	7.91	0.0192
BMI	2	22.45	0.0000
Ethnic group	3	3.98	0.2639
Alcohol	1	3.74	0.0531
Sex	1	0.49	0.4823

TABLE 68 Model summary

Terms in model	df	Chi-squared	*p*-value
Fatty liver at FU1	1	77.76	0.0000
Age group	2	4.19	0.1230
BMI	2	17.74	0.0001
Ethnic group	3	2.86	0.4137
Alcohol	1	1.45	0.2283
Sex	1	0.10	0.7518
ALT	1	8.35	0.0039
Albumin	1	1.94	0.1639

to the 'non-specific' diagnostic group. *Table 69* suggests an association between even small reductions in BMI and improved liver fat.

The association was investigated by an ordinal logistic regression analysis, using a seven-level outcome variable, defined as the difference in the ordinal number of the liver fat category between FU1 and FU2, i.e. a measure of liver fat improvement ranging from –3 (representing a change from 'normal' to 'severe') to +3 (i.e. a change from 'severe' to 'normal'). This outcome was regressed on percentage change in BMI, with a marginally significant result [$p = 0.030$, odds ratio (OR) = 0.76 (95% CI 0.60 to 0.97)] per 10 percentage points change in BMI.

Change in alcohol consumption (represented as a difference in units per week on the square root scale) was added to these models, though without any significant, or near significant, finding (*Table 70*).

These results are necessarily inconclusive, but furnish some confirmatory evidence for the benefits of weight loss on fatty liver.

It is of great interest to explore whether or not the finding of a fatty liver prompts weight loss. There was a non-significant change in BMI in the hypothesised direction. The main weight change in the fatty liver group is –0.4% and in the non-fatty liver group is 0.2% ($p = 0.30$).

Other ultrasound features

At FU1, the liver was abnormal in size in 58 cases (4.5%), and was large in all but two of these. Eight patients had a diagnosis of diffuse cirrhosis.

TABLE 69 Changes in BMI and liver fat within patients[a]

Liver fat from FU1 to FU2	*n*	BMI (kg/m^2), mean (SD)		
		BMI at FU1	Change in BMI from FU1 to FU2	% change (within patient)
Improved liver fat	129	32.4 (6.2)	–0.5 (2.1)	–1.3 (6.3)
Unchanged liver fat	397	28.3 (5.2)	–0.0 (1.9)	0.1 (6.8)
Worsened liver fat	80	32.1 (5.7)	0.1 (2.3)	0.5 (6.4)
Total	606	29.7 (5.8)	–0.1 (2.0)	–0.1 (6.7)

a Liver fat improvement is assessed with reference to the four ordered categories from the sonography reports (normal, mild, moderate, severe).

TABLE 70 Liver fat improvement and BMI[a]

Term	OR	95% confidence limits		*p*-value
% change in BMI (÷ 10)	0.77	0.60	0.98	0.032
Change in alcohol intake (square root)	0.97	0.90	1.05	0.503

a Ordinal regression results incorporating alcohol consumption.

A focal lesion was found in 106 cases (8.2%), but in only 21 cases was the lesion suspicious (20) or obviously malignant (1). The gall bladder was identified in 1150 cases (90%) and gallstones were detected in 191 (17%) of those.

The extrahepatic bile ducts were dilated in 29 of the 1123 patients in whom they were seen (2.4%), and the mean ALP was higher in these cases (271 vs 203). The difference in ALP was even greater when 12 out of 1230 (0.98%) where the intrahepatic duct was dilated (364 vs 203).

Chapter 5

Substudies

Psychology 1: effects of positive tests

Background

Chronic liver disease is often asymptomatic or associated with non-specific symptoms and its early diagnosis is usually through the use of blood-based LFTs, which are routinely requested in primary care. Although the result of an LFT might indicate serious liver pathology, an abnormal result is much more often a chance finding, with a predictive value of < 5%. In fact, as we pointed out earlier, the proportion of people who really benefit from an abnormal LFT result is much smaller than 5%. Most of the cases of haemochromatosis and PBC identified in BALLETS appeared to be progressing at a very slow rate and were likely to have lead times longer than patients' remaining lives. There must be great doubt about the benefits that would have arisen from identifying four cases of metastatic cancer. The 1% of patients with chronic viral hepatitis really did stand to benefit, but over 1300 positive test results is a considerable number when only 13 patients are to benefit. This 'yield' would be especially worrying if it was associated with anxiety sufficient to impair quality of life. On the other hand, long-term benefits might accrue if LFT results prompted people to adopt healthier lifestyles – a situation that could arise if LFT results were used to reinforce behaviour change advice in addition to their rather minimal diagnostic value. The psychological consequences of testing are therefore important. In this section we consider the effect of psychological testing on anxiety. We considered the effects of LFT results on behaviour in *Chapter 4* and will do so again in this chapter (see *Psychology 2*: *effects of results on behaviour*).

A psychological evaluation was added to the main study to monitor any psychological harms created by reporting abnormal LFT test results to patients and informing them of their ultrasound results. Previous evaluations of the process of screening report negative effects on psychological outcomes including anxiety, depression and reduced quality of life in the short term, but little effect in the longer term.[43,44] Screening for potential liver disease, however, has not been investigated and this study therefore examined the psychological effects on patients.

Methods and rationale

Procedures

Participants completed psychological assessment questionnaires at two points: at recruitment, following results of the index test (T_1), and again at 2 years (T_2).

A pilot study was implemented to inform development of psychological questionnaires (T_1 and T_2) for use in the main study. This phase gave the research team a clearer idea of the ways that patients tend to think about and respond to abnormal test results.

The first questionnaire (T_1) was ready to administer 11 months after the recruitment phase commenced, and when the first 250 patients had already been recruited to the study.

There were slight differences in the administration of T_1 at individual practices, as the study clinical process was modified to merge with routine practice. All changes to the clinical process were approved by ethics and local research and development committees.

General practitioners at three practices invited patients to take part in the study and practice administration staff posted information sheets and T_1 questionnaires to patients. At the remaining eight practices, GPs identified patients meeting the study criteria, and provided the research team with a patient list. The research team telephoned listed patients to invite them to a clinic and posted T_1 questionnaires and other study documentation to patients prior to clinics.

Patients who had not completed a questionnaire were offered another opportunity on arrival at the clinic. Reminder letters and additional T_1 questionnaires were sent by study psychologists 1 week following non-response.

Two-year follow-up (T_2) questionnaires (*Appendix 1*, *section 10.10.d*) were posted to patients, along with information concerning their study appointment. Patients were asked to complete the questionnaire and either return it by post or bring it with them to their study appointment. Again a further supply of questionnaires was available at GP surgeries so that patients who had not completed a questionnaire could be offered another opportunity to do so.

Outcome measures

Disease-specific worry

Disease-specific worry was assessed using the item 'How worried are you about the health of your liver?', adapted from Lerman's cancer-specific worry scale.[45] Participants responded on a seven-point scale ranging from 1, 'not at all worried', through to 7, 'extremely worried'.

State anxiety

This was assessed using the short form of the Spielberger State Trait Anxiety Inventory,[46] in which participants are asked to rate six mood states: calm, tense, upset, relaxed, content and worried. Items are scored on a four-point scale ranging from 1, 'not at all', to 4, 'very much'. Scores were transformed to provide a scale ranging between 0 and 100, with higher scores indicative of higher state anxiety.

Self-assessed health

This was assessed using responses to five items from the Short Form questionnaire-36 items (SF-36) health survey[47] comprising a single item rating of self-rated health ('Would you say your health is: excellent, very good, good, fair, poor?') and four further items: 'I seem to get sick a little easier than other people'; 'I am as healthy as anybody I know'; 'I expect my health to get worse'; 'My health is excellent'. Scores were transformed to provide a scale range from 0 to 100, with higher scores indicating higher self-assessed health.

Results

Not all patients were offered a psychological assessment questionnaire to complete, and some declined. Overall, 527 questionnaires were obtained following the index test (T_1). Two years later, T_2 questionnaires were returned by 596, of whom 243 had returned baseline questionnaires.

Table 71 shows the demographic and clinical characteristics of the 527 patients completing the T_1 questionnaire.

Baseline characteristics of patients completing T_1 (as shown in *Table 72*) were compared with those not completing the questionnaire: there were no significant differences. Similarly, there were no significant differences between this cohort and those responding to both T_1 and T_2 questionnaires except for age, with the latter being slightly older (mean age 59.5 vs 56.9 years; *t*-test, $p < 0.01$).

TABLE 71 Demographic and clinical characteristics of the study sample ($n = 527$)

Demographic characteristics	
Age (years), mean (SD)	57.5 (15.5)
Gender, *n* (%)	
Male	296 (56)
Female	230 (44)
Ethnicity, *n* (%)	
White	445 (87)
Other	65 (13)
Social deprivation, IMD score: mean (SD)	36.5 (7.8)
Clinical characteristics	
BMI (kg/m^2), mean (SD)	29.3 (6.3)
Waist–hip ratio	0.93 (0.09)
Alcohol units per week: mean (SD)	14.9 (28.9)
Fatty liver: *n* (%)	
Yes	179 (35)
No	336 (65)
Repeat abnormal blood test result, *n* (%)	
Yes	430 (83)
No	89 (17)

IMD, Index of Multiple Deprivation.

The impact over time of the report of an abnormal LFT was examined in those patients returning questionnaires at T_1 and T_2. *Table 72* shows that both anxiety and worry declined significantly over the 2-year period; there was no change in self-reported health.

As the impact of the report of an abnormal LFT might have been amplified by the subsequent diagnosis of fatty liver following ultrasound, the change in emotional state was examined in those with and without a reported fatty liver. *Figures 16* and *17* show similar declines in anxiety and worry over the 2 years, irrespective of the diagnosis of a fatty liver.

Discussion

The results of this study are in accord with previous research into the impact of screening, that screening might well raise initial anxieties and worries but these soon return to levels within the normal range. Previous research on the emotional impact of screening, however, has largely examined the impact of the screen on the bulk of patients, who are subsequently judged to be negative. In that context, initial anxiety might well be allayed by the reassurance of a negative test. In this study, however, the cohort recruited were those patients who had screened positive albeit with an indicator with predictive significance that was poorly calibrated but not likely to be high. As the GPs had to inform the patients of the results, it is likely that they were reassuring: LFT results were slightly raised, but this was probably of no serious clinical significance. In that sense, patients may have interpreted their abnormal blood test as within normal limits and therefore as a sort of negative.

However, fatty liver, which does suggest the early signs of liver disease, was diagnosed in one-third of patients. This diagnosis was reported to the patients after they had completed their baseline T_1 questionnaire. It might be expected that, if patients had been alarmed by a fatty liver diagnosis, this would have been apparent at 2 years. But again, whatever the initial concerns,

TABLE 72 Emotional and health outcomes at 2 years: mean (SD)

Psychological assessment	T_1	T_2
Self-rated health	68.85 (17.69)	63.37 (15.29)

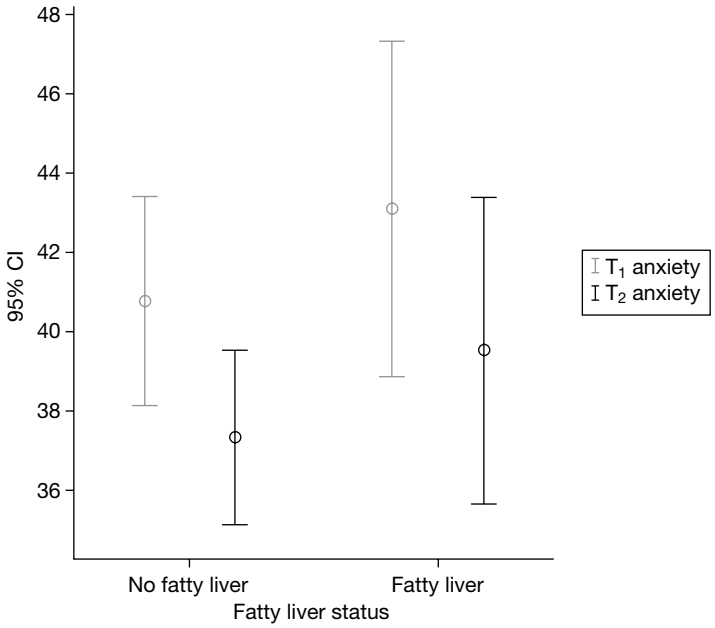


FIGURE 16 Anxiety after the index test and 2 years later in subgroups with and without fatty liver.

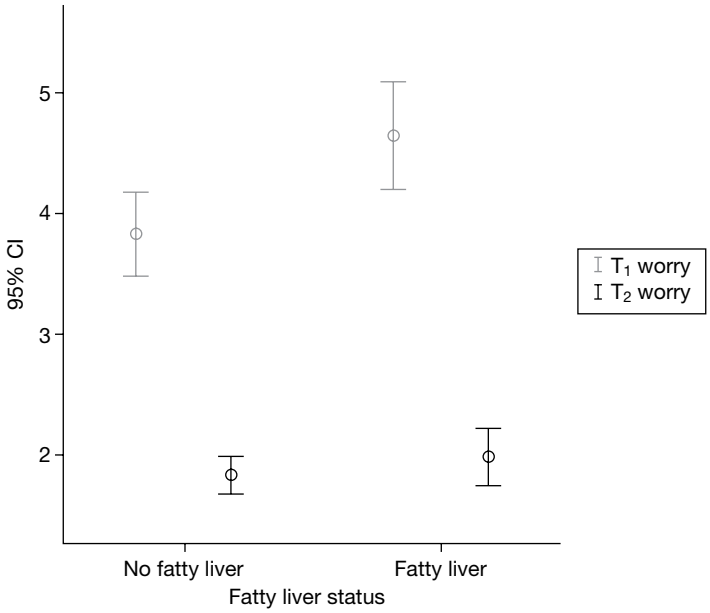


FIGURE 17 Worry after the index test and 2 years later in subgroups with and without fatty liver.

these had clearly dissipated over time. Lastly, if sustained anxiety is a necessary ingredient of behaviour change, then this would suggest that the fatty liver diagnosis would not affect lifestyle to a material degree. If, on the other hand, anxiety is a necessary trigger for change which then becomes self-reinforcing then the initial anxiety may be sufficient. The initial levels of anxiety are slightly higher in those 'with' than in those 'without' a diagnosis of fatty liver.

In conclusion, this study confirms previous research showing that screening has no long-term emotional effects. Where it adds to these findings is that, when the screening result is 'positive' but surrounded by prognostic uncertainty, this too becomes normalised over time and has no long-term effect.

Psychology 2: effects of results on behaviour

Background

The BALLETS study recruited 1300 patients from Birmingham and Lambeth practices. At the initial assessment 40% of study patients were found to have fatty liver on ultrasound. The recognised primary treatments for fatty liver are diet and regular exercise.[48]

The literature on the subject suggests that the 'working alliance' between care provider and patient is important in adopting health behaviours.[49] This alliance is defined by the mutual agreement on goals and objectives and the extent of the emotional bond (liking or trust) between patient and provider.[50] In addition, care providers use a number of verbal compliance strategies in which the subtle use of language can help influence a patient's behaviour.[51] Recently, evidence emerged that moderate- and low-level lifestyle counselling interventions in patients with fatty livers are a practical and effective method of improving health.[52]

At Birmingham follow-up clinics, where patients returned for a repeat USS, BMI and LFT, research nurses had positive anecdotal reports from patients regarding improved drinking, eating and exercise habits. Many reported having had an abnormal first ultrasound. The results were supported by preliminary analysis from the first 277 patients who were followed up at FU2. In the event, an association between a change in mass and a reversal of fatty liver to normal was confirmed in the final analysis (see *Chapter 4*, *The effect of changes in body mass index and alcohol intake on fatty liver*). We therefore sought an extension to conduct a qualitative study of BALLETS patients to better understand a possible modifying effect on behaviour of having an ultrasound showing fatty liver.

Methods and rationale

Based on the above evidence, we conducted a qualitative substudy to explore the patient's experience of participation in the BALLETS study with respect to the finding of fatty liver. We focused on the patient's perception of the results from the initial scan, how the results were imparted to the patient and whether the finding of a fatty liver led to making any lifestyle changes. Therefore, the main aim of this substudy was to understand the overall experience of taking part in the BALLETS study with special reference to the psychological impact of the finding of fatty liver on USS.

Forty patients who participated in the BALLETS study and attended for initial clinics (FU1) and follow-up clinics (FU2) were invited to be interviewed. These 40 patients were divided into four subgroups (*Table 73*) according to whether or not their BMI had reduced and whether or not their ultrasound showed fatty liver.

TABLE 73 Recruitment figures for qualitative substudy of BALLETS

BMI	Fatty liver at FU1	Non-fatty liver at FU1
Unchanged at FU2	10	10
Reduced by ≥5% at FU2	10	10

Patients were randomly selected across the four categories using BALLETS study participant ID numbers. Patients from all GP practices taking part in BALLETS were included. Because it was expected that not all patients would agree to take part in the substudy, 20 patients were randomly selected for each group. Patients were phoned by the substudy research associate, in list order. If patients could not be contacted or were unable to take part, the next patient on the list was invited, until the final sample of 40 was reached. Patients were sent information sheets after providing verbal consent. An appointment was made for the research associate to interview patients if they were in agreement.

All Birmingham, BALLETS study sonographers were invited to be interviewed to determine their opinions on the consultation process, the methods used to impart the results of the scan and possible implications.

Interview process

During the main visit, informed consent was sought by the research associate (DC), who had received informed consent training. The interview was semistructured in nature. Interviews took 30–60 minutes and were audio-recorded with the permission of the patient. A similar process was used for approaching and surveying sonographers, using semistructured interviews.

Data analysis

Once all the interviews were complete, they were transcribed. The transcripts were anonymised. Each transcript was analysed using a qualitative data analysis method of interpretative phenomenological analysis[53–56] as an attempt to unravel the meanings contained in the transcripts.[54] This method recognises that the meanings that people ascribe to events are the product of interactions between people in the social world.[57] The analysis explored the participants' view of the world adopting an 'insider perspective'[58] of the phenomenon under study.[54] This is in accordance with the guidelines of Elliott *et al.*[59] and Parker[60] for good qualitative research, whereby owning up to one's perspective and assumptions helps readers to interpret and understand the researcher's data.

The transcripts were analysed by the research associate who conducted the patient interviews, the research associate who conducted the sonographer interviews and the research fellow who wrote the study protocol. A process of data triangulation took place with constant comparison of emergent themes and discussion within the research team.

Results

We interviewed 40 participants (see *Methods and rationale*) and five key themes emerged. These are described below and in *Table 74*. Five sonographers were also interviewed in order to gain a greater understanding of the consultation process.

Participants

Theme 1: poor recall of BALLETS research study

Most participants exhibited poor recall regarding their involvement in BALLETS including the results from either of the scans. In fact, some appeared unaware that they had participated in a

TABLE 74 Number of individuals within each of the four groups (of 10 patients) who referred to each theme/subtheme

	Research participation				Health beliefs				
		Theme 2: motivation to participate					Theme 5: lifestyle awareness		
Subgroups	Theme 1: poor recall of study participation	Health concerns	Maintenance of relationship with GP/practice	Altruism	Theme 3: normalisation of results such as LFTs, fatty liver	Theme 4: external factors such as other illnesses	Understanding of what affects the liver	Awareness of healthy lifestyle factors	Proactive behaviour
Group A (no change in BMI/normal liver)	5	4	2	5	5	6	5	7	1
Group B (no change in BMI/fatty liver)	5	6	2	3	5	6	5	9	6
Group C (> 5% reduction in BMI/normal liver)	8	2	1	4	3	4	4	4	2
Group D (> 5% reduction in BMI/fatty liver)	5	4	4	5	7	6	4	9	6

research study. One participant suggested that he confused BALLETS with regular visits to the GP for clinical reasons.

> I don't think there was anything different from perhaps going on other occasions.
>
> (Patient 2B)

Most participants seemed to have problems remembering the time scale and sequence of events associated with the study, with many having poor recall of results obtained from either consultation. Two common explanations for poor recall of study results emerged. Participants believed that if an untoward result was reported they would remember the information.

> I would have remembered if … if they'd said anything derog … you know anything that may have been wrong.
>
> (Patient 2B)

In addition, poor recall of scan results was also explained by the sense of trust engendered by their GP or their practice, which meant that the participant would be contacted if any adverse results emerged.

> I just thought it was all right because I thought that is there was something wrong it would come back to the doctor
>
> (Patient 17A)

Theme 2: reasons for participating in the BALLETS study

Reasons for participating fell into one of three subthemes.

Subtheme: health concerns

Participants felt that involvement in BALLETS would benefit their health.

> … as I say I'm willing to do these things because it's helping me as well.
>
> (Patient 13B)

Some participants were concerned about hereditary health conditions which motivated them to consult their GP leading to a greater awareness of their health and their decision to take part in BALLETS.

> Mum was concerned that there may be something hereditary in our family to do with the heart, so he did a load of tests and then he found that my liver reading was borderline or slightly higher (laughs) or lower than it should have been. I don't really understand much about my liver and anyway it was after that the study contacted me and asked if I'd like to take part so I thought I may as well.
>
> (Patient 18C)

Subtheme: maintenance of the patient care–provider relationship

Participants would take part in research if asked to by their GP to help maintain a constructive relationship with the GP or practice.

> … my surgery have been very good to me over the years and looked after me so it's the least I could do really, so yeah.
>
> (Patient 17B)

Subtheme: altruism

The altruistic nature of taking part in research was identified by most participants.

> … the BALLETS study is there to sort of find out the information they need to, sort of, improve people's lives and to improve people's medical side of things.
>
> (Patient 18D)

Theme 3: result interpretation

In many cases patients recalled abnormal ultrasound findings even though they had not recalled detail of the study process.

> I think yeah, but like, basically … I have got a slightly fatty liver.
>
> (Patient 1D)

Sonographers were reported as assuring the participant that the abnormality had minimal implications for participants' health.

> … they said 'yes, everything was OK, that it was 'a little bit fatty but it was OK' …
>
> (Patient 4C)

Theme 4: external causality of BALLETS study results

During follow-up clinics, anecdotal evidence emerged that many participants had lost weight and engaged in making lifestyle changes following the initial consultation. However, sometimes changes in lifestyle or weight loss were perceived to be associated with factors other than the liver ultrasound or LFT results, such as existing health problems or medication.

> … because I was taking my tablets as well … at the time. But since then I've stopped really taking my tablets … I'd kind of lost weight … I felt better in myself and I wanted to be in control of what I was doing, and not want the tablets to be in control.
>
> (Patient 7A)

Theme 5: lifestyle awareness

Three subthemes were identified within the emergent theme of lifestyle awareness. It was noted that more individuals in groups A and C (with fatty liver) discussed their awareness of the factors that contribute to a healthy lifestyle (see *Subtheme*: *conviction regarding own lifestyle*) and engaged in proactive behaviour (see *Subtheme*: *proactive behaviour*) as described in *Table 74*.

Subtheme: personal understanding of what affects the liver

Participants believed that a poor diet or excessive alcohol consumption can cause problems to the liver.

> I suppose in some respects I … could have expected a problem with me liver … because I used to be a very heavy drinker over a long period of time.
>
> (Patient 10A)

Subtheme: conviction regarding own lifestyle

Participants displayed a good understanding of what constitutes healthy living and were confident that they maintained a healthy lifestyle.

> I knew I didn't drink alcohol or have never drunk it very much … I've never been one to eat a great deal of fatty food.
>
> (Patient 1A)

Subtheme: proactive behaviour

For some participants, BALLETS appears to have encouraged beneficial lifestyle changes.

> … he [GP] said that 'a lot of people get these fatty liver cysts … possibly a change in diet might help' … which I've since tried to do.
>
> (Patient 13B)

> They said … that it was a little bit fatty … From the questionnaires … it asks you various questions about your food intake, your alcohol intake … it's one of those things that you take on board and you tend to live that sort of lifestyle … you don't do excessive alcohol and you don't do excess of foods …
>
> (Patient 4C)

Sonographers

Five sonographers were questioned about their role in BALLETS. We were interested in exploring details of the consultation, particularly regarding their interaction with the patient, and whether there was a difference in their approach to the consultation and participant, in comparison with routine hospital consultations. Sonographers reported differences in the equipment and the setting though these made little difference to the scan.

> … considering we weren't in our normal environment we found that we got quite good because every time we recalled a patient to the QE we didn't actually gain any more from it, a lot of the time.
>
> (Sonographer ID S1)

When asked if a scan in the secondary-care environment was comparable to the study scan, attitudes of sonographers varied.

> Well it's a totally different thing really, because with the study the patients were going through questionnaires, blood tests, explanations.
>
> (S3)

> … but I would say perhaps comparable – the patient knows why they're there.
>
> (S1)

Sonographers would inform patients of scan results depending upon their clinical significance, being careful not to exceed their own clinical expertise.

> But if they come in for query 'have you got secondary cancer in your liver', I wouldn't tell a patient they had that for instance, 'cause it's not my place to, you know. They're going to want to ask lots of questions and I don't know the answers … but then you know, in the same way if they've got gallstones, I'd probably say 'yeah you've got gallstones'.
>
> (S3)

Discussion

The growing commitment to patient involvement in research has been reflected by the expanding literature on the aims and core features of research from a patient's perceptive. There is, however,

scant literature on the impact of research participation on patients, particularly regarding beneficial health effects resulting from behavioural and lifestyle changes.

Both compliance and adherence to lifestyle changes are influenced by a number of factors. Our initial hypothesis was that the results from the first ultrasound consultation acted as a powerful driver in motivating people towards improving behavioural and lifestyle factors, reflected in their change in BMI. Across the four groups, recall of participation in BALLETS was poor. Participants were uncertain when and how they received results, if at all. It was evident that in this context the impact of the consultation with the sonographer appeared to be minimal. Evidence elsewhere has indicated that even in the most serious of clinical cases patient recall of clinical information is poor:[61] between 40% and 80% of medical information presented by health-care professionals is forgotten by patients.[62] A number of factors can contribute to this lack of recall, such as complicated medical terminology, educational status of the patient and the means by which the information is presented.[62] Existing literature indicates that participant recall of Central Office of Research Ethics Committees (COREC)-approved, informed consent information is poor, even among those with medical training.[63] The relevant details of BALLETS were contained within the information sheet, although the complicated constraints of a COREC-approved information form may have inhibited understanding and retention of this information and greater engagement with the study by BALLETS participants, including understanding the context of the consultation and implications of the results of the scan.[64]

Individuals can be motivated to participate for several potentially interacting factors, including the likelihood of improved clinical care as a result of their involvement,[65] and social influences such as a desire to please the practitioner.[66] As elsewhere, for the majority of those interviewed, participation was altruistically motivated.[67] The results were seen as of relevance to the study team and not to them as individuals.

Improving communication between patient and care provider, including adopting a less formal approach, can increase the likelihood of adherence to treatment and behavioural regimes.[68] The potentially more relaxed consultation between participant and sonographer exemplified a more informal discourse. Sonographers would impart information of low clinical impact during the scan, and as a result participants, even those in whom abnormalities were observed, reported that results were underplayed and as a consequence they felt no obligation to alter their behaviour.

Participants were aware of the requisites for a healthier lifestyle, some because of existing health conditions and others as a result of their participation in BALLETS. This capacity to obtain, process, and understand basic health information can then lead them to make appropriate health decisions and may account for the observed changes in liver status and BMI. The results within theme 5 indicate that there may be a relationship between being diagnosed with a fatty liver and an increase in awareness of healthy lifestyle factors. However, we did not find strong evidence that patients were powerfully motivated to change lifestyle by the finding of a fatty liver on ultrasound.

Sociology of testing: an exploration of the clinical and non-clinical motives behind the decision to order a liver function test

Background

The numbers of diagnostic tests used in public health systems are increasing in most countries[69] (by 10% per annum in the UK over the last 3 years).[70] The proportion of tests originating from GPs is also increasing; requests from GPs accounted for 37.2% of biochemistry tests in 2002, compared with 41.7% in 2005.[70]

Increases in the number of tests ordered could be because of a number of factors: an older population,[71] increased range of tests available, increased expectations of patients and guidelines promoting multiple test use.[72] Increased testing inevitably produces more positive results, leading to knock-on investigations, adding further to the number of tests ordered.[71,73]

The motivation for ordering a test can be conceptualised under two non-exclusive categories: technical factors related to the diagnosis and management of disease and social factors. The latter include reassurance for patient and/or doctor, patient expectation and maintaining the doctor–patient relationship.[9,74,75] Guthrie[75] found that non-technical motivations behind blood tests were commonly viewed as relevant by GPs, particularly when used to reassure the patient or the doctor, and van der Weijden *et al.*[76] concluded that GPs order tests for many purposes and that non-medical motives were viewed as rational and legitimate.

Liver function tests are a good example of inexpensive tests that are frequently ordered in patients with non-specific symptoms, such as tiredness or upper abdominal discomfort.[77] LFTs are often carried out when the prior risk of disease is low, thereby yielding a high proportion of false-positive results. As LFTs are frequently used despite their lack of specificity, we decided that they would provide an interesting model through which to explore GP motivations behind test ordering.

Methods and rationale

Sample

The study group consisted of GPs participating in the BALLETS study from South Birmingham Primary Care Trust. BALLETS is a National Institute for Health Research (NIHR) HTA study of the value of abnormal LFTs among patients in primary care with non-specific symptoms.

Recruitment

Practice managers in the eight practices participating in the BALLETS study were approached and asked to consult their constituent GPs to ascertain their willingness to take part in the study. GPs from six practices elected to participate. The GPs (29 in total) at each of these practices were supplied with an information sheet and consent form. Interviews were arranged with consenting GPs at a time and date of their choosing.

Interviews

Semistructured interviews with a topic guide and prompts were used. The themes in the topic guide were identified from the existing literature concerning the test-ordering behaviour of GPs and included the impact of a GP's formal and experiential knowledge base, social influences, defensive medicine and characteristics of the test and order process.

Analyses

The interviews were digitally recorded and transcribed by the author. Following initial discussions within the study team the principal codes were determined. The constant comparative method[77] was used, leading to the inclusion of an additional question, addressing the use of LFTs as a tool for modifying patient behaviour. All GPs preferred a telephone interview, usually immediately following morning surgery. Interviews were carried out by the same individual. Saturation was reached after 11 interviews.

As a way of ordering the themes and categories, we adapted the 'attitude–social influence–efficacy' model defined by Kok *et al.*[78] and used by van der Weijden *et al.*[76] The model is based on the assumption that a GP's intention to order a test can be determined by a number of factors, which we placed in one of two broad categories. The first is internal, and includes the

themes of expectation of efficacy and general attitude toward LFTs (positive or negative). The second category contains external influences, and consists of the themes of social influence, test characteristics and defensive medicine.

Results

General practitioner characteristics

Breakdown of GPs by age, sex, duration of service, and part-time versus full-time working is given in *Table 75*. The participating GPs were heterogeneous with respect to these attributes.

Motives behind a decision to order a liver function test

Table 76 shows the themes and subthemes mentioned by each respondent represented by '×' according to Kok *et al.*'s classification.[78]

Internal influences on the decision to order a liver function test

Expectation of efficacy

The expectation a GP has of his or her own ability to correctly diagnose a patient and order the correct test at the apposite time is a function of the knowledge gained from formal training, and knowledge in the form of experience gained as a practising GP.[78]

Formal knowledge

Clinical reasons for test ordering were mentioned spontaneously by all interviewees. These included decisions based on a patient presenting symptoms of liver disease such as jaundice or pruritus, and medicines known to affect (or be affected by) liver function.

> If someone is jaundiced or suffering from weight loss or something like that …
>
> (GP8)

> I would tend to tick someone's LFTs if I was checking someone's cholesterol. If they are going to go on a statin then I am going to need to know what someone's LFTs are like.
>
> (GP2)

TABLE 75 Characteristics of participating GPs

GP study ID	Practice	Gender (M/F)	Age	Part time (%)/full time)	Years practising as a GP (including training)	Years at current practice
1	C	M	31	Full time	2 years 6 months	1 year 8 months
2	D	M	36	Full time	9	8
3	B	M	41	Full time	12	10
4	C	M	52	Full time	20+	20
5	E	M	54	66	25	24
6	A	F	33	Full time	6	6
7	A	F	38	75	11	9
8	F	F	41	55	14	14
9	F	F	43	Full time	15	8
10	C	F	46	77	16	15
11	D	F	58	50	28	28

F, female; M, male.

TABLE 76 Pattern of response by GP to themes and subthemes

	Internal						External						
	Expectation of efficacy				General attitude to LFT		Social influence			Test characteristics			
GP no.	Formal knowledge	Craft knowledge	Personal reassurance	Overordered	Positive	Negative	Colleagues	Patients	Research participation	a	b	c	Defensive medicine
1	×		×	×		×		×					
2	×				×							×	×
3	×	×	×	×	×							×	×
4	×			×	×			×	×				
5	×	×	×	×	×			×		×			
6	×			×	×								
7	×	×			×		×	×	×		×	×	×
8	×	×	×		×			×					
9	×	×		×	×			×	×				×
10	×			×	×			×		×	×	×	×
11	×	×			×								

a, cost; b, invasive nature of test; c, order process.

Craft knowledge

Tests were ordered for a number of personal reasons related to the GP's beliefs and experiences. Evidence emerged during the early interviews that LFTs were used to incentivise certain patients to make behavioural modifications necessary to improve their health. Notably, GPs would order LFTs for patients suspected of drinking too much alcohol, in the expectation that an abnormal test result would provide evidence of impending self-harm and thereby prompt a change in behaviour.

> If someone has got alcohol related problems … and the LFT does come back as abnormal then I would use that as a way of saying 'look, what you're doing is affecting your liver and you're at a stage where you can do something about it'.
>
> (GP8)

> I've got one particular alcoholic who successfully became a teetotaller. His GGTs were up in the sky and then came down to normal or near normal again and with his permission I use a printout of his GGTs going up and down to try and motivate other patients.
>
> (GP11)

Personal reassurance

GPs we interviewed conceded a lack of complete confidence in their ability to identify a condition by using physical examinations and medical history and so sought reassurance from tests such as the LFT.

> … I get the feeling that the more experienced you become the more you do a lot more tests because you know what can happen.
>
> (GP3)

> Rather than just keep saying that 'yes, everything's OK and its just anxiety which is *x*, *y*, *z* and more of a psychological and mental component', sometimes you do the blood test so that you're more reassured …
>
> (GP8)

General attitude to liver function text

Despite the fact that none of the analytes in an LFT can provide a definitive diagnosis nor is necessarily specific for liver complaints, 10 out of the 11 GPs interviewed held positive opinions on the effectiveness of the LFT, though one recently qualified GP was less convinced.

> They are a useful tool, especially for a patient that is unwell and you can't work out what is going on.
>
> (GP10)

> I think they're pretty useless to be honest. I think they throw up a lot of spurious results, most of which don't mean anything at all.
>
> (GP1)

Overordered

It became apparent that those interviewed felt that LFTs were not used as efficiently as they might be. Drawing comparison with other blood tests, they felt that too many were being ordered.

> I think like most tests we order too many.
>
> (GP4)

External influences on the decision to order a liver function test

Social influence

The external sources affecting the motivation to order LFTs included patient influence, defensive medicine, and characteristics of the test and ordering process.

Patient influence

Ordering an LFT can be used as a way of reassuring anxious patients that their concerns are being taken seriously, so maintaining the working alliance between patient and doctor.

> I do think that patients do feel on the whole that they're being taken more seriously if you stick a needle in them.
>
> (GP7)

One of the GPs in our sample used LFTs alongside other blood tests as a way of managing patients who are presenting psychosomatic complaints.

> Sometimes a patient's come and you're sure that they have a psychosocial problem or even depression … but you take a blood test and they're all normal. That's actually quite useful information to feed back to the patient.
>
> (GP10)

The experience of private health care can also serve to raise levels of expectation amongst patients.

> They may go to a private consultation and have panels of blood tests done so they have an expectation that they have regular blood tests.
>
> (GP10)

Research participation

A theme that we had not anticipated was introduced by three of those interviewed, who mentioned the effect of taking part in the primary BALLETS study on their attitude to LFTs. This had led them to question their use of the LFT and helped them focus on the underlying physiology behind the test, increasing the confidence in their evaluation of the LFT.

> In light of the BALLETS study I'll probably find them less useful. If I get a slightly abnormal liver function test I'm probably not going to worry about it.
>
> (GP4)

> Since we've done the BALLETS I feel much more able to understand what's going on.
>
> (GP7)

Defensive medicine

Negative defensive practice was observed in our sample.

> We have to do that [LFT test]. Because if someone ends up with liver disease because they were on statins and you didn't do the test then you can end up in big trouble.
>
> (GP10)

Test characteristics

Cost

Currently there is less financial pressure on investigation than on prescribing and referral. The lower financial impact of ordering a test means that decision can become easier.

> Instead of just doing one, checking renal samples, you might check the whole lot: kidneys, liver, bones, because it doesn't cost any more.
>
> (GP10)

Order process

The ease with which an LFT can be ordered can influence the decision-making process.

> I think one of the reasons [we order too many] is because of the tick box, you end up doing a profile on people and you end up taking them.
>
> (GP3)

Invasive nature of the test

Ordering an LFT has little impact on the patient, particularly if other tests are being ordered, and so lowers the decision-making threshold for ordering LFTs.

> I will do an LFT because it's a relatively non-invasive test isn't it, really, to be honest? It's not like a colonoscopy.
>
> (GP7)

Discussion

Summary of main findings

Our study sample consistently admitted using an LFT for routine monitoring of medication and liver-specific diagnostic reasons. In addition, a number of non-clinical motives behind the test-ordering decision were explored. These include the 'internal' influences stemming from their own expectations of efficacy including their clinical training and the need for personal reassurance. Two novel findings also emerged. First, it became clear that some GPs used LFTs as a means to actively influence unhealthy (eating and drinking) behaviours. The pattern of alcohol consumption in the UK is changing; young people are drinking more and from an earlier age,[79] as are women,[80] with potentially large costs to their health and to the NHS. The use of LFTs to promote lifestyle change among heavy drinkers is an interesting idea that warrants further study. Not only can an LFT provide hard evidence of harm, but repeating the LFT after a period of reduced alcohol consumption can also confirm improvement in a patient's condition. However, there are also potential dangers in using LFTs in this way, as a normal result may have a perverse effect by providing false reassurance. The other novel finding was that active participation in research (i.e. the BALLETS study) led a number of GPs to reappraise their use of LFTs. It will be interesting to observe any effect of the result of the BALLETS study on the test-ordering behaviour of participating and non-participating GPs.

In addition, there are the 'external' influences, for example social interaction with patients, characteristics of the test and the litigative pressure for defensive practice.

Strengths and limitations of this study

This study has for the first time explored the underlying influences behind a GP's decision to order a LFT. LFTs are somewhat unusual in that each 'test' is composed of a panel of five to eight analytes, so it could be seen as a kind of 'catch-all'. Moreover, the tests are fairly sensitive to alcohol abuse and (to a lesser extent) overeating. A limitation is that none of the behaviours documented was observed directly by the research team and recall bias may have been introduced.

Relationship to existing literature

The GPs in our study who had experience of discovering something unexpected said they were more likely to test in the future, a heuristic known as the 'availability bias' in the psychological literature.[81] This may explain a positive correlation between experience and propensity to order seven common blood tests. Another key factor in testing for unexplained complaints is the need to maintain the doctor–patient relationship by meeting user expectations. We found that GPs frequently ordered tests to reassure patients and to signal to them that they were being taken seriously. Also, as reflected in this study, blood tests, such as LFTs, can be used as a way of managing a patient with psychological problems.[82]

The drive towards patient-centred care[83] means that individuals are increasingly aware of their role as customers and may engender a sense of entitlement. Evidence of patient pressure was observed in this study, and it has been reported that GPs are more likely to test if a patient is assertive and actively asks for a test.[84] GPs in our study also acknowledged reassuring a worried or concerned patient by ordering a test. This may increasingly be the case, as many patients now see a blood test as the most reliable diagnostic tool at the GP's disposal.[82,84,85] The countervailing risks of embarking on an investigation 'cascade', triggered by a false-positive test, seem to weigh less highly with patients.

Many in our study group felt an increased need to practise defensively, and other research in the UK has shown that GPs here are now more likely to pursue diagnostic testing as a result of fear of litigation.[86]

A number of GPs in our study provided comments on the ease with which an LFT can be ordered. Studies elsewhere have demonstrated that reducing the options on the test order form can reduce the total number of tests.[87,88] It has also been shown that the design of laboratory request forms can influence the decision to order a test.[87,88] Similarly, the low cost and non-invasive nature of LFTs means that the GP can order with minimum impact on budgets.

Implications for future research or clinical practice

As described, a number of elements interact to prompt frequent orders of LFTs. The need that patients feel for reassurance, and the need for investigation perceived by GPs in our study, could be driven in part by the 'democratisation' of medical information as web-based sources of medical data continue to proliferate. This is a situation unlikely to change soon, and all GPs who participated in this study felt that the number of LFTs ordered was higher than necessary. However, the GP cannot be solely influenced or restricted by formal guidelines and training, as this approach would exclude the social and consultative nature of the doctor–patient relationship and the carefully constructed working alliance that exists between GP and patient. The character and maintenance of this relationship often drives the testing process, beyond narrowly defined clinical need. The GP's acceptance of the need to balance what the patient expects with what the patient requires is further influenced by the low financial and temporal costs of ordering these

tests, their non-invasive nature and the increasing threat of litigation, if failing to use correctly the diagnostic tools at their disposal. The study illustrates that social and behavioural reasons are strong motivators to order a LFT and may even take precedence over clinical motives on some occasions. In particular, the use of LFTs as a tool to increase uptake of health-promoting behaviour could be further explored. Therefore, although an educational change to reduce testing among patients and their doctors might be the theoretical optimal solution, the above range of factors favouring test use suggests that large-scale, rapid change is unlikely to occur.

Chapter 6

Decision analysis

Background

Liver function tests are ordered in large numbers in primary care

Liver function tests comprise a panel of five to eight analytes that are processed inexpensively in large batches. LFTs are one of the most commonly performed 'blood tests' in primary care, such that in 2003 the laboratory at University Hospital Birmingham received 67,182 requests for LFTs from 83 GP practices that serve a population of 300,000 people.[89]

Enigmatic responses to abnormal liver function tests primary-care settings

An abnormal LFT may signify a serious disease that can be identified only through further testing. These conditions include liver diseases (such as PBC), diseases of other organs (such as Paget's disease of bone) and multiorgan diseases (such as haemochromatosis). However, the majority of people with an abnormal LFT result in primary care settings will not have any such previously undetected disease. They will have either no disease at all or will be manifesting the effects of alcohol abuse or obesity. The doctor is likely to be aware, or at least suspicious, of these behaviours when ordering LFTs, but this does not exclude the presence of other diseases that may aggravate liver damage. There is thus a real question about which specific further tests, if any, a GP should order when an abnormal LFT result is obtained in a patient with non-specific symptoms, or as a result of routine testing. In some cases there may be a clear indication for further tests. For example, in a patient with a family history of haemochromatosis, iron saturation should be measured. In some cases, the pattern of LFT abnormality may suggest a diagnosis – for example, an isolated raised unconjugated bilirubin suggests Gilbert's disease, whereas a high blood level of ALP is indicative of PBC. In most cases, however, no unambiguous clinical indication for follow-on testing exists. The literature deals mostly with the pattern of abnormality given a diagnosis, rather than the probability of the various diagnoses given a pattern of abnormal LFTs. It is therefore not surprising that guidelines for GPs[3,10,90–93] confronted with an abnormal LFT result in patients with non-specific symptoms or detected fortuitously are inconsistent, or that the way GPs in which respond has been found to be eclectic.[94] A point on which guidelines do agree is that the LFT panel should be repeated following an abnormal result.

Criteria for selection of a topic for decision analysis

If there is any particular previously unrecognised disease that a patient would wish to have excluded by further testing, then it will have the following features:

1. It is a serious disease.
2. It is treatable in the prodromal phase.
3. Failure to identify the condition can lead to permanent damage.
4. It can be diagnosed with a high specificity by a familiar and inexpensive test.
5. It is among the more prevalent of the serious diseases.
6. It is not a condition, like alcohol misuse or obesity, which can be diagnosed from history and examination.

Viral hepatitis

We discern that chronic viral hepatitis is the prime candidate based on the above criteria. It is a massive problem worldwide[95–97] and *Table 77* shows that it is the most common of the specific liver diseases in the UK population after alcohol damage. Moreover, chronic viral hepatitis can be reliably confirmed or excluded by means of a relatively inexpensive blood test.[98] The disease has a prodromal period lasting many decades and is eminently treatable if caught early, thereby averting cirrhosis and liver cancer.[99]

The purpose of the decision analysis described here is to inform the selection of an efficient strategy for the diagnosis of chronic viral hepatitis. Such a strategy should optimise the trade-off between detection rate and cost.

Methods and rationale

Testing strategies

A simple decision tree was constructed in Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) to enable costs per case detected to be calculated for seven strategies.[100] The strategies were developed in consultation with a GP and hepatologist (PG and JN, respectively), who were aware of the relevant literature and guidelines.

The root decision (or starting point) of the tree is the discovery of an abnormal LFT result in primary care where the patient does not have known or self-evident liver disease. From the root node we identified seven decisions that may be considered by a GP under such a scenario:

- *Strategy A* Repeat the LFT panel and then perform a specific test for viral hepatitis if an abnormality is still present on retesting. This could be considered the intuitive response by a GP on receiving an abnormal LFT result in a patient without the indictors of a specific disease, and is the strategy recommended in the literature.[10,90–93]
- *Strategy B* Perform a viral test in all patients with an abnormal ALT. The rationale for this strategy is that ALT is the most specific indicator of viral hepatitis[10] and has been recommended as the testing criterion by other authors.[28,101,102]
- *Strategy C* Select ALT as the trigger for viral testing, but nominate a higher threshold, at twice the ULN as recommended by Jamali *et al.*[103] This is also the threshold for instigating viral therapy for HBV in certain treatment guidelines.[104–106]
- *Strategy D* Perform a test for viral infection in all patients who originate from a country with an intermediate or high prevalence of viral hepatitis according to WHO criteria.[107–109] Screening has been shown to be cost-effective for people who were born in intermediate- or high-prevalence countries and it is likely that testing would be more cost-effective still in a population with abnormal LFTs.[106,110]
- *Strategy E* Combine the two previous strategies by testing those who have an ALT level exceeding twice the ULN *and* who also originate from an intermediate- or high-prevalence country.
- *Strategy F* Test all patients from prevalent countries as well as those with an ALT level exceeding twice the ULN.
- *Strategy G* Test all patients for viral hepatitis irrespective of the type or extent of abnormal LFT results.

There is also an option to take no action with respect to viral hepatitis, and although this may be a sound decision in some cases, for example when a LFT is ordered in the hope that a positive result will prompt a reduction in alcohol intake, this was not considered here.

TABLE 77 Viral, genetic and autoimmune diseases of the liver (tested for by a 'liver panel'), their prevalence in the British population and diagnostic algorithms[a]

Disease	Prevalence among adult population (%)	Blood tests carried out on all members of the cohort (to diagnose or screen for the disease)	Diagnostic algorithm
Chronic viral hepatitis C	0.42[32]	HCV antibody (HCV Ab)	Viral marker positive
Chronic viral hepatitis B	0.3[33]	Hepatitis B viral markers (HBV surface Ag)	Viral marker positive
Metal storage disease: iron	0.25 (prevalence of phenotype; homozygous plus complex heterozygous)[34]	Iron saturation	Genotype if iron saturation >50%
PBC	0.024[35]	AMA	Raised antibodies and raised ALP level
Autoimmune hepatitis	0.001[36]	SMA	Raised antibodies and raised ALT, AST or globulin exceeding twice the ULN. Confirmed by hepatologist
Metal storage disease: copper	<0.025[37]	Caeruloplasmin	Low levels of caeruloplasmin
A1AT deficiency	<0.025[38]	A1AT	Low A1AT levels followed by phenotype testing

Ab, antibody; Ag, antigen.
a Method by which the diagnosis was made.

In this study, the hepatitis status of all patients was known. Moreover, most had an ALT test result and the results of a repeat LFT panel. Thus, it was possible to evaluate the performance of each of the above strategies.

Populating the decision tree with probabilities/statistical model

All 1236 patients were used in the evaluation of strategy G: but for all other strategies the effective sample size was reduced because of missing data in some of the patient records. Estimates of the proportion of patients undergoing viral tests and the proportion of actual cases detected (sensitivity) were obtained using the sample of patients available for evaluating each strategy. The PPV of a strategy was defined as the proportion of hepatitis cases among those selected for viral testing. Confidence limits for this quantity were calculated using Wilson's method for binomial data.[111]

Estimation of costs

The direct costs incurred at the time of the test were the laboratory costs of the liver function and viral hepatitis tests (Pathology Laboratory Manager, University Hospitals Birmingham NHS Foundation Trust, 2005, personal communication); the GP costs for scheduling each test; and following up on results. Administrative costs were estimated by estimating the time implications for a secretary to add patients to appointment slots and a receptionist to check the patient in for an appointment (MidReC: West Midlands Research Consortium, Department of Primary Care, University of Birmingham, 2006, personal communication; figures correct as of February 2009). The costs are presented in pounds sterling (£) (and were correct for the year 2009). Non-health service costs (patient travel cost and lost earnings) were not measured but are considered in the discussion.

Analysis

The number of cases detected per 100 patients was estimated as the sensitivity of the strategy (cases detected ÷ cases present) multiplied by the prevalence (per 100 patients) of viral hepatitis in the whole sample of 1236 patients. For each strategy, the cost per case detected was then computed as the ratio of the cost per patient to the number of cases detected per patient. The

strategy which minimised this quantity was taken as the base case. For each alternative strategy, the incremental cost-effectiveness ratio (ICER) was computed, defined as the incremental cost per additional case detected compared with the base case. The analysis is deterministic and does not consider the impact of sampling variability. The results of these analyses were compared with published results of cost-effectiveness analysis of screening for chronic viral hepatitis, bearing in mind likely differences between a screening and a diagnostic population. We used this analysis to develop a 'fast and frugal' heuristic,[112] which we offer to readers for their consideration.

Results

Patients

A total of 1344 patients consented to the study; 54 were excluded because they did not match the entry criteria in the protocol, along with a further 54 for whom data on at least one viral hepatitis test were missing (*Figure 18*). This left 1236 patients for this study; 105 of these patients were from Lambeth and 1131 were from Birmingham. The median interval between index and repeat testing was 31 days (IQR 19–52 days).

Chronic viral hepatitis cases

Thirteen of the 1236 patients for whom the test result was available had chronic viral hepatitis – nine had hepatitis B and four had hepatitis C. This gives an estimate of 1.1% (95% CI 0.6% to 1.8%) for the prevalence rate in the primary-care population with abnormal LFTs: only slightly more than the baseline prevalence in the general population (0.7%). The demographic breakdown of patients with and without viral hepatitis is shown in *Table 78*.

The breakdown LFT results in the infected cases is given in *Table 79*. In 10 of these 13 cases, more than one analyte was abnormal. In eight cases, the ALT was abnormal, and it was notably raised in six of those (above twice the ULN). In one case (perhaps detected by serendipity), only protein levels were abnormal and all the enzyme tests (ALT, AST, GGT and ALP) were normal. Eleven of the 13 patients with chronic viral hepatitis had an abnormality on the repeat LFT. In two other cases, there were missing data among the repeat LFT panels. Of the 1113 patients with no viral hepatitis who underwent a complete LFT panel, 169 (15%) reverted to normal.

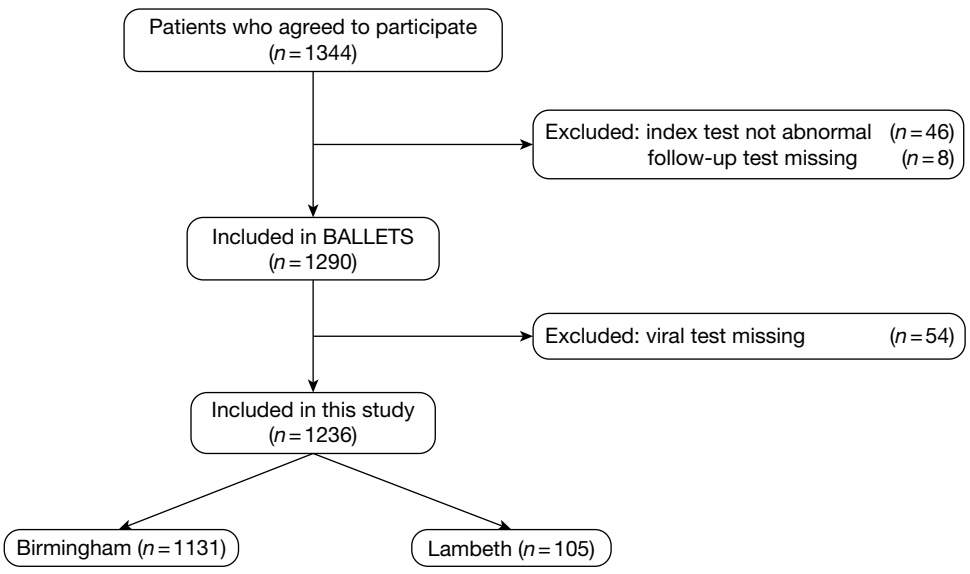


FIGURE 18 Flow diagram of exclusions and inclusions in the study.

TABLE 78 Demographic features of patients with and without viral hepatitis

Feature	Total	Viral hepatitis	Not viral hepatitis
n	1236	13	1223
Age (years), mean (SD)	57.7 (15.2)	54.0 (15.9)	57.7 (15.2)
Sex (*n*, %)			
Male	693 (56.1)	9 (69.2)	684 (55.9)
Female	543 (43.9)	4 (30.8)	539 (44.1)
Ethnic group (*n*, %)			
White	1023 (82.8)	3 (23.1)	1020 (83.4)
Asian	88 (7.1)	5 (38.5)	83 (6.8)
Black	53 (4.3)	3 (23.1)	50 (4.1)
Other	38 (3.1)	2 (15.4)	36 (2.9)
Missing	34 (2.8)	0 (0.0)	34 (2.8)
Reason (*n*, %)			
Abdominal signs/symptoms	69 (5.6)	1 (7.7)	68 (5.6)
Non-abdominal signs/symptoms	302 (24.4)	6 (46.2)	296 (24.2)
Diagnosis – alcohol abuse	17 (1.4)	0 (0.0)	17 (1.4)
Review – CVD	50 (4.0)	0 (0.0)	50 (4.1)
Review – cholesterol	53 (4.3)	0 (0.0)	53 (4.3)
Review – hypertension	147 (11.9)	2 (15.4)	145 (11.9)
Review – diabetes	216 (17.5)	2 (15.4)	214 (17.5)
Review – medication	92 (7.5)	0 (0.0)	92 (7.4)
Medical – review other	290 (23.5)	2 (15.4)	288 (23.5)

CVD, cardiovascular disease.

The country of origin was recorded in 1208 of the 1236 study participants, and of these 170 were born in a country with an intermediate or high prevalence of viral hepatitis (based on WHO definitions of prevalence[107–109]) and 1038 were from low-risk countries. The high-risk group contained 11 out of the 13 patients (85%) with viral hepatitis. None of the 13 cases admitted to use of intravenous drugs at any time.

As expected from the literature, ALT or AST levels when abnormal tended to be more extreme in patients with viral hepatitis than in patients who did not have this disease (*Table 80*).

Diagnostic performance

The sensitivity and PPV of each detection strategy are given in *Table 81*. It can be seen that the recommended strategy (A), of repeating the LFT and then performing a viral test if an abnormality persists, is highly sensitive. However, the predictive value is low (1.15%). Strategy D, simply carrying out a viral test if the patient originates from a high- or intermediate-risk country, detects 85% of cases and has a much higher predictive value (6.47%) than the strategy of repeating the LFT test. The strategy (B) of ordering a LFT if the ALT is raised is not particularly sensitive (67%), nor does it have a high predictive value (1.91%). The more selective strategy (C) of testing if the index ALT is more than twice the ULN has a higher predictive value, but is less sensitive. The best features of strategies C and D are combined in the hybrid strategy F, which achieves high sensitivity (92%) and worthwhile predictive value (5.12%).

Costs and cost minimisation analysis

The cost of the laboratory tests and the practice costs are given in *Table 82*. The average cost per case detected and the incremental costs of detecting each additional case are shown in

TABLE 79 Results of initial LFT for viral hepatitis cases

Case no.	ALT	AST	Bilirubin	ALP	GGT	Albumin	Globulin	Total protein	Repeat LFT	Country of origin (prevalence of viral hepatitis)
HBV										
1	High[a]	High	Normal	Normal	High	Normal	Normal	High	Abnormal	Kenya (high)
2	Normal	Normal	High	Normal	Normal	High	Low	Normal	Abnormal	UK (low)
3	High	Normal	Normal	Normal	High	Normal	Normal	Normal	Abnormal	Pakistan (high)
4	High[a]	High	High	Normal	High	Normal	Normal	High	Abnormal	India (high)
5	High[a]	High	Normal	Normal	Normal	Normal	Normal	Normal	Abnormal	Malaysia (high)
6	High[a]	No result	No result	No result	Normal	Normal	Normal	Normal	Abnormal	UK (low)
7	Normal	Normal	Normal	Normal	Normal	Normal	High	High	Abnormal	Kenya (high)
8	No result	High	Normal	High	No result	Normal	No result	No result	Abnormal	Iraq (high)
9	Normal	No result	High	Normal	No result	Normal	No result	No result	Incomplete[b]	Malta (high)
HCV										
1	High	Normal	Normal	Normal	High	Normal	Normal	Normal	Incomplete	Pakistan (high)
2	High[a]	High	Normal	Normal	Normal	Normal	Normal	High	Abnormal	Hong Kong (high)
3	Normal	No result	Normal	Normal	High	Normal	No result	No result	Abnormal	Jamaica (high)
4	High[a]	High	Normal	Normal	Normal	Normal	Normal	Normal	Abnormal	Somalia (high)

a Denotes where ALT values were greater than twice the ULN.
b Repeat test available for AST and GGT only, both of which were not normal.

TABLE 80 Comparison of ALT and AST results in patients with HBV or HCV or without hepatitis

		HBV or HCV			Non-hepatitis		
Analyte	Upper limit	*n*	Mean	Median	*n*	Mean	Median
ALT	41	8	98.0	89.5	426	65.4	56.0
AST	43	6	94.5	69.5	254	64.5	53.5

Only patients whom the analyte is abnormal are included.

TABLE 81 Yield, sensitivity and PPVs of different detection strategies

Strategy for viral testing	No. of patients[a]	Hepatitis cases[a]	Viral tests	Cases detected	Sensitivity (%)	PPV, % (95% CI)
A. If repeat LFT panel is abnormal	1124	11	955	11	100	1.15 (0.64 to 2.05)
B. If ALT abnormal on primary test	1064	12	418	8	67	1.91 (0.97 to 3.73)
C. If ALT > twice ULN on primary test	1064	12	77	6	50	7.79 (3.62 to 15.98)
D. If patient born in a country of intermediate to high viral hepatitis prevalence	1208	13	170	11	85	6.47 (3.65 to 11.21)
E. If patient born in a country of intermediate to high viral hepatitis prevalence *and* ALT > twice ULN on primary test	1041	12	16	5	42	31.25 (14.16 to 55.60)
F. If patient born in a country of intermediate to high viral hepatitis prevalence, *or* ALT > twice ULN on primary test	1041	12	215	11	92	5.12 (2.88 to 8.93)
G. Test all cases	1236	13	1236	13	100	1.05 (0.62 to 1.79)

Testing patients for viral infection on the basis of country of origin is more sensitive and has much higher PPV.

a The sample of patients available to evaluate each strategy varies because of patterns of missing data, as follows: A requires a complete panel of follow-up LFTs, the missing data in the two cases that were not abnormal might have led to an exaggerated estimate of sensitivity. B and C both require an initial ALT test. D requires information on country of birth. E and F require an initial ALT, together with country of birth. All evaluations require results of viral tests for both hepatitis B and C.

TABLE 82 Cost estimates for resources used

Cost category	Resources (£)
GP consultation cost to check LFT results	12.86[a]
Receptionist to check patient in for appointment (2 minutes)	0.91[a]
Secretary time (1 minute)	0.33[a]
Phlebotomist time (5 minutes)	1.00[a]
Sample analysis: LFT	2.69[b]
Sample analysis: hepatitis B surface Ag and hepatitis C	25.42[b]

AG, antigen.

a Source: West Midlands Research Consortium (MidRec, February 2009).

b Source: University Hospitals Birmingham NHS Foundation Trust (2009).

Table 83. Strategy E (viral test if patient born in an intermediate-/high-risk country *and* ALT is greater than twice the ULN) provides the lowest cost per case detected. This strategy was therefore designated as the base case for calculation of ICERs. Strategy A, the intuitive and widely advocated practice of repeating LFTs, turns out to be the most expensive per case detected. It is dominated by strategy G, in which all patients undergo a viral test. Similarly, strategy B (viral test if the index ALT is abnormal) is dominated by strategy D (perform viral test if patient was born in an intermediate- or high-risk country). Strategy C (viral test if the ALT is greater than twice the ULN) can be eliminated by an extended dominance principle. If strategy C is preferred to strategy E, this can only be because the extra cases detected by strategy C are deemed worth the extra cost. However, strategy D finds yet more cases than strategy C at lower incremental cost. Therefore, either strategies E or D is preferable to strategy C. The cost-effectiveness of the remaining admissible strategies is shown in *Figure 19*. The dotted lines join strategies that cannot be eliminated by dominance principles. The absence of any explicit penalty for missing cases of viral hepatitis in this analysis implies that the costs of strategies E, D and F are underestimated with respect to strategy G. However, strategy F must be regarded as highly competitive with strategy G – it picks up almost as many cases and has very high efficiency in terms of cost per case detected.

The number of detected cases per patient is estimated as (sensitivity of strategy) × 1.05%, where the latter figure is the viral hepatitis prevalence observed in the complete sample of 1236 patients. The number used differs slightly from the actual number of cases detected per patient in *Table 83* because of variation in the prevalence of the condition across the samples in which each strategy was tested. The current approach achieves a more consistent comparison of strategies within our data set; for example, it ensures that the estimate of detected cases per patient for a strategy with 100% sensitivity will always be at least as great as that of any other strategy.

Discussion

Summary of main findings

The BALLETS study is the first GP-based study in which the entire cohort was comprehensively tested for additional diseases (such as viral hepatitis) after an abnormal LFT, using the full analyte panel and normal reference ranges. We have shown that an abnormal LFT alone does not select out a population in which the prevalence rate approaches a threshold that would justify viral screening. We have assessed the validity of the various strategies a GP could adopt, at least as far as viral hepatitis is concerned, when faced with an abnormal LFT of uncertain provenance. The intuitive response for a GP in such a situation would be to repeat the LFT, an approach advocated by current literature. This study shows that this may not be the optimal policy. This strategy is the most expensive, even more so than viral testing all patients, as the costs incurred include repeating the LFT as well as viral testing the majority. The study also shows that, if ALT is notably raised (greater than twice the ULN), then the probability of chronic viral hepatitis is high (nearly 8%), but sensitivity is low. The strategy of testing all people from intermediate- or high-prevalence countries is the second most efficient, in terms of cost per case detected, and detects almost twice as many cases as the most efficient strategy – testing for viral infection when two conditions (birth in an intermediate- or high-prevalence country and an ALT greater than twice the ULN) are satisfied. The relative financial disadvantages of the strategy of repeating the LFT would be even greater if patient cost were included, as the extra visit would have to be factored in.

Strengths and limitations of the study

The main strength lies in the unique nature of the BALLETS cohort, being the only prospective study in a primary-care setting that has looked at the consequences of an abnormal LFT from a full analyte panel. The main limitation of our study relates to the rather small number of cases

TABLE 83 Average cost per case detected for the six scenarios

Strategy	Cost per 100 patients (£)[a]	Cases detected per 100 patients	Cost per case detected (£)	Incremental cost (£) per 100 patients (with base = E)	Incremental cases detected per 100 patients (base = E)	ICER
A	5222	1.05	4965	5159	0.61	Dominated[b]
B	1592	0.70	2270	1530	0.26	Dominated
C	293	0.53	558	231	0.09	(2635)[c]
D	570	0.89	641	508	0.45	1124
E (base)[d]	62	0.44	142	0	0.00	Base
F	837	0.96	868	775	0.53	1473
G	4052	1.05	3853	3990	0.61	6503

a Cost of viral test = £40.52; cost of LFT panel = £17.79.
b Viral testing on the basis of country of origin dominates repeat LFT, being both cheaper and more cost-effective.
c Strategy C is eliminated by extended dominance.
d Patient born in a country of intermediate to high viral hepatitis prevalence and ALT level greater than twice the ULN on primary test is the least expensive strategy and was considered as base case.

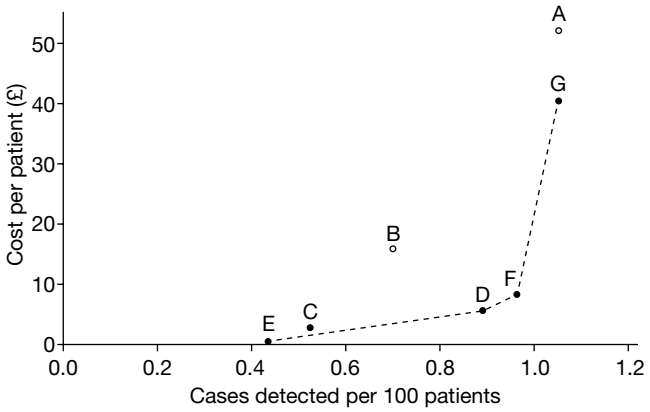


FIGURE 19 Cost per detected case for seven testing strategies.

of chronic viral hepatitis ($n = 13$) and hence wide confidence limits on the results. That said, the results are plausible, in the sense that they are consistent with the pathophysiology of hepatitis and in line with what was found in non-practice settings (see *Table 2*). They are available for meta-analysis with potential future studies.

We deliberately selected multicultural inner city populations in order to provide a sizeable subgroup of people from countries where chronic viral hepatitis is common, as a result of infection during infancy (hepatitis B)[113] and iatrogenic infection (hepatitis C). It turns out that 11 out of the 13 cases originated in medium- or high-prevalence countries. This has two implications. First, the prior probability is low (< 0.2%) and independent of ethnic group in an inner city UK population who do not originate from medium- or high-risk countries. Second, one of the most important questions a doctor can ask of a patient with abnormal LFTs is his or her country of origin – this is likely to apply irrespective of where the patient finally settles, as in most cases hepatitis is acquired soon after birth (hepatitis B) or as a result of iatrogenic infections in countries where has been a real risk (hepatitis C). Inner city populations were selected for this study in order to provide an 'enriched' population with a high proportion of immigrants. Given the biology of hepatitis B and C, there is little or no reason to suspect that an immigrant from a high-risk area will have a different risk according to where they settle in a low-risk country, whereas only 6% of the ethnic minority population had chronic viral hepatitis (see *Table 77*).

Our study considers only one disease type, chronic viral hepatitis, whereas GP decision-making must take into account other diseases, such as haemochromatosis, as well as other behavioural and social motivations for testing.[5,6] That said, our conclusion that repeating the LFT 'offers more than it delivers' may well apply to diseases such as PBC and haemochromatosis.

Lastly, we have presented an analysis for cost minimisation and incremental cost per case detected. This is not a full cost-effectiveness or decision analysis. Donnan *et al.*[24] did attempt a decision analysis. However, this decision analysis was intended to find the most cost-effective strategy in the short term and used a limited time horizon of 1 year. LFTs are often ordered to prevent poor outcome in the long term, with many serious liver diseases, viral hepatitis included, manifesting over decades. Anxiety resulting from a false-positive result was included in the model, whereas long-term health gains as a result of successful case finding and treatment were not captured.

Our results are considered in the context of published cost-effectiveness analyses for screening for viral hepatitis (i.e. studies that found screening was cost-effective in populations with high

prevalence rates, for example migrants) and attempt to produce a 'fast and frugal heuristic'[112] guide to practice.

Implications for practice: a fast and frugal heuristic

The intuitively appealing practice of repeating abnormal LFTs (strategy A) gets little support from our analysis. It is more expensive, both in absolute terms and in terms of cost per case detected, than all five alternative strategies (see *Table 82*), including that of simply testing everyone for viral infection.

The most important question a doctor can ask a patient with abnormal LFTs is his or her country of origin. This holds good whether the person settles in an area of high or low ethnic mix, as infections are acquired in infancy (hepatitis B) or as a result of substandard medical practices, such as needle sharing (hepatitis C). Once infected, people 'take their risk with them' – fewer people will need to be tested in a low-ethnic-mix area, but those from intermediate- or high-prevalence countries still need testing. The strategy of testing people from such countries promises good value for money. In this study, 11 of the 13 patients with chronic hepatitis originated in medium- or high-risk countries. Thus, the prevalence of chronic hepatitis viral infection (PPV) among people with an abnormal LFT who were born in a medium- or high-risk country was 6.5% (11/170; 95% CI 3.7% to 11.2%; see *Table 80*), whereas the prevalence among the home-born population (of all ethnic groups) was < 0.2% (2/1038; 95% CI 0.05% to 0.7%). Our findings support viral testing only in the former group, consistent with the threshold prevalence for both HBV and HCV, of approximately 3%, at which population screening becomes cost-effective.[106,114,115]

Four of the strategies – C, D, E and F – entail viral testing in a population in which the rate of hepatitis exceeds the 3% threshold for which testing has proven cost-effective in screening programmes (see *Table 80*). The cost-effective threshold is probably a little lower in a diagnostic population than in a screening population (costs of inviting people to attend are lower and cases detected might be a slightly higher risk), but no other strategy yields a population with hepatitis rate exceeding even 2%.

Strategy D (test immigrants from prevalent countries) has a better (lower) ICER than strategy C and detects twice as many cases as strategy E. However, strategy F, testing immigrants from prevalent countries *or* any people with a very high ALT, is our preferred strategy, being both sensitive and efficient. We therefore recommend the 'fast and frugal' heuristic described in *Figure 20*. This combines strategy F with normal judgement of clinical indications. For example, a patient who is an intravenous drug user, or who has recently returned from a trip abroad where they had an attack of hepatitis, would be tested notwithstanding the result of the LFTs. Otherwise we recommend testing all patients with an abnormal LFT who were born in a country of intermediate or high prevalence, and all patients for whom the ALT exceeds twice the limit of normal.

The probability of chronic viral hepatitis is low, even when the ALT exceeds this limit and the patient does not originate from a medium- or high-risk country (about 0.2%). Nevertheless, we advocate testing in these patients for the following reasons:

1. It is hard to ignore a level this high, and the wide confidence levels from our data suggest the need for flexibility.[116]
2. The progression for undetected chronic viral hepatitis is worse for patients with ALT levels that are greater than twice the ULN, and this level has been used as a threshold for treatment in guidelines.

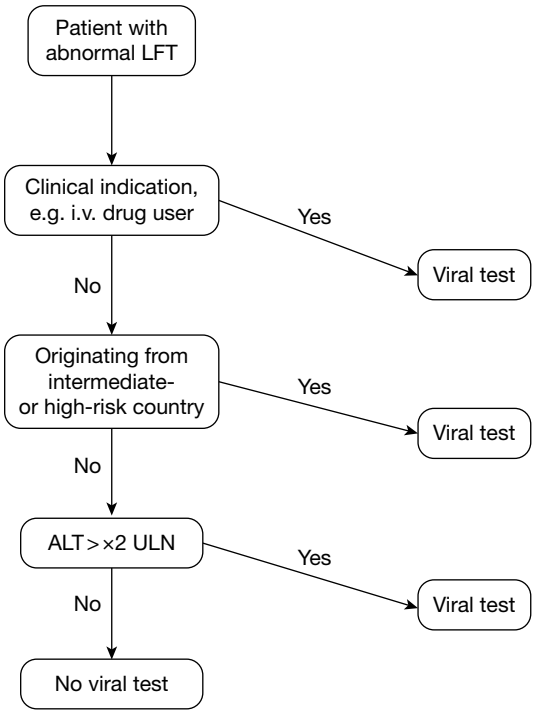


FIGURE 20 Fast and frugal heuristic decision tree. i.v., intravenous.

3. If chronic viral hepatitis is not present at this level, a more in-depth search for other causes of hepatocellular damage is indicated.

We draw the line on further viral testing after this algorithm has been followed, unless of course further clinical indicators emerge. The likelihood of a case of viral hepatitis being present following the exclusions in this algorithm is approximately 0.1% in our study. This is considerably below the UK population prevalence.

Conclusions

This analysis indicates that the strategy of repeating LFTs in asymptomatic patients, advocated by current guidelines, is less sensitive and far more expensive than viral testing those patients born in countries where viral hepatitis is prevalent. Despite few cases of viral hepatitis, the data on costs of the various strategies are strong and the results of prevalence rates within the cohort are consistent with other literature. The finding that a notably raised ALT level was also effective at identifying infected patients inspired the construction of a 'fast and frugal' heuristic that might aid GPs who are faced with abnormal LFTs in asymptomatic patients, with regards to viral hepatitis. Our proposal addresses the diagnostic problem by identifying a clear high-risk population originating in high-risk countries. The residual population who are not immigrants from such countries are at low risk. However, this should not over-ride clinical judgement. Its overall cost in other settings will depend on the relative proportions of patients in these risk strata, but our results suggest that the cost of automatic testing of high-risk individuals will be repaid in terms of additional cases detected.

Clearly, the situation might change as vaccination catches on in developing countries and needle hygiene improves. The key points to emerge are that:

1. It is more efficient to determine country of origin with a view to viral testing, than to simply repeat the LFT.
2. It is more cost-effective to test the whole LFT positive population for viral hepatitis than to repeat the LFT with a view to viral testing if it remains positive.

Chapter 7

Presence and severity of non-alcoholic fatty liver disease

Introduction

The incidence of liver disease is rising throughout the world, and liver disease now accounts for 1.5% of deaths in the UK.[117] In parallel with this, there has been a year-on-year rise in the number of LFTs carried out in primary care. Primary-care practitioners (PCPs) are thus commonly faced with the scenario of abnormal LFT results in patients in whom there are no clinical risks, signs or symptoms of liver disease. NAFLD is now recognised as the most common cause of hepatic dysfunction in the general population; however, this is yet to be confirmed in primary-care practice.[118,119] Furthermore, because of the indolent asymptomatic nature of NAFLD, identifying those with advanced disease in whom specific interventions may be required remains a clinical challenge in primary care.

The prevalence of NAFLD has risen markedly to 14–34% of the general population in Europe,[119,120] Asia[121] and America[122] in recent years. Although patients with simple NAFLD are believed to have benign disease, there is now clear evidence that those who have progressed to NASH and fibrosis are at a much higher risk of developing hepatocellular carcinoma (HCC), liver failure and death.[22,123] The majority of data describing the severity of liver fibrosis in NAFLD arise from selected populations in secondary referral centres.[18,19,21,22,29,124,125] In a large UK prospective study, Skelly *et al.*[19] demonstrated that 19% (23/120) of patients with biopsy-confirmed NASH had significant fibrosis after presenting to their secondary-care centre with unexplained abnormal LFTs.[19] This, and other such studies,[18,29] included patients in whom the decision to refer had been made on clinical grounds by PCPs/consultant colleagues and who were then rigorously screened in liver clinics for other disease aetiologies prior to proceeding to liver biopsy. These studies are therefore influenced by ascertainment bias and may overestimate the severity of NAFLD emerging from primary care.

With the alarming growth of obesity and type 2 diabetes, it is currently expected that the burden of NAFLD on primary care and liver services will continue to rise in the UK.[126] To date, no studies have determined the underlying disease severity of NAFLD in primary care. PCPs remain at the forefront of identifying the patients with advanced NAFLD who require further evaluation, closer surveillance for complications (and interventions where appropriate) and stricter lifestyle modifications. By investigating a large UK primary care sample of patients with incidental abnormal LFTs and absent clinical features of liver disease, this study is the first of its kind to determine the presence and disease severity of silent NAFLD in a primary-care setting.

Methods

Study population

This cross-sectional substudy utilises baseline data from patients enrolled in the BALLETS study from the eight primary-care practices within the Birmingham region only. Patients identified as having significant, positive liver disease aetiology were followed up in the specialist liver

outpatient clinic at the Queen Elizabeth University Hospital, Birmingham. Electronic liver clinic letters were reviewed for this substudy cohort until May 2010 to strengthen the reliability of the initial study finding of liver-specific disease.

Data definitions

The LFT blood profile consisted of ALT, AST, ALP, GGT, total bilirubin, globulin and albumin measurements. LFTs were classified as abnormal according to reference ranges in the local laboratories, which are compliant with quality control standards. All patients were screened for hereditary (Wilson's disease, A1AT deficiency and genetic haemochromatosis), infectious (HBV and HCV), autoimmune (autoimmune hepatitis, PBC and PSC) and drug-induced liver injury.

Body mass index was defined as weight in kilograms divided by the square of the height in metres (kg/m^2). Obesity was defined as BMI ≥ 30 kg/m^2. Alcohol intake was reported as standard units (1 unit = 10 g alcohol) of alcohol consumed on average per week in the 6 months prior to recruitment. The past medical history was also extensively reviewed to identify study participants who had a history of alcohol excess or alcohol-related health problems. Mild (female 1–7 units/week, male 1–11 units/week) and moderate (female 8–14 units/week, male 12–21 units/week) alcohol consumption were defined as drinking within the current UK health guidelines (female ≤ 14 units/week, male ≤ 21 units/week).[127] At-risk alcohol consumption was defined as exceeding these guidelines.

For the purposes of this substudy, type 2 diabetes was defined in patients with a documented history of the disease or a recorded drug history of anti-diabetic medication. Hypertension was defined as a past medical history of the disease or a current recorded drug history of two or more antihypertensive medications.

The diagnosis of NAFLD was based on the following criteria: (1) sonographic features of fatty liver on USS (increased hepatic parenchymal echotexture and vascular blurring); (2) a negative history of alcohol consumption exceeding current UK health guidelines; and (3) exclusion of liver disease of other aetiology including drug-induced, autoimmune, viral hepatitis, cholestatic, metabolic and genetic liver disease.

Non-alcoholic fatty liver disease fibrosis score

The NAFLD Fibrosis Score (NFS)[21] is a simple non-invasive scoring system designed to identify or exclude advanced fibrosis (classified as Kleiner stages F3 and F4[128]) in patients with an established diagnosis of NAFLD on imaging. The NFS was developed and validated by Angulo *et al.*[21] in over 700 liver patients with biopsy-proven NAFLD and is routinely used in liver clinics to select those at risk of disease progression and HCC. The NFS utilises a number of simple clinical (age, hyperglycaemia/diabetes, BMI) and laboratory (platelet count, albumin and AST/ALT ratio) independent predictors of advanced liver fibrosis. The low cut-off score (< -1.455) has a negative predictive value (NPV) of 88–93% and the high cut-off score ($> +0.676$) has a PPV of 79–90% for the presence of advanced fibrosis in NAFLD in secondary-care populations.[21,129] The NFS was calculated using the web-based electronic calculator (http://Nafldscore.com).

As the original BALLETS study protocol did not incorporate a platelet count, retrospective data collection of the electronic haematology laboratory archive at the University Hospital Birmingham enabled platelet counts within 6 months of patient enrolment to be recorded. To avoid false-positive or false-negative NFS, the scoring system was not applied to participants with a past medical history of platelet disorder or an active systemic inflammatory disease or being treated with myelosuppressive medications.

Statistical analysis

After exclusion of a positive blood/drug/alcohol aetiology screen, patients were diagnosed with NAFLD based on the presence of fatty liver on USS. Descriptive statistics were applied to characterise the whole study cohort and the identified NAFLD group. Continuous clinical variables are reported as medians and IQR. Categorical variables are reported as numbers and percentages.

Results

A total of 1118 primary-care patients were included. The majority (38%; 424/1118) of these resulted from routine chronic disease check-ups. In 4.5% (50/1118) of cases no reason was recorded. Liver aetiology screen and ultrasound were successfully completed in 98% (1095/1118) of patients at the study visit.

Causes of abnormal liver function tests

The cause of abnormal LFTs was identified in 54.9% (614/1118) of cases. Detailed testing for viral, genetic and autoimmune causes yielded 33 diagnoses (3.0%). NAFLD was identified as the commonest cause of abnormal LFTs, accounting for 26.4% of all cases, exceeding alcohol excess (25.3%). There were no reported cases of cirrhotic appearances or ascites on USS in the NAFLD cohort. Two or more abnormal LFT analytes were present in 40.7% of NAFLD subjects (120/295), with the remainder having a single analyte abnormality (59.3%; 175/295) on GP sampling. GGT was the most common LFT abnormality in the NAFLD cohort (76.5%; 199/260). The median time difference between GP ordering blood tests and the study visit was 30 days (IQR 18–51 days).

At-risk alcohol consumption was reported in 25.2% (282/1118). The majority of at-risk alcohol consumers were male (44.7%; 126/282) and drank a significantly greater amount of alcohol (units per week) than women [median 42 (IQR 30–56) units/week vs 29 (IQR 21–46) units/week; Mann–Whitney *U*-test, $p < 0.001$]. An echo-bright fatty liver was identified with USS in 44.7% (126/282) of subjects who consumed at-risk levels of alcohol. The majority of excess drinkers (87%; 110/126) had a BMI of $> 25\,kg/m^2$. Cirrhotic appearances (coarse texture with irregular outline) on USS were reported in two patients with at-risk alcohol consumption. The diagnosis of compensated alcohol-induced cirrhosis was confirmed by tertiary liver specialists. No cause for LFT abnormality was identified in the remainder of study subjects (45.1%; 504/1118). Of note, 17.5% (88/504) of the unexplained abnormal LFT cohort were obese with a concurrent diagnosis of either type 2 diabetes and/or hypertensive disease.

Disease severity in the cohort of patients with non-alcoholic fatty liver disease

To calculate the severity of NAFLD in this cohort we used the NFS. The score was calculated in 236 of the 295 patients who met the diagnostic criteria for NAFLD. The NFS was not calculated in the remaining 59 patients with NAFLD as a result of incomplete records of blood platelets ($n = 50$), BMI ($n = 5$) or AST/ALT ratio ($n = 4$). A high NFS ($> +0.676$) was found in 7.6% (18/236) of patients with NAFLD, suggesting the presence of underlying advanced liver fibrosis (stages F3/F4 on Kleiner classification).[130] Advanced fibrosis was predicted to be absent in the majority of NAFLD subjects with a low NFS (< -1.455), being calculated in 57.2% (135/236). The presence of advanced fibrosis, however, could not be confidently excluded in 35.2% (83/236) of the NAFLD patients who scored an indeterminate value with the NFS (-1.455 to $+0.676$).

Discussion

This large prospective primary-care study highlights that NAFLD accounts for over 25% of incidental abnormal LFTs in primary-care consultations in which the consulting GP's suspicion of underlying liver disease is low or absent. In contrast, a specific viral (HBV/HCV), genetic or autoimmune disease was identified on thorough study testing in only 3.0% of all study patients. Application of a simple, non-invasive scoring system suggests that undetected advanced liver fibrosis is present in 7.6% and absent in 57.2% of patients with NAFLD. Incidental abnormal LFTs were most commonly encountered during routine chronic disease reviews (38% cases), including diabetes, hypertension and cardiovascular disease. This study is the first of its kind to report the severity of NAFLD in patients with incidental abnormal LFTs in primary care.

Our study evaluated a primary care-based population with abnormal LFTs rather than a volunteer population from the general community. Nonetheless, the frequency of NAFLD (26%) identified in our study is within the wide range (14–34%) previously reported in general population studies carried out in Italy,[118] Spain,[120] Asia[121] and America.[122] The variation in reported frequencies may be influenced by ethnic diversity[122,130] and differences in study methodologies. These include variable alcohol thresholds that define NAFLD, lack of consistency in screening for other disease aetiologies and variation in risk stratification for liver disease at study enrolment. All the studies nevertheless confirm the strong association between NAFLD and components of the metabolic syndrome,[121,131] the prevalence of which has increased rapidly worldwide.[126] The high proportion of patients with diabetes (38.6%), obesity (60.3%) and hypertension (45.4%) in the NAFLD group in our study is in keeping with population-based studies.[118]

The suspected proportion of advanced fibrosis within our NAFLD cohort is 7.6%. Additionally, from experiences in hospital care[21,129,132] we predict that a subset of the 35.2% of patients with an indeterminate NFS may also have advanced fibrosis. There are currently no data on the severity of NAFLD in primary care. The most relevant studies that best reflect low-risk populations are restricted to biopsy findings in living related liver donors, among whom the prevalence of NASH (± fibrosis) ranges from 1.1% in Japan to 18.5% in the USA.[133] The latter figure is likely to be an overestimate due to the lack of detail on alcohol consumption and full liver aetiology screening in liver donors. Secondary/tertiary centre studies of variable size (range 118–733) and white predominance have reported that 11–27% of patients with biopsy-proven NAFLD and elevated aminotransferases have advanced (stages 3/4) fibrosis.[22,125,132,134,135] The higher rates of advanced fibrosis reported in these liver specialist centres are likely to be due to referral/sampling bias.

Our study has several unique strengths. First, this is the largest prospective cohort of primary-care patients with clinically unsuspected liver disease and incidental abnormal LFTs to be reported. Second, this is the first study to apply the non-invasive NFS to identify patients with advanced NAFLD fibrosis in primary care who are most in need of intensive lifestyle modifications and surveillance for liver-related complications (e.g. HCC detection). Third, the detailed assessment of the liver aetiology screen (alcohol/drug data, serology, genetics and USS imaging) undertaken and high completion rate (98%) mean that a cause for abnormal LFT was identified in the majority of cases (55%). Previous large-scale population-based retrospective analyses of abnormal LFTs have been limited by the absence of USS[119] and the lack of information on alcohol and measured anthropometry[2] to accurately describe the presence of NAFLD. The high rate of liver disease identification in our patient sample that PCPs perceived as a low-risk group may also be explained by the fact that GGT, which has the highest reported sensitivity for

liver disease, above other LFTs,[2] was the commonest LFT abnormality. The finding of an elevated GGT in more than 70% of the NAFLD group, compared with raised ALT in 51.0% and AST in 26.2%, has not been previously reported in adult patients with NAFLD. This finding has also been reported in children with NAFLD.[136]

One limitation of this study is that the application of the NFS was validated against liver biopsy in patients with NAFLD attending hospital,[21,129,132] and so it is possible that the severity of NAFLD may be overestimated in our primary-care cohort. However, our NAFLD cohort has very similar patient characteristics (white, obese, middle-aged, with abnormal LFT results) to those reported by Angulo *et al.*,[21] and in many countries the distinction between primary and secondary care is not as clear. For the purpose of our study, the NFS was chosen over other non-invasive systems[135,137,138] that detect advanced fibrosis as it is an easily applicable tool (web-based calculator) that has the best reported PPV in secondary care,[129] entails minimal extra cost to GPs (i.e. platelet sampling) and incorporates blood and clinical parameters that are routinely available in primary care. We were not able to validate the NFS against other non-invasive modalities,[137–139] as these had not been developed or sufficiently studied by the time our study had started. Moreover, there are issues about how to validate such modalities in primary care, as it is unlikely that liver biopsies would ever be performed in such a large sample of patients or in this setting (and would also be unethical).

Despite a thorough non-invasive aetiology screen and detailed alcohol history, 45% had unexplained abnormal LFTs in our cohort. However, as we targeted the more problematic patients in primary care, who have incidental abnormal LFTs in the absence of a clinical suspicion of underlying liver disease, this is not a surprise. Furthermore, unlike previous general population studies[118,119] that utilised only ALT, AST and/or GGT, our study recruited patients with a wider spectrum of LFT analytes to reflect common practice in primary care. It is therefore possible that some of the unexplained abnormal LFTs represent transient viral illness, Gilbert syndrome, under(self)-reported use of alcohol/over-the-counter medications or non-liver-related disease (i.e. bone, muscle).[119] Although USS is the most readily available imaging tool in primary care, the fact that 18% of the 'unexplained' group had co-existing obesity with diabetes and/or hypertension raises the possibility that reliance on ultrasound alone will miss a proportion of cases of NAFLD. The difficulty in detecting the presence of fatty liver with USS is well reported in the morbidly obese and when the degree of fat infiltration is < 33% of the hepatic content.[140] Furthermore, biopsy reports have shown that fat content is lost towards the more advanced stages of NAFLD, with the resultant fibrotic tissue being undetectable on USS.[140] The lack of markers of insulin sensitivity and lipid profile in the study meant that we were unable to non-invasively quantify hepatic fat,[141] and hence potentially determine the numbers of patients with undetected NAFLD on USS within the 'unexplained' group.

Our findings have important clinical and public health implications. This study raises awareness that NAFLD accounts for a significant proportion of incidental abnormal LFT results commonly encountered by PCPs, in the absence of a clinical suspicion of liver disease. We have identified a potential subset of patients with NAFLD with advanced fibrosis (7.6%) who require further follow-up and management in secondary care. We would advocate reassurance and lifestyle modifications to patients with a low NFS (57.2%). In the absence of validated scoring systems, at present patients with an indeterminate NFS require close surveillance in primary care with referral to secondary care as deemed appropriate by the PCP.

In conclusion, we provide novel information on the severity of NAFLD in a primary-care setting, as well as guidance on the triaging of such patients for further investigation and management.

Chapter 8

Interpretation and discussion

The BALLETS study did what it set out to do – recruit a cohort of patients with abnormal LFT results in primary care, characterise them comprehensively and follow them up for 2 years. It is a prognostic study but not a standard diagnostic study. It is possible to calculate 'sensitivity' ('true-positive rate'), as there are seven independent analytes and therefore many negative results for each analyte. However, it is important to remember that this will be an overestimate, as the results refer to a selected population of patients with at least one abnormality at index LFT (see *Chapter 4*, *Selection effects*). The degree to which this is representative of a wider population cannot be determined from within the study. However, the predictive value of abnormal LFTs can be confidently estimated from the study data.

Comparison of our results with previous literature

Research questions

Many of the BALLETS findings reinforce existing understanding concerning LFTs (or corroborate, in a primary-care setting, what is known in hospital care). Some findings reinforce ideas that many had long suspected but for which the evidence was scanty. A small but important number of findings are new or could be taken to contradict current understanding. We shall consider our findings with respect to the following questions:

1. Which LFTs (or combination of LFTs) predict what diseases (or disease classes)?
2. Which LFTs contribute most to diagnosis and which are more marginal?
3. What is the utility of the standard advice to repeat abnormal LFT results?
4. What is the overall contribution of LFTs to diagnosis in a primary-care setting?
5. What are the psychological sequelae of being told that LFT results are abnormal?
6. Why do doctors do so many LFTs and what are the different reasons for doing them?
7. What are the implications of 1–6, above, for ordering and interpreting LFTs?
8. How is fatty liver affected by change in weight for obese and non-obese patients?

Which tests predict which disorders?

There is a large literature on factors other than liver injury affecting LFTs, as summarised by Dufour *et al.*[14,15] Levels of both aminotransferase enzymes (ALT, AST) were higher in men, and levels of ALP were higher in women, in both the BALLETS study and Dufour's review. Both studies find an inverted U-shaped relationship between ALT and age, but Dufour did not find the age-related decline in albumin levels that we found in BALLETS. We found that globulin was higher in certain ethnic groups than others, but Dufour does not comment on this relationship. Laboratory reference ranges should be designed to take these factors into account, although, in practice, ethnicity is not considered. The positive association between ALT, AST and GGT with alcohol intake is well known but we have found an interesting negative association with ALP.

As far as the relationship between LFT levels and disease is concerned, our results are again in line with findings from Dufour *et al.*'s[15] systematic review both in the univariate analysis (see *Chapter 4*, *Patterns of abnormality and disease classes*) and in the various multivariate analyses (summarised in *Chapter 4*, *Discussion*).The major distinction typically drawn between diseases that damage hepatocytes directly (e.g. alcohol, viral infection) and those that cause intrahepatic

bile obstruction (namely PBC and PSC) was confirmed in the BALLETS cohort. As expected, the first group (disease category 1a) was associated with increased levels of aminotransferase enzymes (ALT and AST), whereas the second (disease category 1b) was characterised by high levels of ALP, the production and release of which from cell membranes is stimulated by cholestasis. ALP was also the analyte that was most strongly associated with type 2 diseases (that includes metastatic cancer); it was the only analyte for which abnormality was significantly associated statistically with this category.

Gamma-glutamyltransferase is by far the most frequently abnormal analyte (it has a strongly positive skewed distribution) and is raised across disease categories 1a and 1b. However, the high sensitivity of GGT is a function of its high overall positivity and it has lower predictive values for type 1a diseases than ALT and lower predictive values for 1b and category 2 diseases than ALP (see *Chapter 4*, *Diagnostic performance of alternative liver function test panels*, and *Figure 10*). Moreover, GGT was less sensitive than ALT for the most important 1a disease – viral hepatitis – and the cases of PBC that showed impending cirrhosis were all associated with abnormal ALP. Curiously, there is no tendency for higher GGT levels to achieve higher sensitivity as results become more extreme (see *Chapter 4*, *Patterns of abnormality and disease classes*, and *Figure 9*). We discuss the implications of the very low 'specificity' of GGT below.

Albumin levels are known to decline in decompensating cirrhosis and in many other late-stage diseases.[15] However, albumin measurement did not emerge as a useful test in our sample of low-risk patients with non-specific symptoms or attending for review of chronic diseases. It was the analyte that was least predictive of any other analyte being abnormal. It was not statistically associated with any disease category, nor did it emerge as an independent predictor for any disease class. We discuss this topic further below.

In summary, this study confirmed the well-known finding that aminotransferases are associated with 'hepatocellular' (1a) diseases, ALP with cholestatic (1b) diseases and systemic diseases involving the liver (type 2 diseases). It confirms that GGT is the most commonly abnormal analyte but its predictive value is relatively low. Albumin emerges as unhelpful for the diagnosis of liver disease in a non-high-risk population.

Which analytes are most useful and which are candidates for relegation?

Many laboratories use only five analytes and few use all eight deliberately included in BALLETS. It is reasonable to suppose that adding an analyte to a set that is already fit for its discriminatory purpose will add marginal diagnostic value at the cost of greater anxiety, patient inconvenience and health service expense. It is therefore important to use the BALLETS study to define the default set of analytes. If one wished to select only one analyte, GGT would be a very strong candidate, especially if sensitivity were regarded as a more important goal than specificity. However, as soon as two analytes can be selected, two prime candidates emerge – ALT and ALP (see *Chapter 4*, *Patterns of abnormality and disease classes*). The former ascertains most category 1a diseases and the latter most cases of 1b along with many in category 2. There was no correlation between ALP on the one hand and either ALT or AST on the other. In so far as they portend disease, they portend different diseases: ALT and AST are indicators of category 1a diseases (most often viral hepatitis, haemochromatosis), whereas ALP is a sign of intrahepatic biliary disease and, to a lesser extent perhaps, tumours in the liver. The univariate analysis showed that ALT is most strongly associated with 1a diseases and ALP with 1b and with category 2. The discriminant analysis confirms that these analytes are the strongest independent predictors of 1a and 1b diseases respectively [see *Chapter 4*, *Results (complete case analysis)* and *Chapter 4, Analysis of imputed data*]. The data show clearly that these two analytes, used together, are, by a considerable margin, the most discriminatory combination of tests in the extended LFT panel investigated. Bilirubin did not emerge as a strong discriminator. However, bilirubin is a marker of

acute liver disease and there were few such cases in our sample. We can see an argument to retain bilirubin in the panel so that it can be used in both acute and chronic liver disease.[15] Candidates for removal from the LFT panel (when used to diagnose disease of the liver – see below) are therefore GGT, AST and the protein fractions.

These will be considered in turn. Our conclusions regarding the first and third of these analytes are the opposite of those proposed in a companion study also funded by the HTA programme. This study was based on record linkage in Tayside, Scotland, and was conducted by Donnan *et al.*[24] Donnan *et al.*'s study[24] did not break down liver disease by type – the outcomes were 'liver disease diagnoses, 'liver disease mortality' and 'all-cause mortality'. As with BALLETS, the target population consisted of 'patients with no obvious liver disease'. Patients were excluded if they were referred to hospital following an abnormal test, had an abnormal LFT within the 6 months preceding the index test or had 'clinically obvious liver disease' recorded in the electronic patient records. No distinction was drawn between ALT or AST, which were combined into an aminotransferase category. Follow-up was for a median of 3.5 years. Comprehensive testing for type 1 disease was not carried out, nor was ultrasound examination, as this was a retrospective study. These points are mentioned because we think they go some way to explaining the diametrically opposite conclusions we have reached with respect to GGT and albumin.

- *Gamma-glutamyltransferase* Donnan *et al.*[24] recommend that GGT should be retained by laboratories when it is routine and that it should be incorporated in laboratories when this is not the case. This argument is based on the finding that that GGT was the most sensitive test for liver disease (sensitivity of 62%), whereas the next most sensitive test (transaminase) had a sensitivity of 36%. We contest their argument that GGT should be retained for this reason on two grounds:
  - The clinically important factor is the *added* value of a test *given* other information available to the clinician at the time. It is not overall sensitivity that should drive decision-making but *marginal* sensitivity given other information. There is no suggestion from any quarter that ALT and ALP should not be included in the LFT panel. So the salient question can be stated thus: given ALT and ALP, what are the marginal gains from adding GGT and what is the cost in terms of false-positives? ALT identified the majority of category 1a diseases, whereas ALP was raised in all but four of the 12 category 1b diseases. We provide an argument (see *Chapter 5*, *Summary of main findings*) that the most important result not to miss is chronic viral hepatitis. In this regard ALT was increased in 8 out of 12 cases in which the measurement was available from the index test; raised GGT identified only one additional case, and was seen in only 5 out of 11 cases for which the analyte was available, a smaller proportion than in the remainder of the study patients. It therefore appears that ALT and ALP in combination provide a sensitive strategy for detection of serious disease, and that the marginal yield from GGT is modest.
  - Any gains from GGT must be offset against the 'costs' contingent on a high false-positive rate. GGT was by far the most commonly abnormal analyte in BALLETS, with a high false-positive rate, and Donnan *et al.*[24] also found it to be the least specific of all the analytes they tested. As positive results create anxiety and risk a cascade of further tests, this undesirable feature of GGT must be offset against the small marginal gains in detection.
  - The expected relationship between true-positive rate and threshold (found for the other enzymes) is not present for GGT when the threshold is increased threefold (*Figure 9*). This casts further doubt on the value of GGT as a diagnostic test for liver disease in a general practice population at the existing threshold.

  GGT was less sensitive than ALT for identification of fatty liver, but the relationship to alcohol use was confirmed. We discuss the use of LFTs in relation to alcohol misuse below.

- *Albumin* Donnan *et al.*[24] note that albumin is a sensitive marker for *'all-cause mortality'*, while being more specific than GGT in this regard. The PPV for albumin, for death over 5 years, was 50% compared with 15% for GGT and 10% for ALP (the analyte with the lowest PPV). However, the marginal contribution to decision-making might have been less than these results might imply, as patients with low albumin were older (mean age of 69 years) than patients with normal albumin (mean age of 53 years) and they had considerably more comorbidities. The fact that these patients were frail and at high risk would therefore not have come as 'news' to the doctor in many cases. Donnan *et al.*[24] also found that albumin was the least sensitive marker for *liver* disease. We will discuss the role of LFTs as a general marker for health below, but the BALLETS results confirm that albumin is a very poor marker for type 1 and 2 liver diseases. It was associated with fatty liver but to a lesser extent than either BMI or ALT and does not emerge as a sensible screening test for people needing further investigation of this condition. In our opinion albumin could be dropped from the standard panel of tests for diagnosis of diseases involving the liver, although it might have utility as a marker for general health, not withstanding the point we make above. The issue of LFT markers for non-liver diseases is discussed below.
- *Aspartate aminotransferase* AST seems to be the 'poor relation' of ALT. It has little independent diagnostic precision, and misses more type 1a diseases generally, and more cases of viral hepatitis specifically, than ALT. It seems to be a very strong candidate for relegation and this is not highly controversial; most laboratories in the UK incorporate either one or the other of the aminotransferase enzymes in their standard panel. As with all analytes, it is in no way part of our argument that analytes excluded from the standard panel should not be available in special circumstances and such an argument applies to AST, which has a proven utility in distinguishing between alcohol damage and other type 1a liver diseases.[15]

In summary, we think that the standard testing for liver diseases should be built around just two analytes (ALT and ALP). We realise that this flies in the face of convention and may be too radical for immediate implementation. Bilirubin has a role when acute liver disease is suspected[15] and it may be reasonable to include it in the standard panel for this reason. We think that protein measurements should, if possible, be kept in reserve for situations in which there is general ill health, rather than specifically disease of the liver, which prompts investigation as discussed below. GGT and AST also have potential special roles in relation to alcohol, also as discussed below.

It could be argued that we have not demonstrated, in a formal sense, that the marginal gains from an extended panel are outweighed by the contingent losses, in terms of anxiety, additional visits and further tests. However, a full decisions analysis to 'prove' this point would be a massive and tedious undertaking, as there are so many inputs to the model (clinical features and patterns of LFT abnormality) and so many outputs (all category 1 and 2 diseases along with NAFLD). For each disease the added improvement in outcome would have to be modelled in the face of extremely uncertain transitional probabilities (viral hepatitis being something of an exception). Moreover, the drop-off in sensitivity gain and rise in specificity loss when extending the default panel (beyond the two analytes we recommend) is so stark that we think an intuitive response to the data is not only necessary but desirable. One claim that we cannot make is that there are significant savings to be made directly from reducing the number of analytes processed. This is because the tests are performed on the same analytical platform and directly from the same sample so that only reagent costs of about £0.17 could be saved. That said, volumes are high so the laboratory would save over £11,000 per year.

Repeating abnormal liver function tests versus testing for a specific disease forthwith

The standard advice when the LFT result is abnormal is to repeat it.[3,10,90,91,93] We have no quarrel with the conclusion of Donnan *et al.*'s decision analysis[24] – repeating the test is better than sending the patient to hospital for further investigation. However, this is not the only alternative; the result can be simply ignored (on the grounds, say, that it is only marginally abnormal) or a more specific follow-on test can be arranged as an alternative to simply repeating the LFT panel. The latter policy is precisely what our decision analysis indicates with respect to chronic viral hepatitis.

We question the standard default advice to repeat the LFT panel. LFT results remained abnormal on retesting in 84% of BALLETS patients (see *Chapter 4*, *Correlation analysis of liver function tests*) and in 66% of patients in the National Health and Nutrition Examination Survey.[142] The corollary is that, for the great majority of patients, the decision to take more specific action is merely deferred by the retesting policy. Furthermore, the low specificity of the repeat panel (15.6%) means that there is little opportunity for reassuring healthy patients (see *Chapter 4*, *Diagnostic performance of alternative liver function test panels*). Following an initial abnormal result, we suggest that either the patient should be reassured immediately or a further, more specific, test should be conducted in most cases. This point is reinforced by the decision analysis (see *Chapter 6*). Although conducted with respect to viral hepatitis, it may serve as a specimen for LFTs in general. An LFT work-up for category 1 genetic, autoimmune and viral diseases cost £67 in the study. Based on the costs shown in *Chapter 6* (see *Costs and cost analysis*), it is less expensive to the service to proceed directly to a LFT work-up than to repeat the LFT panel with a view to a full work-up if still positive, provided that the probability of a positive repeat LFT panel exceeds 67%. This result would be more extreme if the patient costs for an extra test were also factored in. The advice to proceed directly to a specific test might make yet more sense if the panel used to exclude/diagnose diseases of the liver could be restricted to just two or three analytes (ALT, ALP and bilirubin), as suggested above. In that case, a broad default framework for investigation of an abnormal LFT result could be described, as shown in *Figure 21*.

As ALP is also associated with category 2 diseases, a more complete algorithm for this scenario is posited in *Figure 22*. Likewise, excluding type 1a diseases when ALT is raised requires further elucidation, and a default guideline for this scenario is offered in *Figure 23* – this builds on the fast and frugal heuristic developed on the basis of the decision analysis in *Chapter 6*.

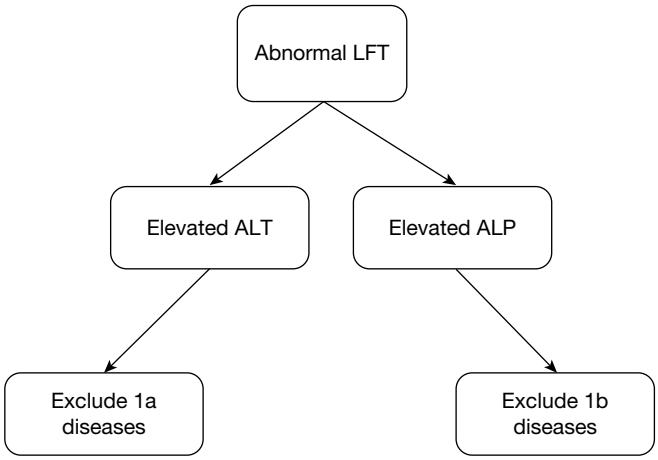


FIGURE 21 Broad default framework for investigation of an abnormal LFT result.

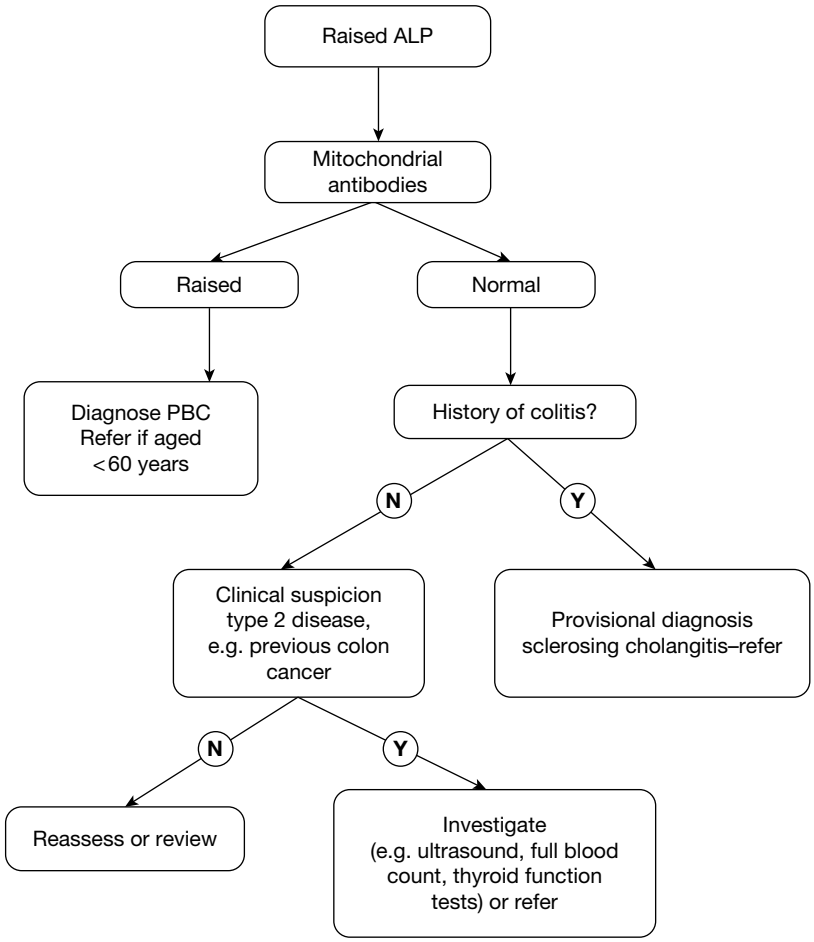


FIGURE 22 Diagnostic algorithm for patients with raised ALP. N, no; Y, yes.

This scheme is indicative rather than prescriptive, must be tailored to individual circumstances and does not trump clinical judgement. It could be adapted to cope with a more extended range of analytes, if these are retained in the panel used for the diagnosis of liver disease. The algorithms deal with the default scenario and do not preclude repeating the LFT. This would certainly be the appropriate course of action if, for example, it transpired that the patient had taken vigorous exercise shortly before testing, as it is known that this can cause up to a threefold elevation in ALT.[14] Likewise, repeating the test would make sense in cases of suspected drug reaction or transient viral infection (such as infectious mononucleosis).

In summary, the standard advice to routinely repeat a LFT test following an abnormal result can be strongly questioned on the basis of BALLETS results. In 84% of cases this will simply defer the decision. The decision analysis (in which viral hepatitis is used as a specimen – see *Chapter 6*) suggests that moving directly to the test that would have been carried out had the LFT remained abnormal is the most cost-effective option from both a health service and societal perspective. Such a policy will also avoid cases of false reassurance that a recidivist test can provide (see *Chapter 4*, *Persistence of liver function test abnormality from index to first follow-up*).

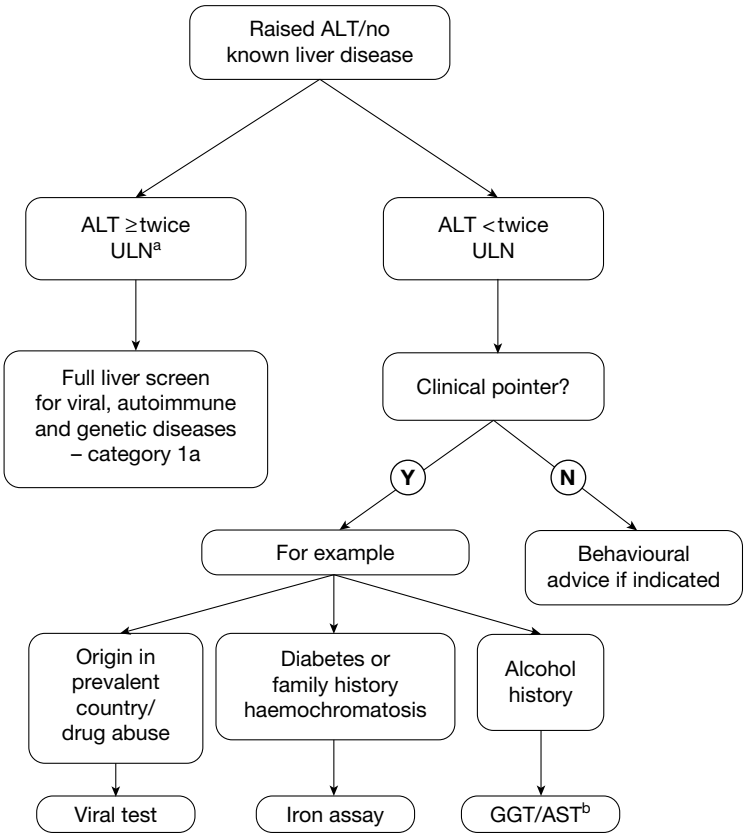


FIGURE 23 Diagnostic algorithm for default management of patients with raised ALT. a, The risk of disease increases with ALT level (see *Chapter 6*) and, in the case of hepatitis B viral infection, an ALT twice the ULN is an indication for antiviral treatment. b, If patient suspected of being in denial or to motivate patient to reduce consumption – see text. N, no; Y, yes.

The overall contribution of liver function tests to diagnosis in a primary-care setting

Donnan *et al.*[24] state that one of the four main findings of their record linkage study is that 'liver disease is not common in those with abnormal LFTs'. These authors would have underestimated disease frequency because they followed patients for a median of 3.5 years, whereas many diseases, such as viral hepatitis, PBC and haemochromatosis, emerge over decades. BALLETS, on the other hand, must overestimate the frequency of clinically relevant disease, as many of the diagnoses were pathophysiological entities with lead times that would exceed life expectancy in many cases. This applies, in particular, to haemochromatosis and PBC but even, to some extent, to viral hepatitis. Despite this potential 'over-ascertainment' in BALLETS and underestimation in the Donnan *et al.* study,[24] both studies reach the same conclusion: the incidence of serious liver disease in patients with abnormal LFT results in primary care is low. Only 59 patients in nearly 1300 had a *potential* serious disease affecting the liver, including metastatic cancer – a predictive value of < 5%. Only 45 (< 3.5%) had one of the diseases that are captured by a comprehensive liver screen for viral, autoimmune and genetic diseases. Viral hepatitis (B and C) is the most common of the serious liver conditions for which a highly effective medical treatment is available. Only 1% of people had this condition – similar to the UK population prevalence of 0.7% (and the Birmingham prevalence in antenatal clinics) (see *Chapter 5*, *Discussion*). The risk of cirrhosis is also under 1%, at 0.7% (see *Chapter 4*, *Disease categories*). The results therefore confirm the prevailing opinion that LFTs are carried out in circumstances in which serious preventable

disease is unlikely to be detected. Category 1 diseases were rare in the BALLETS cohort, and even when detected the majority seemed to be following a benign course – increased availability and use of testing in developed countries identifies diseases that would not have presented but for the testing. Many of these category 1 diseases were subclinical and likely to remain so. Only two cases of PBC were likely to produce clinical sequelae – the remainder were in elderly patients with no evidence of incipient cirrhosis. Serious cases of PBC usually present in mid-life. Likewise, the majority of the 10 haemochromatosis cases are unlikely to come to harm and none was started on chelating treatment as a result of the diagnosis. Fewer than 1% of patients were started on a course of treatment that was likely to extend their life as a result of an abnormal LFT. These were patients with viral hepatitis, and all but two of them could have been detected not by LFT testing, but by simply testing all people from intermediate- or high-risk countries for the virus. In short, the BALLETS study confirms what many have long suspected: LFTs deliver much less than they promise, at least as far as detecting disease in the nominated organ is concerned.

Group 2 includes certain conditions that might not have been in the GP's mindset when ordered; for example, Paget's disease of the bone, infectious diseases such as glandular fever and leptospirosis, and thyroid disease. Analyses excluding these cases from Group 2 have been published in a *BMJ Open* paper (Lilford RJ, Bentham LM, Armstrong MJ, Neuberger J, Girling AJ. What is the best strategy for investigating abnormal liver function tests in primary care? Implications from a prospective study. *BMJ Open* 2013; in press.) The conclusions are essentially unchanged after these exclusions.

The 59 cases of putative liver disease do not include patients who have fatty liver and/or exceed safe alcohol limits. ALT is the most sensitive analyte for the identification of fatty liver. The high incidence of fatty liver (38%) is consistent with findings reported in the literature.[17,18,23,26–30] The finding nevertheless reinforces the high prevalence of ultrasonically detected liver fat. The ultrasound diagnosis of fatty liver is not an 'exact science', but the finding that obesity and ALT results increased with the presence and degree of fatty liver provided evidence of criterion validity for the interpretation of the ultrasonic images. That said, the value of making the ultrasound diagnosis of fatty liver is questionable. The probability of identifying incipient cases of fat-induced cirrhosis in this way must be small; the incidence of fatty liver is very much higher than that of fat-induced cirrhosis (38% vs < 0.5%). Moreover, the main argument for detecting fatty liver by ultrasound must rest on the expectation that a positive result will prompt behaviour change and motivate the identified person to eat and/or drink less. If this is true, then there may also be a risk that a negative result will provide false reassurance and hence an insouciant attitude to an unhealthy lifestyle. There are good arguments for behaviour change in people with high calorie or alcohol intake irrespective of LFT and liver ultrasound results. The use of LFTs or ultrasound as a method to encourage healthy lifestyles is an unproven intervention and arguably both should be used with circumspection pending further study – a topic to which we return below.

In summary, we conclude with Donnan *et al.*[24] that the proportion of cases in which LFTs lead to the diagnosis of a previously unsuspected liver disease, for which an evidence-based treatment is indicated, is very low – < 1% in BALLETS. Most of these cases relate to viral hepatitis and the value of LFTs would be further attenuated in a population in which all patients from high-risk countries had been screened. A more circumspect/discriminatory attitude to LFTs is recommended.

Psychological effects

The psychological consequences of being informed about a positive LFT were measured in nested studies in both the Donnan *et al.*[24] and BALLETS projects. Both studies formed a measurable adverse effect on anxiety (state anxiety in BALLETS and both state and trait anxiety in Donnan *et al.*[24]). The BALLETS study (see *Chapter 5*, *Psychology 1*: *effects of positive tests*) also found

that disease-specific worry was markedly increased and self-assessed health slightly decreased after testing when compared with results 2 years later. An ultrasound diagnosis of fatty liver was associated with slightly worse scores on all three dimensions but this result did not reach statistical significance. The qualitative study (see *Chapter 5*, *Psychology 2*: *results on behaviour*) produced results that were consistent with the dissipation of anxiety and disease-specific worry seen in the quantitative study. The hypothesis that ultrasound detection of fat in the liver would be a powerful motivating factor in behaviour change did not gain support from the quantitative data. This is consistent with poor recall of findings and their significance found in the qualitative study as discussed below. The effect of finding abnormal LFTs and/or ultrasound on unhealthy behaviour, and whether or not repeating these tests can nudge people towards healthy lifestyles, remain unanswered questions. On the other hand, there can be little question that the tests are anxiety provoking in the short term and this must be included in the deficit column in the balance sheet of potential benefits and harms of testing. Demonstrable anxiety following an abnormal result forms part of the argument for removing GGT from the default list of analytes in the LFT panel (it adds little information at the margin and increases the probability of a positive test with all that entails) (see *Chapter 4*, *Diagnostic performance of alternative liver function test panels*). The idea of restricting the panel for suspected liver disease to just two (or possibly three) analytes in the first instance is merely an extension of this argument.

In summary, the documented negative effects of an abnormal LFT mean that false-positives must be taken seriously. It is an argument against simply advocating including analytes with the highest sensitivity with no regards to predictive value. The idea that, in some circumstances, the disability of anxiety can be offset against positive effects on behaviour only applies to circumstances where there is an indication for behaviour change and even then the net benefit of using LFT results for this purpose is unproven. Taken in the round these considerations reinforce arguments for:

1. being more circumspect about doing LFTs in the first place
2. excluding analytes (e.g. GGT) that add only small marginal sensitivity to the LFT panel at the expense of a big increase in false-positives.

Doctors' motivations in ordering large numbers of liver function tests

The low probability of making a timely diagnosis of an important disease needing treatment suggests that LFTs have limited value in people with vague symptoms or as part of the monitoring of non-liver diseases. The time has come to re-examine the widespread use of these tests. Four motivations for testing can be discerned from the sociological substudy:

1. to diagnose a serious disease affecting the liver (i.e. a category 1 or 2 disease)
2. to test for a non-liver disease
3. to promote/reinforce behaviour change and/or to elucidate a suspicion that the patient may underestimate alcohol consumption
4. to reassure the patient and/or signal that the complaint was being taken seriously, to 'buy time' or as an 'insurance policy' against potential complaints.

We appreciate that these may be overlapping motivations and that doctors are not necessarily conscious of explicit motivations in practice; as one respondent said, 'Ordering an LFT may have become a type of "tick-box" response.' However, that might be part of the problem – failure to think through the purpose of testing can be blamed for the current situation where large numbers of people present with abnormal LFTs, the meaning of which is unclear; uncertain provenance generates low prognostic significance.

In summary, it makes sense to consider tailoring the LFT to the reason for testing rather than adopting a 'one-size-fits-all' approach.

Building a new testing paradigm

In this section, we combine the conclusions from the six preceding sections. First, we take as our starting point a hypothetical scenario in which LFTs have only recently been discovered and have not yet come into routine use. Second, we consider the rational response under such a scenario, given both biological knowledge (summarised by Dufour *et al.*[14,15]) and the results of the BALLETS study. Third, we attempt an answer to this question in terms of the four broad motives for doing LFTs described above:

1. Concern over disease affecting the liver. We recommend that recourse to LFTs be more circumspect and that when carried out a panel of just two analytes (ALT and ALP) is used, with bilirubin added where an acute liver event is suspected. Alternatively, for simplicity, a three-panel test may be used. Such a panel would be suitable for monitoring liver-toxic drugs, in cases of suspected acute poisoning (e.g. paracetamol, mushroom) and in patients with infectious diseases, such as hepatitis A and leptospirosis.[15]
2. Concern over general (non-liver) conditions. The standard LFT panel is not fit for purpose. It will produce a crop of positive results that do not point clearly to the next step, and simply repeating the LFT is unlikely to advance the diagnosis. We recommend a dropdown list of tests from which the clinician can select according to circumstances. Pending further research we suggest the following candidates for inclusion on such a list: thyroid function tests (TFTs), the full blood count (FBC), an inflammatory marker (such as C-reactive protein) and albumin. Which tests are most propitious in these circumstances is unclear. The FBC is useful in patients with non-acute abdominal complaints in general practice,[143] whereas TFTs, FBC and an inflammatory marker are advocated for chronic fatigue.[144] TFTs and the FBC are 'tractable' in the sense that the required actions contingent on a positive result are reasonable well defined.
3. As a means to promote behaviour change or to confirm suspicion of alcohol misuse. If the patient is suspected to be in denial about alcohol misuse, then AST and GGT would be a sensible choice of test. Using LFTs to reinforce behaviour change may be a reasonable option pending further evaluation but it gets no support from this study (see *Chapter 7*, *The effect of changes in body mass index and alcohol intake on fatty liver*). GGT is an obvious choice in the case of alcohol misuse. It is important to be aware that behaviour change is still warranted, even if the test is normal, and that false reassurance should be avoided.
4. Meeting the patient's perceived need for a blood test. Here, it seems that the very last thing required is an 'open-ended test' – that is to say a test that has a high positive rate but whose meaning is obscure. We would recommend selection of tests that cover frequently missed diagnoses and where further action is well defined by a positive test result. TFTs and the FBC meet these requirements. Again, the clinician may be aided by a dropdown list or check list.

In summary, we think the LFT panel has outlived its usefulness and should be replaced by a more nuanced approach. Above, we outline such an approach based on the results of the BALLETS study and review of the literature.

Non-alcoholic fatty liver disease

We found a high prevalence of fatty liver disease as described above. BMI and ALT were the strongest predictors of this condition. Seventy per cent of patients who had fatty liver on their initial scan were found to have a fatty liver 2 years later, whereas only 14% patients did not have a fatty liver at the outset had this finding at 2-year follow-up. We found an interesting J-shaped curve relating alcohol intake to probability of fatty liver in men. This has been discovered before in secondary care,[145,146] and moderate alcohol consumption was associated with lower risk of metabolic syndrome in a community-based study.[142]

BALLETS found a significant ($p = 0.032$) association between change in BMI over 2 years and change in liver fat (see *Chapter 7*, *The effect of changes in body mass index and alcohol intake on fatty liver*). The improvement in liver fat was found to be sensitive to relatively small reductions in BMI. These findings should be encouraging to patients with metabolic syndrome/fatty liver, although the extent to which they may translate into clinical outcomes is conjectural. One important question relates to the effect of having a fatty liver on behaviour. The qualitative study did not suggest that the finding of a fatty liver was a sustained, powerful motivating factor (see *Chapter 5*, *Psychology 2*: *effects of results on behaviour*). Examination of weight change in BALLETS showed that, although patients with fatty livers lost some mass, on average, over the period of the study, those without fatty liver experienced a small gain. The difference does not reach statistical significance (see *Chapter 4*, *The effect of changes in body mass index and alcohol intake on fatty liver*). The possibility of a small but worthwhile effect remains an intriguing possibility.

Lastly, although most cases of fatty liver do not progress to cirrhosis, finding out why a small proportion do so is a priority, given the rising incidence of obesity and metabolic syndrome.

In summary, we provide yet further evidence that a small amount of alcohol is associated with healthy outcomes but proving a cause and effect relationship remains elusive. There is an intriguing hint that informing a person that they have a fatty liver will prompt weight loss but this needs confirmation, as does the hypothesis that this putative benefit is not vitiated by a countervailing effect in people who test negative. Using LFTs/ultrasound to promote behaviour change is unproven and clinicians should be cautious in doing so.

Strengths and limitations

Birmingham and Lambeth Liver Evaluation Testing Strategies is a unique study comprising patients presenting in primary care who have been investigated by means of a 'full' panel of LFTs who have then been comprehensively screened for liver disease and followed up for 2 years.

An important strength of the BALLETS study is that it investigates not only the psychological sequelae of testing, but also the reasons for ordering LFTs in the first place. This has enabled the authors to analyse the implications of the results not just in some general context, but in the context of what turned out to be very different motivations for testing. We have also conducted an analysis of cost per case detected for the most important liver disease – viral hepatitis – which resulted in a provocative finding that contradicts the current guidelines. Lastly, we have created a cohort of patients, many with fatty liver, for further follow-up. We had planned at one stage to develop a consensus statement regarding the practical implications of the BALLETS study. However, the results suggested that radical changes in practice were indicated and the groups of primary-care clinicians to whom the data have been presented were reluctant to immediately accept the radical corollaries that we believe flow from the data. Rather than sublimate our views in contemporary consensus we decided on a completely different philosophy – an interpretation that can be debated over time and gradually assimilated into practice as required. In economic terms, we felt that a form of supplier generated demand was needed, that this would take time, and that a 'consensus development' approach was likely to be excessively reactionary in the circumstances.

Not all eligible patients were recruited to the BALLETS study, but the substudy of patients who were not recruited in the integral pilot (see *Chapter 3*, *Integral pilot*) provides reassurance that the population studied is broadly representative of the target group.

Many patients diagnosed with specific conditions (especially PBC) represented pathophysiological entities rather than patients destined to suffer clinical effects. However, by providing details of each case we were able to distinguish between cases where clinical consequences were more or less likely to occur in a transparent way.

The number of patients in certain disease categories was small, limiting the statistical power of some of the analyses. However, this low incidence of disease emerges as an important finding in its own right. The populations chosen were deliberately rather high risk with a skew towards the inner city, rather than wealthier suburban or rural locations (see *Chapter 6*, *Strengths and limitations of the study*). The low predictive values observed in the BALLETS study would be, in all likelihood, lower still in more middle-class neighbourhoods.

Implications for research

1. A pilot study of a 'customised' approach to test ordering should be considered. The clinical value of different tests when patients have vague symptoms, such as tiredness or upper abdominal pain, should be evaluated. Likewise, the need to carry out more blood tests when patients are on treatment for chronic disease, such as hypertension, is unclear. There is a mismatch between the frequency with which blood tests are used to monitor chronic diseases and investigate symptoms, on the one hand, and scientific exploration of this subject, on the other. We have made the point (see *Chapter 5*) that LFTs are sometimes (or even usually) ordered not because liver disease is suspected, but because of a less focused suspicion of disease interacting with a perceived societal expectation to perform a test of some type. We propose a research project aimed at better defining situations in which different blood tests are done for vague symptoms (such as tiredness). On the basis of the BALLETS results, we hypothesise that tests such as TFTs and the FBC, for which the meaning of a positive result is rather clear cut, will offer more than LFTs, for which the predictive meaning of a positive result is so uncertain. In addition to studying the yield from various tests we propose an evaluation of a specific dropdown menu comprised the FBC, TFTs and the various individual components of the LFTs.
2. The BALLETS cohort should be followed up over time to find out whether or not it is possible to identify the minority of cases of fatty liver that are likely to progress to cirrhosis and to evaluate the fibrosis score in a primary-care setting. We are seeking permission to obtain death certificates for the cohort and also to obtain funding to follow up patients with special reference to the ultrasound diagnosis of NAFLD.
3. A controlled study of the net effects of using serial LFTs (including liver ultrasound) as part of a package to reduce unhealthy behaviours should be seriously considered, especially in light of the rising incidence of obesity. The hypothesis that using test results to promote behaviour change will do more good than harm is unproven. The simpler solution – tackling unhealthy behaviours directly and irrespective of test results – may be more effective all round.

Chapter 9

Conclusions

Our conclusion, derived by integrating statistical findings with motivations for testing, is that we have reached the beginning of the end of the pervasive LFT panel. A more rational response is to blend biological knowledge and the statistical results from the BALLETS study, to create testing heuristics appropriate to the very different purposes for testing. The ubiquitous and frequently used LFT panel has been the subject of prolonged scepticism. What has been lacking hitherto was a sufficiently large empirical study of patients similar to the bulk of those encountered in clinical practice and an intellectual framework that started with the objective of testing. We offer our study as an example of the insights that can be achieved when biological knowledge, quantitative field work and qualitative interviews are combined. The conclusions are radical, only because existing practice has evolved as a type of mneme, with little empirical justification.

Acknowledgments

Contributions of authors

All authors read and approved the final manuscript. In addition:

Richard Lilford (Professor of Clinical Epidemiology, Director of Birmingham Clinical Research Academy) participated in the design of the study and acquired funding, supervised the research group and drafted the final report manuscript.

Louise Bentham (BALLETS Study Co-ordinator) participated in study management, data collection in Birmingham, the design of the sociological and qualitative substudies, and contributed to writing final report.

Alan Girling (Senior Research Fellow) participated in the design of the study, performed the statistical analysis and wrote the first draft of the results section.

Ian Litchfield (Senior Research Fellow) conducted and co-authored qualitative and sociological substudies.

Robert Lancashire (Computer Officer) participated in design of analysis plan and provided critical revision of final report.

David Armstrong (Professor of Medicine and Sociology) participated in the design of the study and provided critical revision of final report.

Roger Jones (Professor of General Practice) participated in the design of the study, was lead GP at the Lambeth site and provided critical revision of final report.

Theresa Marteau (Professor of Psychology) participated in the design of the study, conducted psychology substudy and provided critical revision of final report.

James Neuberger (Honorary Consultant Physician) participated in the design of the study and provided critical revision of final report.

Paramjit Gill (Clinical Reader in Primary Care Research) participated in the design of the study and provided critical revision of final report.

Bob Cramb (Consultant Chemical Pathologist) participated in the design of the study, wrote a subsection of the report and provided critical revision of final report.

Simon Olliff (Consultant Radiologist) participated in the design of the study, conducted and drafted quality assurance of ultrasound subsection and provided critical revision of the final report.

David Arnold (Medical Student) conducted main study literature search and co-authored the decision analysis substudy.

Khalid Khan (Professor of Women's Health and Clinical Epidemiology) participated in the design of the study, the analysis plan and the final analysis, and provided critical revision of final report.

Matthew Armstrong (Clinical Research Fellow) conducted and co-authored the fibrosis substudy and provided critical revision of the final report.

Diarmaid Houlihan (Clinical Research Fellow) conducted and co-authored the fibrosis substudy.

Philip Newsome (Senior Lecturer in Hepatology) conducted and co-authored the fibrosis substudy.

Peter Chilton (Research Associate) co-authored the decision analysis substudy and provided critical revision of the final report.

Karel Moons (Professor of Clinical Epidemiology) participated in the design of the study, the analysis plan and the final analysis, and provided critical revision of the final report.

Doug Altman (Professor of Biostatistics) participated in the design of the study, the analysis plan and the final analysis, and provided critical revision of the final report.

All authors read and approved the final manuscript.

Contribution of others

The authors are grateful to the following people:

Study collaborators Rosalind Raine (Professor of Health Care Evaluation) and Ramasamyiyer Swaminathan (Consultant Chemical Pathologist).

Sonographers Clair Powell, David Hill, Sharon Breedon, Vicky Conway, Judith Phillips and John Leddy.

Research associates Susan Hoult Robinson, Emily Benson, Ruth Collins and Divya Chadha.

Research nurses and clinical researchers Carol Wheeler, Anne-Marie Goddard and Ugochi Nwulu.

Lead GPs Dr Ewan Hamnett, Dr Bill Strange, Dr David Taylor, Dr Lakh Jhass, Dr Kirstie Blackford, Dr Adam Fraser and Dr Peter Clarke.

Practice managers and administrative support Chris Jenkins, Joyce Marriott, Sheila Jones, Fay Knight, Janet Barrass, Julie Brown, Julie Walker, Karen Leslie, Angela Styring, Vicky Chambers, Suzy Hill, Maria DelVecchio, Melita Shirley and Catharine Hill.

Laboratory staff Dr Peter Lewis, Timothy Plant, Barbara Swann, Dr David Vallance, Dr Mourad Labib, Jean Shaw and Gillian Muirhead.

Database and IT support Phill Martin and Ciaron Hoye.

Publications

1. Arnold D, Bentham L, Jacob R, Girling A, Lilford J. Should patients with abnormal liver function tests in primary care be tested for chronic viral hepatitis: cost minimum analysis based on a comprehensively tested cohort. *BMC Fam Pract* 2011;**12**:9.
2. Lilford RJ, Bentham LM, Armstrong MJ, Neuberger J, Girling AJ. What is the best strategy for investigating abnormal liver function tests in primary care? Implications from a prospective study. *BMJ Open* 2013; in press.

References

1. Office for National Statistics (ONS). *Health service quarterly*; Winter 2008:59–60.
2. McLernon DJ, Donnan PT, Ryder S, Roderick P, Sullivan FM, Rosenberg W, *et al.* Health outcomes following liver function testing in primary care: a retrospective cohort study. *Fam Pract* 2009;**26**:251–9.
3. Sherwood P, Lyburn I, Brown S, Ryder S. How are abnormal results for liver function tests dealt with in primary care? Audit of yield and impact. *BMJ* 2001;**322**:276–8.
4. Milne S, Sheeran P, Orbell S. Prediction and intervention in health related behaviour: a meta-analytic review of protection motivation theory. *J Appl Soc Psychol* 2000;**30**:106–43.
5. Petrie KJ, Cameron LD, Ellis CJ, Buick D, Weinman J. Changing illness perceptions after myocardial infarction: an early intervention randomized controlled trial. *Psychosom Med* 2002;**64**:580–6.
6. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. *Health Educ Behav* 2000;**27**:591–615.
7. Chatwin T. Diagnosing liver disease in asymptomatic patients. *JAAPA* 2001;**14**:39–47.
8. Moseley RH. Evaluation of abnormal liver function tests. *Med Clin North Am* 1996;**80**:887–906.
9. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, *et al.* Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998;**2**(3).
10. Pratt DS, Kaplan MM. Evaluation of abnormal liver-enzyme results in asymptomatic patients. *N Engl J Med* 2000;**342**:1266–71.
11. Rochling FA. Evaluation of abnormal liver tests. *Clin Cornerstone* 2001;**3**:1–12.
12. Younossi ZM. Evaluating asymptomatic patients with mildly elevated liver enzymes. *Cleve Clin J Med* 1998;**65**:150–8.
13. Green RM, Flamm S. AGA technical review on the evaluation of liver chemistry tests. *Gastroenterology* 2002;**123**:1367–84.
14. Dufour DR, Lott JA, Nolte FS, Gretch DR, Koff RS, Seeff LB. Diagnosis and monitoring of hepatic injury. I. Performance characteristics of laboratory tests. *Clin Chem* 2000;**46**:2027–49.
15. Dufour DR, Lott JA, Nolte FS, Gretch DR, Koff RS, Seeff LB. Diagnosis and monitoring of hepatic injury. II. Recommendations for use of laboratory tests in screening, diagnosis, and monitoring. *Clin Chem* 2000;**46**:2050–68.
16. Bonacini M. The value of hepatic iron index in cirrhosis. *Gastroenterology* 1997;**113**:1052–3.
17. Whitehead MW, Hawkes ND, Hainsworth I, Kingham JG. A prospective study of the causes of notably raised aspartate aminotransferase of liver origin. *Gut* 1999;**45**:129–33.
18. Daniel S, Ben Menachem T, Vasudevan G, Ma CK, Blumenkehl M. Prospective evaluation of unexplained chronic liver transaminase abnormalities in asymptomatic and symptomatic patients. *Am J Gastroenterol* 1999;**94**:3010–14.
19. Skelly MM, James PD, Ryder SD. Findings on liver biopsy to investigate abnormal liver function tests in the absence of diagnostic serology. *J Hepatol* 2001;**35**:195–9.

20. Sorbi D, McGill DB, Thistle JL, Therneau TM, Henry J, Lindor KD. An assessment of the role of liver biopsies in asymptomatic patients with chronic liver test abnormalities. *Am J Gastroenterol* 2000;**95**:3206–10.

21. Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, *et al.* The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* 2007;**45**:846–54.

22. Ekstedt M, Franzen LE, Mathiesen UL, Thorelius L, Holmqvist M, Bodemar G, *et al.* Long-term follow-up of patients with NAFLD and elevated liver enzymes. *Hepatology* 2006;**44**:865–73.

23. Kim HC, Nam CM, Jee SH, Han KH, Oh DK, Suh I. Normal serum aminotransferase concentration and risk of mortality from liver diseases: prospective cohort study. *BMJ* 2004;**328**:983.

24. Donnan PT, McLernon D, Dillon JF, Ryder S, Roderick P, Sullivan F, *et al.* Development of a decision support tool for primary care management of patients with abnormal liver function tests without clinically apparent liver disease: a record-linkage population cohort study and decision analysis (ALFIE). *Health Technol Assess* 2009:**13**(25).

25. Dore GJ, Freeman AJ, Law M, Kaldor JM. Is severe liver disease a common outcome for people with chronic hepatitis C? *J Gastroenterol Hepatol* 2002;**17**:423–30.

26. Bellentani S, Tiribelli C, Saccoccio G, Sodde M, Fratti N, De MC, *et al.* Prevalence of chronic liver disease in the general population of northern Italy: the Dionysos Study. *Hepatology* 1994;**20**:1442–9.

27. Pendino GM, Mariano A, Surace P, Caserta CA, Fiorillo MT, Amante A, *et al.* Prevalence and etiology of altered liver tests: a population-based survey in a Mediterranean town. *Hepatology* 2005;**41**:1151–9.

28. Yano E, Tagawa K, Yamaoka K, Mori M. Test validity of periodic liver function tests in a population of Japanese male bank employees. *J Clin Epidemiol* 2001;**54**:945–51.

29. Hultcrantz R, Glaumann H, Lindberg G, Nilsson LH. Liver investigation in 149 asymptomatic patients with moderately elevated activities of serum aminotransferases. *Scand J Gastroenterol* 1986;**21**:109–13.

30. Mathiesen UL, Franzen LE, Fryden A, Foberg U, Bodemar G. The clinical significance of slightly to moderately increased liver transaminase values in asymptomatic patients. *Scand J Gastroenterol* 1999;**34**:85–91.

31. Kim KM, Kim YJ, Lee KH, Paek D. Clinical characteristics of factory workers with asymptomatic liver function test abnormalities found on serial health examination. *Korean J Hepatol* 2005;**11**:144–56.

32. Health Protection Agency. *Hepatitis C information*. Health Protection Agency, 2003 (cited 22 February 2011). URL: www.hpa.org.uk/infections/topics_az/hepatitis_c/phlsgen_info.htm

33. Health Protection Agency. *Hepatitis B information*. Health Protection Agency, 2003 (cited 22 February 2011). URL: www.hpa.org.uk/infections/topics_az/hepatitis_b/gen_info.htm

34. Worwood M. Haemochromatosis. *Clin Lab Haematol* 1998;**20**:65–75.

35. Metcalf JV, Bhopal RS, Gray J, Howel D, James OF. Incidence and prevalence of primary biliary cirrhosis in the city of Newcastle upon Tyne, England. *Int J Epidemiol* 1997;**26**:830–6.

36. AutoImmune Hepatitis Support Group. *Autoimmune Hepatitis*. AutoImmune Hepatitis Support Group website, 2009 (cited 3 April 2009). URL: www.autoimmunehepatitis.co.uk

37. Olivarez L, Caggana M, Pass KA, Ferguson P, Brewer GJ. Estimate of the frequency of Wilson's disease in the US Caucasian population: a mutation analysis approach. *Ann Hum Genet* 2001;**65**:459–63.

38. de Serres FJ. Worldwide racial and ethnic distribution of alpha1-antitrypsin deficiency: summary of an analysis of published genetic epidemiologic surveys. *Chest* 2002;**122**:1818–29.

39. Office for National Statistics, 2001 Census: Digitised Boundary Data (England and Wales). ESRC/JISC Census Programme, Census Geography Data Unit (UKBORDERS), EDINA (University of Edinburgh).

40. Targher G, Montagnana M, Salvagno G, Moghetti P, Zoppini G, Muggeo M, *et al.* Association between serum TSH, free T4 and serum liver enzyme activities in a large cohort of unselected outpatients *Clin Endocrinol* 2008;**68**:481–4.

41. van Burren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;**18**:681–94.

42. Royston P, Carlin JB, White IR. Multiple imputation of missing values:new features for mim. *Stata Journal* 2009;**9**:252–64.

43. Heshka JT, Palleschi C, Howley H, Wilson B, Wells PS. A systematic review of perceived risks, psychological and behavioral impacts of genetic testing. *Genet Med* 2008;**10**:19–32.

44. Shaw C, Abrams K, Marteau TM. Psychological impact of predicting individuals' risks of illness: a systematic review. *Soc Sci Med* 1999;**49**:1571–98.

45. Lerman C, Trock B, Rimer BK, Jepson C, Brody D, Boyce A. Psychological side effects of breast cancer screening. *Health Psychol* 1991;**10**:259–67.

46. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *Br J Clin Psychol* 1992;**31**:301–6.

47. Ware JE, Jr, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) project. *J Clin Epidemiol* 1998;**51**:903–12.

48. Grattagliano I, Portincasa P, Palmieri VO, Palasciano G. Managing nonalcoholic fatty liver disease: recommendations for family physicians. *Can Fam Physician* 2007;**53**:857–63.

49. Fuertes JN, Mislowack A, Bennett J, Paul L, Gilbert TC, Fontan G, *et al.* The physician–patient working alliance. *Patient Educ Couns* 2007;**66**:29–36.

50. Bordin E. The generalizability of the psychanalytic concept of the working-alliance. *Psychother Theory Res Pract Train* 1979;**16**:252–60.

51. Smith VA, DeVellis BM, Kalet A, Roberts JC, DeVellis RF. Encouraging patient adherence: primary care physicians' use of verbal compliance-gaining strategies in medical inteviews. *Patient Educ Couns* 2005;**57**:62–76.

52. St GA, Bauman A, Johnston A, Farrell G, Chey T, George J. Effect of a lifestyle intervention in patients with abnormal liver enzymes and metabolic risk factors. *J Gastroenterol Hepatol* 2009;**24**:399–407.

53. Smith JA. *Semistructured interviewing and qualitative analysis. Rethinking methods in psychology*. London: Sage; 1995. pp. 9–26.

54. Smith JA. Beyond the divide between cognition and discourse: using interpretative phenomenological analysis in health psychology. *Psychol Health* 1996;**11**:261–71.

55. Smith JA. Identity development during the transition to motherhood: an interpretative phenomenological analysis. *J Reprod Infant Psyc* 1999;**17**:281–300.

56. Smith JA, Osborn M. Interpretative phenomenological analysis. In Smith JA, editor. *Qualitative psychology*. London: Sage; 2003.

57. Willig C. *Qualitative research in psychology: a practical guide to theory and method*. Buckingham: Open University Press; 2001.

58. Conrad P. *The experience of illness: recent and new directions. Research in the sociology of health care*. In Roth JA, Conrad P, editors. London: JAI Press; 1987. pp. 1–31.

59. Elliott R, Fischer CT, Rennie DL. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *Br J Clin Psychol* 1999;**38**:215–29.

60. Parker I. Criteria for qualitative research in psychology. *Qual Res Psychol* 2004;**1**:95–106.

61. Ley P. Memory for medical information. *Br J Soc Clin Psychol* 1979;**18**:245–55.

62. Kessels RP. Patients' memory for medical information. *J R Soc Med* 2003;**96**:219–22.

63. Fortun P, West J, Chalkley L, Shonde A, Hawkey C. Recall of informed consent information by healthy volunteers in clinical trials. *QJM* 2008;**101**:625–9.

64. Edwards SJ, Lilford RJ, Braunholtz DA, Jackson JC, Hewison J, Thornton J. Ethical issues in the design and conduct of randomised controlled trials. *Health Technol Assess* 1998;**2**(15).

65. Lawton J, Fox A, Fox C, Kinmonth AL. Participating in the United Kingdom Prospective Diabetes Study (UKPDS): a qualitative study of patients' experiences. *Br J Gen Pract* 2003;**53**:394–8.

66. Raftery J, Bryant J, Powell J, Kerr C, Hawker S. Payment to healthcare professionals for patient recruitment to trials: systematic review and qualitative study. *Health Technol Assess* 2008;**12**(10).

67. Lovato LC, Hill K, Hertert S, Hunninghake DB, Probstfield JL. Recruitment for controlled clinical trials: literature summary and annotated bibliography. *Control Clin Trials* 1997;**18**:328–52.

68. Cegala DJ, Marinelli T, Post D. The effects of patient communication skills training on compliance. *Arch Fam Med* 2000;**9**:57–64.

69. Beastall GH. The impact of the new General Medical Services contract: national evidence. *Bull R Coll Pathol* 2004;**128**:24.

70. Galloway MJ, Osgerby JC. An audit of the indications for the reporting of blood films: results from the National Pathology Benchmarking Study. *J Clin Pathol* 2006;**59**:479–81.

71. Croal BD. Laboratory demand management. *Bull R Coll Pathol.* No. 139, July 2007.

72. Wolf JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med* 2002;**162**:2269–76.

73. Xiong T, McEvoy K, Morton DG, Halligan S, Lilford RJ. Resources and costs associated with incidental extracolonic findings from CT colonogaphy: a study in a symptomatic population. *Br J Radiol* 2006;**79**:948–61.

74. Axt-Adam P, van der Wouden JC, van der Does E. Influencing behaviour of physicians ordering laboratory tests: a literature study. *Med Care* 1993;**31**:784–94.

75. Guthrie B. Why do general practioners take blood? *Eur J Gen Pract* 2001;**7**:142.

76. van der Weijden T, van Bokhoven MA, Dinant GJ, van Hasselt CM, Grol RP. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract* 2002;**52**:974–80.

77. Strauss A, Corbin J. *Basics of qualitative research, grounded theory procedures*. London: Sage; 1990.

78. Kok G, de Vries H, Mudde NA, Strecher VJ. Planned health education and the role of self-efficacy: Dutch research. *Health Edu Res* 1991;**6**:231–8.

79. Jefferis BJ, Power C, Manor O. Adolescent drinking level and adult binge drinking in a national birth cohort. *Addiction* 2005;**100**:543–9.

80. Dantzer C, Wardle J, Fuller R, Pampalone SZ, Steptoe A. International study of heavy drinking: attitudes and sociodemographic factors in university students. *J Am Coll Health* 2006;**55**:83–9.

81. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* 1974;**185**:1124–31.

82. Wileman L, May C, Chew-Graham CA. Medically unexplained symptoms and the problem of power in the primary care consultation: a qualitative study. *Fam Pract* 2002;**19**:178–82.

83. Department of Health (DoH). *Creating a patient-led NHS: delivering the NHS Improvement Plan*. London: DoH; 2005.

84. van der Weijden T, van VM, Dinant GJ, van Hasselt CM, Grol R. Unexplained complaints in general practice: prevalence, patients' expectations, and professionals' test-ordering behavior. *Med Decis Making* 2003;**23**:226–31.

85. van Bokhoven MA, Pleunis-van Empel MC, Koch H, Grol RP, Dinant GJ, van der Weijden T. Why do patients want to have their blood tested? A qualitative study of patient expectations in general practice. *BMC Fam Pract* 2006;**7**:75.

86. Summerton N. Positive and negative factors in defensive medicine: a questionnaire study of general practioners. *BMJ* 1995;**310**:27–9.

87. Bailey J, Jennings A, Parapia L. Change of pathology request forms can reduce unwanted requests and tests. *J Clin Pathol* 2005;**58**:853–5.

88. Zaat JO, van Eijk JT, Bonte HA. Laboratory test form design influences test ordering by general practitioners in The Netherlands. *Med Care* 1992;**30**:189–98.

89. Cramb R, Chemical Pathologist. University Hospitals Birmingham laboratories. 2005. Personal communication.

90. Cobbold JF, Anstee QM, Thomas HC. Investigating mildly abnormal serum aminotransferase values. *BMJ* 2010;**341**:c4039.

91. Giannini EG, Testa R, Savarino V. Liver enzyme alteration: a guide for clinicians. *CMAJ* 2005;**172**:367–79.

92. Limdi JK, Hyde GM. Evaluation of abnormal liver function tests. *Postgrad Med J* 2003;**79**:307–12.

93. Theal RM, Scott K. Evaluating asymptomatic patients with abnormal liver function test results. *Am Fam Physician* 1996;**53**:2111–19.

94. Butt AA. Hepatitis C virus infection: the new global epidemic. *Expert Rev Anti Infect Ther* 2005;**3**:241–9.

95. Lavanchy D. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. *J Viral Hepat* 2004;**11**:97–107.

96. Trepo C, Pradat P. Hepatitis C virus infection in Western Europe. *J Hepatol* 1999;**31**:80–3.

97. Weber B, Taelem-Brule N, Berger A, Simon F, Geudin M, Ritter J. Evaluation of a new automated assay for hepatitis B surface antigen (HBsAg) detection VIDAS HBsAg Ultra. *J Virol Methods* 2006;**135**:109–17.

98. Kanwal F, Farid M, Martin P, Chen G, Gralnek IM, Dulai GS, *et al.* Treatment alternatives for hepatitis B cirrhosis: a cost-effectiveness analysis. *Am J Gastroenterol* 2006;**101**:2076–89.

99. Thompson-Coon J, 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).

100. Weinstein M, Fineberg H. *Clinical decision analysis*. Philadelphia, PA: WB Saunders; 1980.

101. Lu SN, Wang JH, Kuo YK, Kuo HL, Chen TM, Tung HD, *et al.* Predicting the prevalence of antibody to hepatitis C virus (HCV) in a community by the prevalence of elevated levels of alanine aminotransferase: a method to identify areas endemic for HCV. *Am J Trop Med Hyg* 2002;**67**:145–50.

102. Wang CS, Wang ST, Chou P. Using the prevalence of an elevated serum alanine aminotransferase level for identifying communities with a high prevalence of hepatitis C virus infection. *Arch Intern Med* 2001;**161**:392–4.

103. Jamali R, Khonsari M, Merat S, Khoshnia M, Jafari E, Bahram KA, *et al.* Persistent alanine aminotransferase elevation among the general Iranian population: prevalence and causes. *World J Gastroenterol* 2008;**14**:2867–71.

104. D'Souza R, Foster GR. Diagnosis and treatment of chronic hepatitis B. *J R Soc Med* 2004;**97**:318–21.

105. Persico M, Persico E, Suozzo R, Conte S, De SM, Coppola L, *et al.* Natural history of hepatitis C virus carriers with persistently normal aminotransferase levels. *Gastroenterology* 2000;**118**:760–4.

106. Veldhuijzen IK, Toy M, Hahne SJ, De Wit GA, Schalm SW, de Man RA, *et al.* Screening and early treatment of migrants for chronic hepatitis B virus infection is cost-effective. *Gastroenterology* 2010;**138**:522–30.

107. World Health Organization (WHO), Geneva, Dept of Vaccines and Biologicals. *Introduction of hepatitis B vaccine into childhood immunization services: management guidelines, including information for health workers and parents*. Geneva: WHO; 2001.

108. World Health Organization (WHO), Geneva, Marketing and Dissemination. *International Travel and Health. Situation as on 1 January 2005*. Geneva: WHO Press; 2005.

109. World Health Organization (WHO), Geneva, Marketing and Dissemination. *International travel and health. Situation as on 1 January 2010*. Geneva: WHO; 2010.

110. Hutton DW, Tan D, So SK, Brandeau ML. Cost-effectiveness of screening and vaccinating Asian and Pacific Islander adults for hepatitis B. *Ann Intern Med* 2007;**147**:460–9.

111. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998;**17**:857–72.

112. Todd PM, Gigerenzer G. Precis of Simple heuristics that make us smart. *Behav Brain Sci* 2000;**23**:727–41.

113. Poland GA, Jacobson RM. Clinical practice: prevention of hepatitis B with the hepatitis B vaccine. *N Engl J Med* 2004;**351**:2832–8.

114. Castelnuovo E, Thompson-Coon J, Pitt M, Cramp M, Siebert U, Price A, *et al.* The cost-effectiveness of testing for hepatitis C in former injecting drug users. *Health Technol Assess* 2006;**10**(32).

115. Stein K, Dalziel K, Walker A, Jenkins B, Round A, Royle P. Screening for hepatitis C in genito-urinary medicine clinics: a cost utility analysis. *J Hepatol* 2003;**39**:814–25.

116. Nakamura J, Terajima K, Aoyagi Y, Akazawa K. Cost-effectiveness of the national screening program for hepatitis C virus in the general population and the high-risk groups. *Tohoku J Exp Med* 2008;**215**:33–42.

117. Office for National Statistics (ONS). *Mortality statistics: deaths registered in 2008*. London: ONS; 2008.

118. Bedogni G, Miglioli L, Masutti F, Tiribelli C, Marchesini G, Bellentani S. Prevalence of and risk factors for nonalcoholic fatty liver disease: the Dionysos nutrition and liver study. *Hepatology* 2005;**42**:44–52.

119. Clark JM, Brancati FL, Diehl AM. The prevalence and etiology of elevated aminotransferase levels in the United States. *Am J Gastroenterol* 2003;**98**:960–7.

120. Caballeria L, Auladell MA, Toran P, Miranda D, Aznar J, Pera G, *et al.* Prevalence and factors associated with the presence of non alcoholic fatty liver disease in an apparently healthy adult population in primary care units. *BMC Gastroenterol* 2007;**7**:41.

121. Hamaguchi M, Kojima T, Takeda N, Nakagawa T, Taniguchi H, Fujii K, *et al.* The metabolic syndrome as a predictor of nonalcoholic fatty liver disease. *Ann Intern Med* 2005;**143**:722–8.

122. 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**:1387–95.

123. Bugianesi E, Leone N, Vanni E, Marchesini G, Brunello F, Carucci P, *et al.* Expanding the natural history of nonalcoholic steatohepatitis: from cryptogenic cirrhosis to hepatocellular carcinoma. *Gastroenterology* 2002;**123**:134–40.

124. Neuschwander-Tetri BA, Clark JM, Bass NM, Van Natta ML, Unalp-Arida A, Tonascia J, *et al.* Clinical, laboratory and histological associations in adults with nonalcoholic fatty liver disease. *Hepatology* 2010;**52**:913–24.

125. Soderberg C, Stal P, Askling J, Glaumann H, Lindberg G, Marmur J, *et al.* Decreased survival of subjects with elevated liver function tests during a 28-year follow-up. *Hepatology* 2010;**51**:595–602.

126. Ahmed MH, Abu EO, Byrne CD. Non-Alcoholic Fatty Liver Disease (NAFLD): new challenge for general practitioners and important burden for health authorities? *Prim Care Diabetes* 2010;**4**:129–37.

127. British Medical Association (BMA). *Alcohol: guidelines on sensible drinking.* London: BMA; 1995.

128. 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**:1313–21.

129. McPherson S, Stewart SF, Henderson E, Burt AD, Day CP. Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. *Gut* 2010;**59**:1265–9.

130. Petersen KF, Dufour S, Feng J, Befroy D, Dziura J, Dalla MC, *et al.* Increased prevalence of insulin resistance and nonalcoholic fatty liver disease in Asian-Indian men. *Proc Natl Acad Sci USA* 2006;**103**:18,273–7.

131. Marchesini G, Brizi M, Bianchi G, Tomassetti S, Bugianesi E, Lenzi M, *et al.* Nonalcoholic fatty liver disease: a feature of the metabolic syndrome. *Diabetes* 2001;**50**:1844–50.

132. Wong VW, Wong GL, Chim AM, Tse AM, Tsang SW, Hui AY, *et al.* Validation of the NAFLD fibrosis score in a Chinese population with low prevalence of advanced fibrosis. *Am J Gastroenterol* 2008;**103**:1682–8.

133. Tran TT, Changsri C, Shackleton CR, Poordad FF, Nissen NN, Colquhoun S, *et al.* Living donor liver transplantation: histological abnormalities found on liver biopsies of apparently healthy potential donors. *J Gastroenterol Hepatol* 2006;**21**:381–3.

134. Angulo P. Long-term mortality in nonalcoholic fatty liver disease: is liver histology of any prognostic significance? *Hepatology* 2010;**51**:373–5.

135. Harrison SA, Oliver D, Arnold HL, Gogia S, Neuschwander-Tetri BA. Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. *Gut* 2008;**57**:1441–7.

136. Feldstein AE, Charatcharoenwitthaya P, Treeprasertsuk S, Benson JT, Enders FB, Angulo P. The natural history of non-alcoholic fatty liver disease in children: a follow-up study for up to 20 years. *Gut* 2009;**58**:1538–44.

137. Parkes J, Roderick P, Harris S, Day C, Mutimer D, Collier J, *et al.* Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. *Gut* 2010;**59**:1245–51.

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

139. Castera L, Forns X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. *J Hepatol* 2008;**48**:835–47.

140. Saadeh S, Younossi ZM, Remer EM, Gramlich T, Ong JP, Hurley M, *et al.* The utility of radiological imaging in nonalcoholic fatty liver disease. *Gastroenterology* 2002;**123**:745–50.

141. Kotronen A, Peltonen M, Hakkarainen A, Sevastianova K, Bergholm R, Johansson LM, *et al.* Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. *Gastroenterology* 2009;**137**:865–8, 72.

142. Freiberg MS, Cabral HJ, Heeren TC, Vasan RS, Curtis ER. Alcohol consumption and the prevalence of the Metabolic Syndrome in the US: a cross-sectional analysis of data from the Third National Health and Nutrition Examination Survey. *Diabetes Care* 2004;**27**:2954–9.

143. Muris JW, Starmans R, Fijten GH, Crebolder HF, Schouten HJ, Knottnerus JA. Non-acute abdominal complaints in general practice: diagnostic value of signs and symptoms. *Br J Gen Pract* 1995;**45**:313–16.

144. Harrison M. Pathology testing in the tired patient: a rational approach. *Aust Fam Physician* 2008;**37**:908–10.

145. Dunn W, Xu R, Schwimmer JB. Modest wine drinking and decreased prevalence of suspected nonalcoholic fatty liver disease. *Hepatology* 2008;**47**:1947–54.

146. Moriya A, Iwasaki Y, Ohguchi S, Kayashima E, Mitsumune T, Taniguchi H, *et al.* Alcohol consumption appears to protect against non-alcoholic fatty liver disease. *Aliment Pharmacol Ther* 2011;**33**:378–88.

Appendix 1

BALLETS study protocol, version 13.0, March 2010

Contents

5.4 A Qualitative Investigation into Patients' Experience of Taking Part in the BALLETS Study and the Impact of the Finding of a Fatty Liver on Ultrasound

5.4.1 Methodology

5.4.1.1 Participants

5.4.1.2 Process

5.4.2 Data analysis

6.0 CONTRIBUTION: EXPERTISE

7.0 CONSUMERS

8.0 JUSTIFICATION OF SUPPORT REQUIRED

9.0 REFERENCE LIST

10.0 APPENDIX

10.1 BALLETS Database

10.1.a Sonographers report

10.1.b Patient details

10.1.c Lab results

10.2 BALLETS Study Patient Process

10.2.a Hall Green Health

10.2.b Lordswood House & Shenley Green Surgery

10.2.c Greenridge Surgery

10.2.d Yardley Wood Health Centre

10.2.e Lambeth

10.2.f Woodland Road Surgery, Wand and Cofton Medical Centres

10.3 Patient Information Sheet

10.3.a Hall Green Health

10.3.b Lordswood House Medical Practice

10.3.c Greenridge Surgery

10.3.d Yardley Wood Health Centre

10.3.e Woodland Road Surgery

10.3.f Cofton Medical Centre

10.3.g Wand Medical Centre

10.3.h Lambeth

10.4 Consent Form

10.4.a Birmingham participants

10.4.b Lambeth participants

10.5 Lordswood House

10.5.a Receptionist script

10.5.b Letter to patients

10.6 Lambeth Letter to Patients

10.7 Ultrasound Examination Data Collection Form

10.8 Consolidated Report Form

10.9 Hepatology Referral Guidelines

10.10 Testing Strategies

- 10.10.aPatient interview information sheet
- 10.10.bPatient interview consent form
- 10.10.cPatient views questionnaire 1
- 10.10.dFollow-up psychology and lifestyle questionnaire

10.11

- 10.11.aData collection form
- 10.11.bLong-term follow-up protocol for patient
- 10.11.cExtra patient information sheet
- 10.11.dExtra consent form

10.12

- 10.12.aGP consent form
- 10.12.bInformation sheet for General Practitioners
- 10.12.cReferences
- 10.12.dList of practices

10.13

- 10.13.aPrimary biliary cirrhosis letter to GPs
- 10.13.bWilson's disease letter to GPs
- 10.13.cHaemochromatosis letter to GPs
- 10.13.dAutoimmune hepatitis letter to GPs

10.14

- 10.14.aPatient information sheet
- 10.14.bPatient consent form
- 10.14.cBALLETS qualitative sub-study: semi-structured patient interview questions
- 10.14.dSonographer information sheet
- 10.14.eSonographer consent form
- 10.14.fBALLETS qualitative sub-study: semi-structured sonographer interview questions

BALLETS - Birmingham And Lambeth Liver Evaluation Testing Strategies

1. Project Reference and Title: 03/38 Investigations following abnormal liver function tests.

2. Planned Investigation

2.1 The Problem. Liver Function Tests (LFTs) are a good example of inexpensive tests (modern auto-analysers process large batches of samples using inexpensive reagents) that are frequently ordered as a 'test of exclusion' in patients with non-specific symptoms, such as tiredness or upper abdominal discomfort. The tests are also non-specific in the sense that none of the four to seven analytes included in the LFT panel points directly to a specific diagnosis and many are not even specific to the liver. A doctor may order a laboratory test because a patient has features of a particular disease, for example, the gradual onset of jaundice in the user of injectable substances points to hepatitis C. The prior risk of hepatitis in such a person would be high; many positives would be true positives. Frequently, however, LFTs are ordered without such a tractable link between symptoms and a specific diagnosis. For example, when patients have vague symptoms for which no specific organ system appears likely to be responsible, or where serious, but still nebulous diagnosis such as disseminated malignancy is part of the differential diagnosis. Such tests are often offered as a type of insurance policy. The prior risk of disease is low and other things being equal, the proportion of positives which are false positives will be high. LFTs are interpreted by reference to population norms, rather than explicit calculus of the relative benefits and harms of false positive and negative diagnoses. Many patients have a positive test, but it is not clear what proportion of these are true positives, especially when the test is only mildly abnormal.

It is clear that very large number of tests are ordered (in 2003 the laboratory at University Hospital Birmingham received 67,182 requests for LFTs from 83 GP practices representing 210 GPs and 9,779 (15%) were abnormal in the sense that at least of one of the analytes on the LFT panel exceeded the reference range). Since LFTs are inexpensive and easy to organise as one of the standard 'blood tests' in the general practitioner's repertoire, their widespread use has occurred without careful study of their meaning in a general practice setting. Since the meaning of the various combinations of possible test results and clinical features is unclear, different practitioners have responded in different ways to the same test profile – the eclectic nature of practitioners' responses to the same scenarios has been well documented[1].

On the one hand, many mildly abnormal LFTs are false positives. On the other, irreversible and progressive liver damage is frequently put in train well before the liver disease becomes clinically apparent. LFTs then have the ability to detect diseases when they are most treatable, for example by reducing overload in patients with metal storage diseases or by administering modern anti-viral agents in those with chronic viral hepatitis. Furthermore, theory-based interventions designed to modify behaviour which leads to liver damage, while clearly far from a panacea, are effective [2-4]. This evidence confirms the common sense notion that people are most likely to adopt healthy lifestyles when they perceive that their health is threatened and that engaging in the recommended behaviour will reduce this threat. Moreover, new hypotheses for improved behavioural and medicinal treatment

for alcohol and calorie induced liver diseases are likely to be proposed in the future. The study we propose will therefore serve as a platform for long term observational and possibly interventional studies for the treatment of liver diseases.

The incidence of many liver diseases is rising, for example with migration from places with high rates of chronic hepatotoxic viral infection and as a result of alcohol and calorie excess. Co-morbidity is becoming more common as alcohol misuse and calorie excess unmasks other diseases of the liver, such as haemochromatosis. Thus, three interacting factors create an urgent need to better understand the clinical epidemiology of abnormal LFTs:

i) frequent use of these tests
ii) lack of clarity about the meaning of the results
iii) increasing treatability and rising incidence of liver diseases

2.2 Objectives. Our objectives have been formulated to undertake a rigorous assessment of value of abnormal LFTs among patients in primary care with non-specific symptoms based on methodologically robust frameworks for evaluation of tests[5;6], taking into account the recently published STARD statement [7] and the QUADAS quality instrument [8]. The research objectives are to:

1. Determine the value of LFT abnormality in predicting the risk of serious treatable disease.
2. Generate, using multivariable modelling, the probabilities of serious treatable disease according to:
   - the type and severity of LFT abnormality and
   - clinical and demographic features of patients in GP settings.
3. Discern how much the various initial and follow up tests which a GP may order (including ultrasound) contribute to the final diagnosis.
4. Measure the psychological effects of test results on patients.
5. Generate a cohort of patients for long term follow up beyond the HTA funded study.

These objectives are defined at a finer level of granularity in section 3.5. At this point, we emphasise that the research objectives, consistent with the commissioning brief, focus on information to inform GP decision-making. Clearly, a crucial decision concerns whether or not to pursue a mildly abnormal test by referral or further testing.

2.3 Existing Research

2.3.1 Systematic review. There is considerable literature on the laboratory measurement of analytes. Dufour and colleagues carried out a systematic review of this topic [9]. This review contains much useful information on biological variability and how it is affected by sex, age, race, use of the oral contraceptive pill (and other medicines), pregnancy, exercise, delay in analysis, and time of day. The study also reviews the patterns of abnormality of each analyte in the different diseases. In a second article, Dufour and colleagues systematically review the literature or the relationship between LFTs and disease and formulate guidelines for the interpretation

of LFTs [10]. Most of the 220 useful references describe test results in given diseases, rather than the probabilities of the various diseases given test results – for example '… diagnosing advanced fibrosis in cirrhosis in patients with chronic hepatitis C infection' [11]. A number of articles report results of follow-up of patients with specific LFT abnormalities, such as 'notably raised aspartate aminotransferase' [12]. A small number of papers deal with raised aminotransferase or transaminase levels in asymptomatic patients [13-15]. A number of papers deal with the results of liver biopsy in people with chronically abnormal results [16-18], but this does not tell us whom to investigate further, because none of the studies follow up all patients with LFT abnormalities. A number of important papers have been published in the years since the Dufour review, perhaps most notably a recent analysis of insurance data [19], which showed a correlation between aminotransferase concentration, even in the normal range, and outcome in the Korean population (where hepatitis B carriers are common). A further systematic review which distilled 14,000 references was commissioned by the American Gastroenterology Association Clinical Practice Committee [20].

In the absence of the necessary primary study, a number of authors have nevertheless produced diagnostic algorithms for the investigation of people with abnormal LFTs[21-26]. These provide sensible advice – for example stressing the importance of taking a careful family history, or of responding to tests which suggest obstructive biliary disease – but they do not provide a clear probabilistic basis for their reasoning. This is what we will provide. This is not to say that we will remove the need for judgement – rather, we will provide probabilities, such as the probability that a certain combination of clinical and laboratory features are benign, on which rational judgements may be based.

2.3.2 Pilot work. In addition to the statistical work mentioned above (by Alan Girling, see section 2), the design of our proposed study required an estimate of the frequency and nature of abnormal LFTs in General Practice. We have reviewed LFTs recorded in Birmingham. In a single calendar year (2003) 67,182 LFTs originated from 210 GPs in 83 practices. The Birmingham LFT panel consists of 4 analytes (alkaline phosphatase (ALP), transaminase (AST), total bilirubin and albumin) and 9,779 (fifteen percent) of patients had at least one abnormal result, while 11,277 of the approximately 270,000 individual analytes were abnormal. Seventeen percent of tests were among people who had had a previous test within the last 3 months (ie they are likely repeat tests). Thirteen percent of the remaining tests were abnormal (for one or more analytes). The percentages of these patients with 1, 2, 3 or 4 tests for individual analytes in the abnormal range were 11.1, 1.4, 0.4 and 0.1 respectively. The great majority with 'abnormal LFTs', have only one abnormal result within the panel. Thus it will be possible to calibrate the risks of different disease conditions across a wide spectrum of both normal and abnormal LFT results (see 2.5.2). The ALP result was abnormal in 5.7% of cases, AST in 4.4%, bilirubin in 3.2% and albumin in 1.6% of cases. The high rates of abnormal AST and ALP suggest that GPs are genuinely selecting for a population at higher risk than age-matched population as a whole. Seventy and 73% of patients with raised AST and ALP respectively had no other abnormal tests on the panel. One quarter of AST results were more than twice the upper limit of the normal range.

2.4 Justification for Prospective Study: Avoidance of Bias. A prospective study on the scale required offers good value for money:

- Secondary research and guideline development have been taken as far as they can go (see 2.3 'Existing research'); hence the HTA call for 'primary research'.
- Retrospective analysis is limited by biases which can be largely avoided by a prospective study. The documented lack of uniformity of testing practice[1], means that any database compiled from past practice would be not only massively incomplete but potentially biased. For example, whether or not a patient has a repeat test might depend not only on variables recorded in the database, but also other evidence that the doctor might have discerned. By following up all cases within the population we define, a prospective study will avoid bias associated with difference in disease severity or progression. In a retrospective study, clinical data would have to be retrieved from case notes or databases where they would not be recorded in a consistent or comprehensive way. Outcomes are not only coded, but diagnosed, inconsistently in liver disease; there may be difficulty in distinguishing between alcoholic hepatitis or non alcoholic steatosis for example. Only a prospective study can include measures standardised to ensure that these are diagnosed in a transparent and consistent way (and hence avoid bias due to variation in the reference standard). Our proposed prospective study allows us to ensure that all patients receive the same basic tests (see 2.5.3 and 3.3) so that spectrum variation can be explored by modelling when extrapolating to populations in different times and places with different disease profiles. Our study design will allow us to ensure uniformity of measurement and interpretation of index tests. Moreover, performing a comprehensive package of tests on all patients ensures that diseases such as hepatitis C and haemochromatosis, which may not become overt for 10 years or more, are identified by specific testing.

We have consulted the Medicines and Healthcare products Regulatory Agency (MHRA) about the possibility of linking LFT results (downloaded electronically in many practices since 2000) with outcomes recorded on the GPRD database by 2007. This is feasible, but given the above reasons, we conclude that this will offer little or no marginal benefit.

What is missing from the literature is a substantial prospective study of a well documented population given a standardised diagnostic work up in general practice and then followed up for a period of time; as Green and Flamm state[20]: 'Unfortunately … there are no long term prospective studies to define the natural history of liver disease in patients with abnormal liver chemistries tests'.

According to this statement, a follow up study of patients with abnormal LFTs is needed and the HTA call for proposals also defines the patient group as 'Patients… found to have abnormal liver function tests.' Therefore as outlined in our original application we shall start with a population of patients with an abnormal test and hence we will not have negatives (whether false or true) in the sense that all the analytes in the test fall in the normal range (and many people with normal tests are likely to decline follow up for study purposes). Thus the evaluation of patients with abnormal LFTs is different to standard evaluation of diagnostic tests in that:

1. We start with a population of test positives, not all patients who might have a disease
2. The LFT panel is a panel of tests (4 to 7) rather than just 1 test
3. The diseases of interest are very large in number and vary considerably from common to extremely rare

These differences raise some fascinating methodological problems to which we now describe our approach.

2.5 Methodological Issues

2.5.1 Negative results within the panel of LFTs. Although we will not include those cases negative on all tests in the LFT panel, we will have very large numbers who are negative on most of the tests (see 2.3.2). This means that LFTs/risk relationships can be estimated by regression analyses across the full range of LFT results. The fact that most patients test abnormal for one analyte only will allow us to evaluate the predictive value of negative tests on other analytes.

2.5.2 Comprehensive and standardised testing. In order to be comprehensive the initial (index) LFT panel must include all analytes which may be used (see 2.5.3). Secondly, in accordance with a STARD requirement (to avoid disease progression and severity bias) the diagnostic algorithm must apply to an entire population of interest and be as comprehensive as possible within the bounds of good practice. In the context of this study this means that the initial (basic) package of investigations must apply to all participants in the study (we get around the problem of not inconveniencing patients by carrying out multiple tests on single blood samples or at single visits – see study design). This means that not only will all patients have repeat LFTs (consistent with prudent practice and to track LFT profiles over time), but their blood sample will be subject to further testing and they will have an ultra sound examination. The panel of LFTs can indicate that the liver may be affected by a disease, and if so, suggest broad groupings of disease, but they are nearly silent on what the particular underlying disease (if any) may be. There are, however, a number of relatively inexpensive blood tests which are much more specific (Table 1) and which will be done in each case.

Table 1: Diseases for which relatively selective blood tests exist.

Disease	Approximate prevalence in English population (%)
Chronic viral hepatitis C	0.4 [27]
Chronic viral hepatitis B	0.3 [28]
Metal storage disease – Iron	0.25 [29]
Primary biliary cirrhosis (PBC)	0.001 [30]
Autoimmune hepatitis	0.001 [31]
Metal storage disease – Copper	<0.025
Alpha-1 antitrypsin deficiency	<0.025 [32]

2.5.3 Making the diagnosis – the reference standard. The design of the study has to take into account 3 particular features of the clinical epidemiology of diseases (some not even affecting the liver) which may cause abnormal LFTs.

1. There is a very wide range of diseases of different severities. The practical implication of this is that while it is eminently feasible for all patients to have the same initial testing protocol, it would be completely inappropriate for all patients to receive identical further follow up. Some will appear to have no disease, others infiltrating disease of the liver, for example. Secondly, while we shall seek to standardise subsequent testing as much as possible (see 2.5.3), it will not be possible to ensure that every case with, for example, space occupying lesion, is investigated in precisely the same way. This does not matter provided it is accepted that the setting is 'primary care' as stated in the HTA call. We can then answer the question 'What is the meaning of LFTs obtained in a primary care setting in terms of the likelihood of serious disease of various types?'. If one wanted to answer questions of the sort: 'How should specialists investigate patients with obvious serious disease to more efficiently reach a diagnosis?' we would need to start with a different population – those referred from general practice and then standardised testing could elucidate the role of, for example, MRI versus liver biopsy. It is sufficient for our purposes that patients needing further investigation will be identified and investigated by experts.
2. Gold standard versus less clear cut diagnosis In many cases a gold standard diagnosis will be reached in the sense that a test which is independent of the LFT panel will clinch the diagnosis – for example, many of the tests for table 1 diseases or the result of a biopsy showing cancer. In the case of people who fall into three large groups, however, such a gold standard diagnosis is not possible. These are, people with no disease, alcoholic steatohepatitis (ASH) and non-alcoholic steatosis (steatohepatitis) (NASH). These must be diagnosed by consensus methods and there is a risk of incorporation bias in such cases. However, the 2 years follow up will make diagnosis a little more solid than would be the case if outcome was assessed earlier and this is also the reason why we shall use the cohort generated in the study for longer term follow up.
3. Very variable prevalences We discuss this in detail in the section on power calculations, but here point out that some diseases will be very rare, so that precise probability estimates would not be possible, even in a sample of tens of thousands and our approach here is to amalgamate these conditions in groupings which makes sense from established clinical and pathophysiological principles.

2.5.4 Generalisablity. Normally in clinical epidemiology efforts are made to recruit a population of average risk, i.e. representative patients of the country at large. However, the incidence of liver disease is unstable in both place and time. For example, it changes with migration and immunisation practice. We have more chance of finding out how the probabilities vary by clinical/demographic features if we study populations which have sizeable sub-groups at high risk of the individual conditions. Hence, we have selected two sites with high-risk populations. This will enable us to model the relationship between test results and risk by patient characteristics, rather than produce a static measurement which can never be representative of the diverse circumstances in which study results should have relevance.

2.5.5 Probabilistic information and algorithms. Judgement is required in moving from information to guidelines. Firstly, the importance of not missing a diagnosis varies. Failure to diagnose NASH is less serious than failure to make a timely diagnosis of haemochromatosis where delay is associated with a sizeable risk of permanent harm. Secondly, some diagnoses, especially NASH/ASH are not completely clear cut. Thirdly, we

are not starting from scratch - certain patterns are already known. For example, it is known that persistently raised bilirubin in a patient whose tests are otherwise normal can be a sign of the relatively common and benign genetic condition called Gilbert's Syndrome[33]. Raised alkaline phosphatase and bilirubin is a sign of one of the biliary or infiltrative diseases, while raised transaminases are a sign of diseases mainly of the hepatocyte. Such insights will be a crucial concept when deciding which variables to analyse in connection with particular diseases (see also 3.6). We have therefore built in consensus development into the work programme to prioritise these variables as well as to develop consistent approaches to diagnosis (section 3.9).

2.5.6 The patient's perspective. The call for proposals asks for this perspective to be taken into account. We propose an approach based on measuring patients' cognitive, emotional and behavioural responses to an abnormal liver function test and subsequent investigations. When we consider that the 'treatment' for many patients with early liver disease is behaviour change, the need to explore the "meanings" that patients attach to LFT investigations by measuring emotional and cognitive impact becomes clear. Such change is not likely to be forthcoming if patients are insouciant about their test results: in an emotional sense and when the patient considers the relevance and importance of the results to their ongoing health. It is also noted however that generating anxiety and concern without coupling this with support to assist a change in lifestyle, poses double jeopardy – psychological anguish with no countervailing health benefit. By looking at patient engagement in health-related behaviours over the investigation schedule we will be able to explore the psychological factors that may be related to an increased likelihood to engage in behaviour which promotes liver function. Such as a reduction in alcohol consumption.

This line of research will also provide some indication of the varying ways in which different types of investigation, such as ultrasound, scans impact on patients. For example, these tests may create more motivation to engage in risk reducing behaviours than blood tests, net of their clinical portent. The results in this regard ill also inform algorithm production where preferences need to be taken into account. For example, if imaging tests have strong motivating effects on patients they may need to be included in the algorithms even if, in a purely statistical sense, they do not contribute greatly to diagnostic precision – at least such a finding would suggest the need for more empirical work on the point.

Lastly, a word on our philosophy on the use of clinical algorithms. We believe these should guide practice and that they should not trump the clear preference a particular patient may have (say, for more extensive testing in someone with an affected relative). For an algorithm to be useful it is not necessary that it should be followed with total fidelity in every case. To summarise, we bring in the patient perspective at four levels:

a. To evaluate the duration and intensity of experience at different stages of the testing programme.
b. To understand how patients 'construct' test results in order to develop propitious strategies to convert any engendered anxiety into improved health.
c. To develop algorithms to use in the typical (default) situation in such a way as to be highly sensitive to the needs of patients as a whole.

d. To draft algorithms in such a way as not to foreclose on any particular preferences an individual may have.

3. Research Methods

3.1 Overview. A database will be assembled prospectively and with patient consent in the two centres (Lambeth and Birmingham) (see Appendices 1a, 1b, and 1c). The number of tests included in the LFT panel and the number and range of subsequent tests ordered in primary care will be as comprehensive as possible within the constraints of cost and convenience and will fall within the range of (currently diverse) acceptable clinical practice. For example, the current four panel LFT screen in Birmingham will be extended to seven (see 3.4.1). This is necessary to evaluate the discriminatory powers of algorithms less inclusive than the study algorithm. The pathway of care has been based on discussions with general practitioners in both locations and has been carefully constructed so that it is often less taxing (and seldom more taxing) for patients than existing pathways of care. Patients (n = 1,500) from four large practices (i.e. two from each centre) will be invited to participate during an 18-month recruitment phase. They will receive a standardised package of tests, and will then be followed up for two years, at which point a standardised method will be used to determine the reference standard (diagnosis). The database thus constructed will be interrogated by discriminant analysis to answer a number of specific questions, most identified in advance. These results will then be fed into two guideline development groups, which will each follow a Delphic process to develop sample algorithms for care. We will seek patient consent for follow-up beyond the lifetime of the HTA funded study itself.

3.2 Study Locations (Centres). We have planned our study around two major centres (Lambeth and Birmingham) which have:

1. Large throughput of LFTs generated in primary care and tested in a single central laboratory.
2. A diverse population, including migrants and inner city dwellers, providing sub groups at risk of certain diseases.
3. Well established liver disease services providing expert oversight and advice.

The aim is to generate a sufficient population of patients with abnormal tests in a relatively short period of time, enabling us to produce precise estimations of risk, even for fairly rare conditions such as viral hepatitis. We need 1,500 consenting patients with abnormal LFTs (see 3.5). Use of two centres allows us to have a replication sample (see 3.5.3) and allows the specialists' work to be shared (see 3.4.2).

3.2.1 Practices. In a single year in Birmingham (2003), 210 GPs requested 67,182 LFTs of which 55,761 (83%) were new tests i.e. tests of previously untested patients and of these 13% were abnormal (7,248). Thus 'the average' GP (the database does not distinguish full and part time) generates a mean of 266 new tests each year, of which 35 are abnormal. Using Birmingham as a surrogate for Lambeth, if we assume that 10 cases in 35 will either have obvious liver disease or not wish to participate, then 1,500 patients will be recruited in 18 months from six practices [Lordswood House Medical Practice, Yardley Wood Health Centre, Greenridge Surgery, Woodland Road Surgery, Cofton Medical Centre, Wand Medical Centre Shenley Green Surgery and Hall Green Health] in Birmingham and 2 [Lambeth Health Group, Waterloo and Harley Clinic] in Lambeth) with a mean of 10 GPs. We allow 24 months to include a start up phase and leave some time for contingencies.

The practices submit all samples to the Birmingham or Guys/St Thomas' Trust laboratories. Use of a limited number of practices allows us to standardize procedures cost effectively while ensuring a reasonably broad spectrum of patient and clinical features.

We will be working with the Midlands and Lambeth Research Practices Consortia (MidRec and STaRNet) that have successfully recruited subjects to a number of major studies funded by the MRC and HTA. PG and RJ will be leading on this and if we are successful, we have secured 'Support for Science' funding to enable professionals to participate in this research.

3.3 Clinical Protocol

3.3.1 Initial testing. The call for proposals is based on abnormal LFTs in primary care. General Practitioners are busy people, and to expect them to remember the study at the point of initial testing will result in losses from the study. We shall maximise ethical recruitment as follows:

1) The practices will be visited by the site study co-ordinators and applicants at the start of the study, and will remain in frequent contact throughout.

2) With permission, posters informing about the study will be displayed in the public and other areas of the practice.

3) Blood will be taken in the normal way for each practice.

3.3.2 Formal enrolment in subsequent testing protocol: defining of the patient population and seeking consent.

At a routine appointment GPs will request routine liver function tests (LFTs). The pathology laboratories involved in the study routinely transmit blood test results electronically to the originating GPs and during the study will also provide a weekly list of abnormal index LFTs for each surgery.

At this point in the clinical protocol some changes have been made to bring the study into line with routine practice at the surgeries involved. Individual practice procedures/patient process flow charts are detailed in appendix 10.2.a to e.

a) <u>Birmingham Practices</u>

Hall Green Health, Greenridge Surgery Yardley Wood Health Centre and Woodland Road Surgery

1. A designated secretary at these surgeries will inform GPs if their patients appear on the pathology laboratory list of abnormal LFTs in order to remind GPs about the study. GPs will decide on the suitability of their patients for inclusion (if no previous LFT analysis in the past 12 months, over 18 years old, with no sign of liver disease – self-evident jaundice or signs of liver failure - or disseminated cancer, and not pregnant).
2. If GPs feel that the patient is deteriorating or has obvious disease and should be referred for specialist care at this stage, referral will take place in the normal way. Such patients who are clearly sick are not the focus of

this study which is concerned with the very much more common scenario of abnormal LFTs of unknown provenance.

3. Within 6 weeks of the blood test results arriving, GPs will telephone patients to provide their LFT results, as they would routinely do at these practices. They will then explain the study and invite patients to the nurse practitioner clinic (HGH) or 'Designated GP' clinic. If patients have agreed to take part it will be noted on a shared electronic drive by GPs. The GP will also record whether a patient speaks English, and if not, whether he/she lives with someone who does so and whom the patient agrees may speak on his/her behalf;
4. Each week the secretary will contact patients listed on the shared electronic drive, to invite them to one of the weekly BALLETS study clinics and the standard Patient Information Sheet (see Appendix 10.3.a, c & d) for the study will be sent.
5. At the clinic the research nurse will:
    - Explain the study further;
    - Ask patients to sign a Study Consent Form (Appendix 10.4.a) if they are interested in taking part;
    - Invite the patient to attend for an ultrasound appointment at the surgery;
    - Complete a brief template with clinical details for the study covering alcohol and transfusion history, substance abuse, drugs, chronic disease such as diabetes, travel and immunisation history, demographic details, any acute illness, height, weight and abdominal girth, reasons for the original test being ordered;
    - Document the patient's NHS number (for follow-up if the patient moves);
    - Take blood samples for microbiology, biochemistry and immunology analysis. Some would argue that further blood testing should be offered anyway to all such patients and few if any would say that it would be inappropriate. When it reaches the laboratory, it will be comprehensively tested – repeat LFTs, tests for diseases in Table 1 and additional tests of future scientific interest (lipid profile and haemoglobin A1c).
6. Patients will visit the practice for an ultrasound scan. Blood test results and ultrasound summaries will be collated by the study coordinator and given to practice secretary who will invite patients to attend for a follow-up appointment with their GP (described in 3.3.5).
7. The study coordinator will confirm that the patient has consented before despatching the first psychological questionnaire by mail and asking the participant to bring the completed form to the GP appointment.

The only additional tasks for GPs at these practices will be to explain the study and invite patients to participate.

Lordswood House Medical Practice

1. GPs at this practice routinely access electronic pathology results for their patients each day. Patients who meet the inclusion criteria for the study (if no previous LFT analysis in the past 12 months, over 18 years old, with no sign of liver disease – self-evident jaundice or signs of liver failure - or disseminated cancer, and not pregnant) will have a short message placed alongside their results by the GP to notify receptionists that patients should be invited to join the study.

2. Patients from this practice are routinely asked to telephone for blood test results. When patients who are identified as eligible telephone the practice, the receptionist will provide blood test results and a brief explanation of the study (see copy of script attached as Appendix 10.5.a). The patient will also be invited to attend for a GP appointment and will be sent a letter by GPs (Appendix 10.5.b) at the practice and the Patient Information Sheet (see Appendix 10.3.b) for the study.
3. At the appointment the GP will:
    - Give an explanation of the study;
    - Ask the patient to sign a study Consent Form (see Appendix 10.4.a) if they are interested in taking part;
    - Invite the patient to attend for an ultrasound appointment and further blood test at the practice – these tests will take place at the second appointment;
    - Complete a brief electronic study template with clinical details for the study covering alcohol and transfusion history, substance abuse, drugs, chronic disease such as diabetes, travel and immunisation history, demographic details, any acute illness and reasons for the original test being ordered;
    - Record whether the patient speaks English, and if not, whether he/she lives with someone who does so and whom the patient agrees may speak on his/her behalf;
    - Document the patient's NHS number (for follow-up if the patient moves).
4. When patients visit the practice for an ultrasound scan, they will have a blood sample taken by a phlebotomist for analysis at the microbiology, biochemistry and immunology laboratories. Some would argue that further blood testing should be offered anyway to all such patients and few if any would say that it would be inappropriate. When it reaches the laboratory, it will be comprehensively tested – repeat LFTs, tests for diseases in Table 1 and additional tests of future scientific interest (lipid profile and haemaglobin A1c). During this visit they will also have their height, weight and abdominal girth measured.
5. The designated study secretary at the practice will invite patients to attend for a follow-up appointment with their GP (described in 3.3.5).
6. The study coordinator will confirm that the patient has consented before despatching the first psychological questionnaire by mail and asking the participant to bring the completed form to the GP appointment.

The only additional tasks for GPs at this practice will be explaining the study, inviting the patient to participate and providing a few more clinical details on a study request form than otherwise might be included.

Cofton Medical Centre, Shenley Green Surgery and Wand Medical Centre

1. GPs at this practice routinely access electronic pathology results for their patients each day. Patients who meet the inclusion criteria for the study (if no previous LFT analysis in the past 12 months, over 18 years old, with no sign of liver disease – self-evident jaundice or signs of liver failure - or disseminated cancer, and not pregnant) will have a short message placed alongside their results by the GP to notify receptionists that patients should be invited to join the study.

2. Patients from this practice are routinely asked to telephone for blood test results. When patients who are identified as eligible telephone the practice, the duty or triage doctor will provide blood test results and a brief explanation of the study. The patient will also be invited to attend for a BALLETS study appointment and will be sent a letter by GPs (Appendix 10.6) at the practice, the Patient Information Sheet (see Appendix 10.3.g & 10.3.h) and a psychology questionnaire for the study (see Appendix 10.10.c). The GP will also record whether the patient speaks English, and if not, whether he/she lives with someone who does so and whom the patient agrees may speak on his/her behalf.
3. At the appointment the BALLETS research nurse will:
    - Give an explanation of the study;
    - Ask the patient to sign a study Consent Form (see Appendix 10.4.a) if they are interested in taking part;
    - Complete a brief electronic study template with clinical details for the study covering alcohol and transfusion history, substance abuse, drugs, chronic disease such as diabetes, travel and immunisation history, demographic details, any acute illness and reasons for the original test being ordered;
    - Document the patient's NHS number (for follow-up if the patient moves).
    - The patient will have blood sample taken for analysis at the microbiology, biochemistry and immunology laboratories. (Some would argue that further blood testing should be offered anyway to all such patients and few if any would say that it would be inappropriate.) When it reaches the laboratory, it will be comprehensively tested – repeat LFTs, tests for diseases in Table 1 and additional tests of future scientific interest (lipid profile and haemaglobin A1c).
    - The patient will also have their height, weight, and abdominal and hip girths measured.
    - Finally the patient will have an abdominal ultrasound performed by an ultrasonographer.
4. Patients will be asked to contact the duty or triage doctor in 3 weeks for their results. The doctor will advise the patient whether a follow up GP appointment is required.
5. The study coordinator will confirm that the patient has consented before advising the psychology assistant to despatch the first psychological questionnaire by mail.

The only additional tasks for GPs at this practice will be explaining the study, inviting the patient to participate and providing a few more clinical details on a study request form than otherwise might be included.

b) <u>Lambeth Practices</u>

1. Patients having LFTs will be asked at the time of the blood test to consent to a study researcher contacting them if they meet the study criteria. (See copy of Lambeth Pre-consent form – awaiting COREC Substantial Amendment approval – Appendix 10.6) Results are transmitted electronically to the originating GP and patients will be informed of their results according to the GPs normal practice. In addition, a summary of all abnormal results for each practice will be provided by the laboratories on a weekly basis.

2. Patients identified as eligible for the study will be confirmed with reference to patient records and their GP where appropriate. Inclusion criteria will be patients of at least 18 years of age, not pregnant and none of the following: known liver disease (self evident jaundice or other signs of liver failure) or known disseminated cancer.
3. A study researcher will then make contact with the patient (usually by telephone) and will:
    - Explain the study;
    - Invite patients to attend an appointment for an ultrasound examination and second blood test;
    - Confirm the appointment in writing, including a patient information sheet (Appendix 10.3.e); full study consent form (Appendix 10.4.b) and the first psychology questionnaire, asking the patient to complete them and bring to their appointment.

   Any patient requiring an interpreter will be identified at this point and appropriate arrangements made.
4. At the clinic, a study researcher will collect data from consenting patients on height, weight and abdominal girth as well as substance abuse (including alcohol), drugs and travel and immunisation history. In addition patients will be asked about socio-demographic details and clinical history and NHS number (for two year follow-up) will be extracted from patient records. Patients will have a further blood test and ultrasound examination.
5. Results of the second blood test and scan will be reviewed by the GP and normal clinical practice followed to inform patients of the results.
6. A study researcher will contact participating patients after 2 years to invite them for a repeat blood test.

7. In all practices participants undergo:

i) Ultrasound

ii) Data collection: height, weight and recording of the NHS number (for follow up if the patient moves).

iii) Consultation with the GP to receive the result of the blood tests and ultrasound and discuss implications.

We will now describe each of these in more detail.

3.3.3 Ultrasound.

1. Personnel. We will appoint a panel of part time radiographers (2 or 3 per centre) who will each do 2 to 5 sessions each week. They will be paid a basic rate for availability and an additional component for each patient they see and payment and quality control will be through the host ultrasound department. Our enquiries have confirmed that there are ultrasound personnel who would wish to be employed on this basis and who would enjoy the research, outreach nature of the work and degree of autonomy. Two portable ultrasounds will be available in each locality. This will be sufficient for approximately 3 ultrasound tests in each centre per day. Quality control is discussed below.
2. The patient will be asked to omit the meal preceding the ultrasound.
3. The ultrasound will cover the following basic features - size, echogenicity (texture), any space occupying lesions, state of the biliary system, ascites and spleen. A copy of the Ultrasound Examination data collection form is included as Appendix 10.7.

4. Ultrasonographers will share images with patients.
5. If an abnormality of serious portent e.g. space-occupying lesion, is discovered, this will be disclosed to the (or a) GP immediately after the scan.

3.3.4 Collection of patient information.

The ultrasonographer will be taught to collect the basic information under 3.3.2, point 7 ii, above (Appendix 10.7). It is easier to teach an ultrasonographer to collect this information than to teach another member of the clinical team to do ultrasound. Also, having one person do both tasks (ultrasound and data collection) makes co-ordination much easier.

If the patient cannot speak and read English, then an interpreter will be invited to attend, as per normal practice, but an extra payment will be made to cover the ultrasound and data collection phases of the visit, in accordance with good research practice and in recognition of the diverse ethnic make up of large English cities[34;35]. The psychological questionnaires will not be translated into languages other than English as there is insufficient time to allow such questionnaires to be validated.

3.3.5 Consultation with the GP.

Amendments have been made to this section of the protocol in consultation with principal investigators, to bring the study into line with routine practice at the surgeries involved.

Birmingham and Lambeth practices

1. Blood test results and ultrasound summaries will be collated by the study coordinator and given to GPs in order to provide feedback to patients. (A copy of the Consolidated Report form is attached as Appendix 10.8.) Sonographers will provide immediate feedback to GPs if any abnormality is detected so that a more detailed ultrasound examination can be arranged. The designated study secretary will invite patients to attend for a follow-up appointment with their GP.
2. GPs will continue to access blood test and ultrasound scan results electronically as they routinely do.
3. The study hepatologist will provide training sessions for GPs on feedback to patients when all the study results have been collated. The hepatologist has supported these teaching sessions by providing a flow chart of patient referral guidelines, which can be used as a guide by GPs. Please see flow chart attached as Appendix 10.9.

The timing of the final visit to the GP will take place within 2 weeks of the results from the second visit becoming available and will be timed on GP availability and patient convenience, but, in most cases, this will leave a number of alternatives, so that more than one patient can be scheduled for a particular surgery session. This will be a little more complex to arrange then, say, a rapid-access hospital clinic, but it is more convenient for patients and will engender greater participation and it is therefore a cost effective solution.

In order to facilitate this consultation and to assist the GP by reducing the need to toggle between screens, the research team in Birmingham will produce a consolidated report (Appendix 10.8), which will be faxed or transported to the relevant practice (it is not acceptable to transmit electronically due to NHS Patient Confidentiality code of conduct).

The consolidated report will contain the clinical history including alcohol consumption, BMI, transfusion and travel history, all blood tests collected under BALLETS and the ultrasound report.

3.3.6 Long term follow up. The follow up protocol for patients is illustrated in a diagram – see Appendix 10.11.b. All consenting patients will receive follow up after 2 years. They will have a blood test in primary care unless they are still under continuing hospital supervision, in which case they will be followed up from the hospital case notes (see below). Patients who have moved will be followed up through the Office for National Statistics (ONS), and their GPs will be contacted before they are approached.

Patients who have moved will be referred to as P2 and patients who have not moved and who are not under continuing hospital supervision, are described as P1. The P1 follow up sub-protocol is more extensive than the P2 sub-protocol. The case notes of all P1 and P2 patients will be scrutinised by the researcher. In the case of P2 patients the last date of attendance will be noted. It is recognised that data will be incomplete in these cases. A data collection form will be completed (Appendix 10.11.a) for all P1 and P2 patients.

The coordinator will find out from the practice if patients who have not moved are under ongoing hospital supervision. If not he/she will be invited by telephone to attend a clinic appointment at the practice for their follow-up LFT. The P1 patient will also be asked if they would consent to having an additional blood sample taken for storage and later testing (Birmingham only), and a follow up (T4) psychology and lifestyle questionnaire (Appendix 10.10.d). The extra patient information sheet (Appendix 10.11.c) regarding the additional blood sample will be posted to Birmingham patients.

At the follow up clinic Birmingham patients will be asked if they have questions about their appointment before being given the extra Consent Form to sign (Appendix 10.11.d). For all P1 patients, weight, waist and hip measurements will be taken; and the completed psychology questionnaire (T4) will be collected. If not collected patients will be invited to answer questions at this point, strictly avoiding any hint of coercion. The LFT (and in Birmingham additional blood sample for storage and later testing) will also be collected. The blood for storage and later testing will be taken to the Institute for Biomedical Research, Medical School, University of Birmingham (see Section 5 Add-on study). Blood for repeat LFTs will be sent to the local laboratory and the results will be retrieved after 3 weeks.

P2 patients will be contacted by telephone with their GPs permission, and asked to attend their local GP surgery for the follow up blood test. They will receive the T4 questionnaire by post and given a stamped addressed

envelope for the return of the questionnaire. The research team will contact their local GP surgery to ask for the LFT results.

Patients under ongoing hospital supervision will not receive a blood test or additional questionnaire. Their hospital notes will be retrieved and the data collection form (Appendix 10.11.a) will be completed. It will be recorded if patients have moved and are under ongoing supervision from their new local hospital. These patients will be identified when the overture is made to their new GP. Their diagnosis will be noted and they will not be followed up further.

3.4 Standardising and specifying index tests and reference standards

3.4.1 Index tests. Poor description of LFT index tests in terms of measurements and thresholds for defining abnormality are associated with bias when extrapolating to other populations where different procedures are used. We will standardise all testing procedures so that there are no ambiguities about testing and our criteria for interpretation and thresholds will be determined *a priori*. Measurements of these analytes will be undertaken by standard laboratory methods. Both laboratories at University Hospital Birmingham and Guys-St Thomas NHS Trust are accredited by Clinical Pathology Accreditation (CPA), and both laboratories participate in recognised external quality assurance schemes.

Table 2: Analytes in the LFTs panel

Analyte	Reference range
Albumin	34 – 51 g/L
Total Protein	60 – 80 µmol/L
Bilirubin	1 - 22 µmol/L
AST	3 – 43 U/L
ALT	5 – 41 U/L
ALP	70 – 330 U/L
GGT	9 – 50 U/L (male) 9 – 40 U/L (female)

Table 2: Reference ranges of analytes to be measured in the LFT panel. Abnormal LFTs will be defined as any result falling outside these ranges.

The ultrasound images will be recorded digitally and a random selection of 100 verified by the two radiologist applicants.

3.4.2 Standardising the final diagnosis/reference standard. In many cases, diagnosis by tests independent of index tests (e.g. hepatitis C or cancer) will be available. However, many cases will be more complex – e.g. dual diagnosis and cases where there is no gold standard, especially NASH/ASH. We will therefore use consensus diagnosis as a reference standard. This will be applied to individual cases at the 2 year follow up and will be

based on all information available at the time which will be assembled by the co-ordinator at each site. In the case of patients undergoing continual hospital care, the hospital notes will be used as further reference. In other cases, the general practice notes will be used. Diagnosis will be based on guidelines produced before the first patients complete their 2 year follow up – we describe this in section 3.9.2. The hepatologist at each centre will assign cases to diagnostic groups, when they are undergoing continued hospital supervision. The general practitioner collaborator at each site will do this for patients who have either not been referred to hospital or who have been discharged from hospital care.

We will also form 2 outcome committees. A random 200 cases from each centre (total 400) will also be referred to the outcome committee of which the centre hepatologist/ GP is not a member. Each outcome committee will be constituted of one GP applicant, one radiologist applicant, 2 hepatologists (one an applicant and another peer) and one chemical pathologist applicant. Each committee will contain similar numbers from each centre. Experience shows that a committee can review 42 cases per day, bearing in mind that many of the cases will be straightforward (e.g. transient transaminase abnormality with no disease, hepatitis C, metastatic cancer). This means that each committee will meet 5 times, starting when the first patients complete their 2 year follow up. The results will be fed back to the hepatologists and GPs as this work unfolds and we will make measurements of inter-rater reliability.

We are aware that consensus diagnosis entails the risk of incorporation bias, but:

1. Many diagnoses are made on the basis of independent tests so that this risk will be small in these cases.
2. Where the tests under evaluation are heavily incorporated in the diagnostic outcome (NASH/ASH) we will have 2 year follow up data to make the diagnosis more solid than if we made an immediate assessment and we will use our study, as we have said, as a platform for longer term follow up until an objective outcome materialises. When the outcome committee reviews a case it will do so with and without knowledge of the findings of the initial ultrasound carried out in each practice. This will provide a direct measurement of the extent to which this test contributes to diagnosis to back up the results of statistical analyses.

3.5 Protocol for the Analysis of the Results

3.5.1 Broad aim. We aim to stratify the risks of diagnosable conditions from results of LFTs and other available tests by logistic regression methods. We will perform analyses to elaborate any loss in diagnostic precision consequent on dropping certain tests in certain circumstances and in different intensities of follow up (see 2.5). Intrinsic to such an analysis is the concept of test interaction (which we shall measure) and hence possible test redundancy. The US Food and Drug Administration has recently produced draft 'Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests' which we have consulted [36]. Our analytic approach takes into account its recommendations and goes well beyond to accommodate for the complexity inherent in this project. We now identify the salient questions, which can be answered from the database, along with the algorithm/guidelines they inform.

Inevitably not every patient with an abnormal LFT will be included – clinicians may fail to invite patients and some may even decline. This could bias the study with respect to the total population of patients with mildly abnormal LFTs in general practice. In order to determine whether such bias is likely and to evaluate the extent of any potential bias we shall proceed as follows. Firstly the abnormal LFTs from patients not included in the study will be anonymised in the practices. Next we will enter the anonymised data on the database. Then, when we come to the analysis, we will compare the pattern and extent of abnormalities in the participating and non-participating patients. If necessary, modelling within the entire data-set will enable us to extrapolate, with due caution, to a totally unselected population.

3.5.2 Specific questions identified in advance

1. What would be the effect, if any, of dropping one or more of the tests in the comprehensive study LFT panel, on identification of:

a) Serious treatable diseases (Table 1 diseases and/or alcoholic hepatitis and/or systemic disease)

b) All patients who have persistent abnormal results at the 2 year examination.

This informs the decision on what should be included in the standard panel of LFTs. For example, what (if anything) is gained by measuring both alanine and aspartate aminotransferase?

2. What profile of initial test results and clinical features suggest higher and lower risks of:-

a) Having one of the serious and specific, but treatable diseases of the liver in Table 1. For example, to what extent, if any, do patients with only very mildly elevated transaminases have a higher risk of chronic viral hepatitis than 'controls' who have only mildly abnormal alkaline phosphatase. If so, is this restricted to people at risk from their demographic/clinical histories?

b) Persistent liver disease of any type (over 2 years)?

c) Disease of other systems?

d) What is the relationship between ascertained alcohol intake and the risks and severity of various diseases?

Informs guideline on who to investigate for specific diseases in Table 1 and who to re-test more than once

3. How does the probability of abnormal ultrasound vary by the initial clinical features and test results? In certain cases, the probability will be no higher than in age-matched controls. Even though we will not test people with normal results, we will still be able to impute this figure approximately by comparing the probability of ultrasound abnormalities in people with different patterns of results through logistic regression analysis – e.g. the risk of an echogenic liver might vary with level of transaminases, down to a threshold where it is no longer higher than in people with isolated alkaline phosphatase abnormality or borderline reduced serum protein. For example, is there a steep gradient in the risk of an echogenic liver by test results? Do people with an echogenic liver but no gold standard diagnosis have a worse prognosis, net of index (initial) LFTs and history? The cohort we will

follow up with funding from elsewhere will answer this question with respect to cirrhosis, liver failure etc in the longer term.

Informs the guideline on which patients a GP should invite to have ultrasound testing.

4. What proportion of tests are abnormal at 24 months in people without a clear-cut diagnosis and how does this vary by clinical features, by ultrasound results and by initial tests results? To what extent are 2 abnormal tests within 2 months indicative of diseases of different types and class?

Informs guideline on when a second normal test should be treated as reassuring, e.g. if transaminase is slightly elevated, but normal on re-testing, can a transient abnormality be inferred.

5. In patients without gold standard diagnoses (ie ASH/NASH or no apparent diagnosis), what are the rates of deterioration or improvement in results of LFTs by initial features, results of further tests and reported changes of lifestyle?

Informs guidelines on repeat testing and feedback, especially in light of patient wishes and how they 'construct' medical knowledge.

6. How often can patients be put in secure diagnostic categories; how often is dual diagnosis present; and how often are patients left with indeterminate or insecure diagnoses, such as persistent abnormal tests with no apparent cause?

We may find in our psychology study (see below) that patients value test results, not only for their medical utility, but also for their 'newsworthiness'. In this case, the ability to provide better 'news' will need to be included in guideline development. For example, ultrasound in certain categories of people may not add a great deal in distinguishing between alcohol or dietary induced damage, but patients may wish to know whether their disease has yet produced ultrasonically detectable morphological change and may be highly motivated thereby to change behaviour.

7. Given how tests are currently interpreted to make non-gold standard diagnosis, what is the contribution of the different tests in making the final diagnosis?

Although not normative in selecting the strategy to most accurately identify the 'true' underlying diseases, this can inform the decision on how to use tests most parsimoniously given existing diagnostic practice. One of the reasons we wish to maintain the cohort beyond the HTA funded study period, is to be able to repeat statistical analyses to see how different testing strategies perform in identifying the long-term outcome for patients – see 2.5.3.

8. What are the cognitions, emotions and threat-related behaviours generated in patients by the process of assessing their liver functioning? What is the relationship between thoughts, anxiety and the testing process? What is the relationship between thoughts, feelings and intentions to engage in behaviour change to improve liver functioning (addressed under 3.7)?

3.5.3 Analysis – model building. Our basic approach is to use logistic regression methods to explore the relationships between diagnostic outcome and index test results taking account of clinical variables. Binary regression methods – appropriate for analysing a single disease or group of diseases – will be supplemented as necessary by polytomous (nominal) logistic methods and classical discriminant analysis if these methods prove useful in the simultaneous modelling of separate risks of different diagnostic categories. The choice of predictor variables when considering a particular disease will be informed by existing knowledge, since we already know which variables are most likely to be predictive of particular diseases, or groups of disease. Though we expect to make some use of statistical variable selection methods, we will not be trawling through the data in an undirected fashion, giving some protection against overfitting. Our approach for prioritising variables for statistical testing is outlined below (section 3.9). Multiple imputation and/or maximum likelihood methods will be employed to deal with missing data[37-41]. One of the limitations associated with a multivariable analytic approach lies in its generalisability to other data sets or clinical practices even with similar base-line risk. Cross validation techniques will be employed including bootstrapping to enhance generalisability and estimate shrinkage factors [42;43]. We will also compare results from the two centres. We aim to show that any differences in predictive features of the test results can be related to clinical/demographic differences in the population.

3.6 Power and Sample Size. The state of the art on sample size calculations provides less clarity on how these should be done in prognostic/diagnostic studies than in the case of intervention studies. There are a number of different approaches for diagnostic studies, but no single agreed formula, particularly as there are many designs possible. Even the STARD statement has deliberately left out this item from its checklist. We have adopted 2 broad approaches: the first is based on a widely recognised empirical rule; the second uses an explicit evaluation of the level of uncertainty in the probabilities of disease given by a logistic regression method.

The first approach is based on an 'events by variable rule'. Several such rules have been proposed, ranging from 5 to 1 for logistic regression analyses[44] to 10-25 to 1 for proportional hazards regression[45-48]. The 10 to 1 rule is a widely accepted standard in prognostic studies. In our diagnostic study, using logistic regression, a 5-10 to 1 rule, at the lower end of these recommendations, may be appropriate (see 3.5.3). The prevalence of disease in our study population will be increased (enriched) above the national background rate of undiagnosed cases in the community by three factors. Firstly, we have selected high risk populations. Secondly, the subjects have risk factors which have led to the test being ordered – we have seen that the prevalence of abnormal results exceeds those expected in unselected populations. Thirdly, they will have tested positive, e.g. those positive for AST should have increased risks of hepatocellular disease, while among those positive for ALP, an increase risk of

biliary or space-occupying lesions may be expected. Indeed, the study by Kim and colleagues[19] shows an exponential increase in risk of death from liver disease with increasing levels of AST. A sample of size 1,500 should provide an ample number of cases to analyse the common diagnoses NASH, ASH and No Disease. Our review of the literature suggests that the consensus committee (see section 4.9) may recommend an initial grouping of the remaining conditions into primary hepatocellular diseases (all Table 1 diseases except PBC) and biliary obstructive/infiltrative disease. These groups might account for 4% and 2% of our sample – i.e. 60 and 30 cases respectively in 1,500 – giving an ability to analyse 6-12 and 3-6 variables respectively. In fact, the prevalence of undiagnosed hepatitis C in the US is about 1.7%[49]. In England as a whole it is 0.4%[27], and in our study population it is likely to be intermediate; say 0.7%. Thus, it is possible that with four-fold enrichment we might find a 3% risk of hepatitis C alone, enough to support a separate analysis of this single disease. Other conditions for which a separate analysis might be feasible include hepatitis B and cancer. Our approach is to analyse individual diagnoses when they are sufficiently common for this to be feasible, though for rare conditions, such as primary biliary cirrhosis – about 0.25 in 100,000 – this will clearly not be possible.

For the second approach, suppose that a logistic regression model will be used to assess the diagnostic capacity of the LFTs in relation to a particular disease, or groups of disease, and that further testing will be indicated if an individual is perceived to carry a risk of the condition – as determined by the LFTs – greater than a pre-specified threshold level. Having specified the threshold risk, the sample size needed to construct the testing algorithm can be determined with reference to the performance of the algorithm at a second (higher) level of risk at which a high probability of 'follow-up' – i.e. further testing – is desired. The follow-up probability at the higher risk level is akin to a statistical power. For example, we may choose the sample size so as to be 80% certain that further testing will be triggered at this level.

Various choices for the threshold risk are possible. For example, it can be argued that this need be set no lower than the average population risk of the diagnosis over, say, one year. In any year, we estimate that between 1% and 2% of all GP-registered patients will meet the inclusion criteria for our study. Thus for each condition in Table 1, the population risk relevant to our study will be of the order of 1% or 2% of the proportion of cases in the study. In the following table the threshold is set at 2% of the average risk in the study, and the second risk-level, at which the power is computed, is taken to be twice the threshold risk. To make the calculations, a working value for the variation in risk explainable by LFTs was derived from data in Kim et al[19]. This amounts to an assumption about the predictive value of the LFTs. The details of the calculations are to be found in Girling[50].

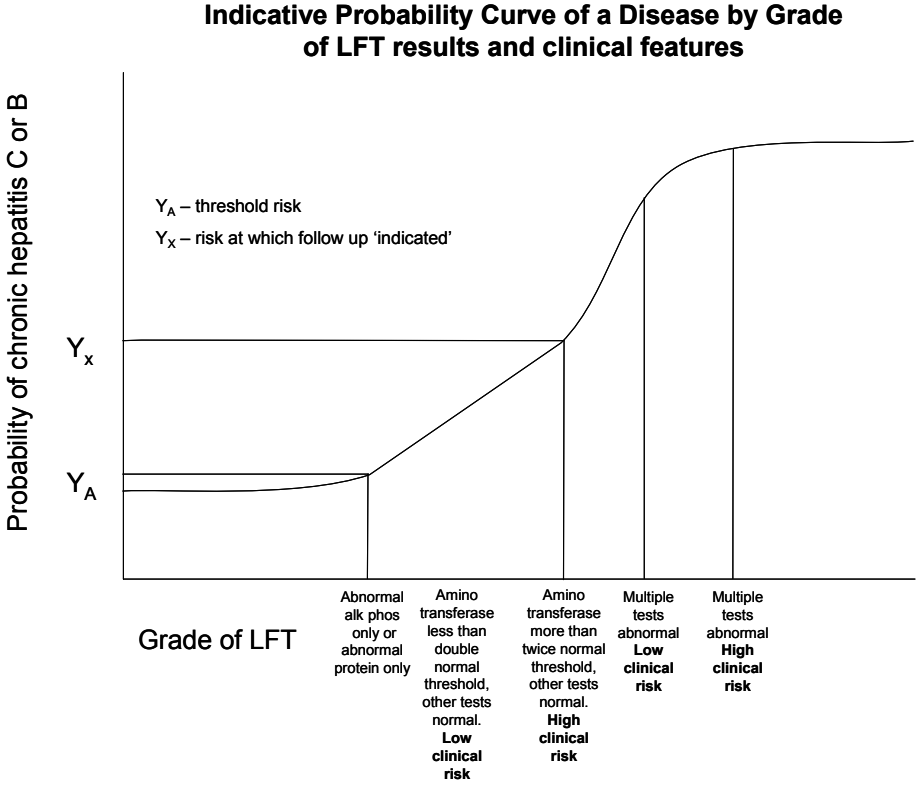
Table 3 suggests that 1,500 participants will be sufficient to construct testing algorithms with over 80% power for conditions which occur in 1.5% or more of our sample. For more prevalent conditions, with average sample risk of 4% or greater, 1500 subjects would be sufficient to achieve 90% power. These calculations are indicative only since the choice of the second risk-level is arbitrary. For any population risk, additional power is available if the second risk-level is set higher than twice the threshold risk.

Table 3: Statistical power at different sample sizes and average risk levels.

Average Risk in Sampled Population	Sample Size					
	1000	1500	2000	3000	4000	5000
0.001	0.57	0.59	0.60	0.62	0.64	0.66
0.005	0.66	0.69	0.71	0.76	0.79	0.81
0.010	0.71	0.75	0.79	0.83	0.87	0.89
0.015	0.75	0.80	0.83	0.88	0.91	0.93
0.02	0.78	0.83	0.86	0.91	0.94	0.96
0.03	0.82	0.87	0.90	0.95	0.97	0.98
0.04	0.85	0.90	0.93	0.97	0.98	0.99
0.05	0.88	0.92	0.95	0.98	0.99	0.995
0.1	0.94	0.97	0.98	0.996	0.999	0.9997
0.2	0.98	0.99	0.997	0.9997	0.99996	0.999995

For example, suppose that hepatitis C was diagnosed in 3% of our study sample. Then the threshold risk would be set at 0.06% (= 3% x 0.02) and the LFT criterion estimated from a sample of 1,500 would have an 87% chance of recommending follow-up for a patient whose true risk was 0.12%, or twice the threshold risk. This analysis can incorporate more than one LFT and additional clinical factors[50]. Note that we get a similar increase in precision by increasing from 1,000 to 2,000 participants, to that obtained by a further doubling to 4,000 participants, i.e. the cost per unit of precision has doubled. The concept of discriminating between probabilities of diseases by LFT abnormality and clinical features is shown diagrammatically in Figure 1.

Figure 1



3.7 The patients' perspectives

3.7.1 Study plan. Prior to the commencement of recruitment for the main study, we will conduct semi-structured pilot interviews on a sample of up to ten patients who have already received an abnormal LFT result from their doctor. Due to the prospective nature of the research, these patients will not be included in the main study. These interviews will enquire about participants' attitudes to testing, their interpretation of their results, and their plans to change behaviour along with perceptions of the barriers to doing so. Analysis of these interviews will inform the development of a structured questionnaire. At a later date, once recruitment is underway, further semi-structured interviews with up to 30 main study participants, sampled to reflect different patterns of results, will take place to further inform questionnaire development. The Patient Information Sheet and Consent Form for this pilot phase are included as Appendix 10.10.a and b.

We propose to measure patients' cognitions, emotions and behaviour using questionnaires at different stages of the investigation schedule (Table 4). Patients will complete the first questionnaire prior to the ultrasound appointment. They will then be asked (by post, phone or email: depending on their indicated preference) to complete a second questionnaire a week after this study appointment (when we are particularly keen to examine the impact of ultrasound scanning on participants) and then again 3 months post-appointment. On each occasion, a lack of response within 10 days will lead to a telephone reminder, and where possible an attempt to administer the questionnaire over the phone. At 2 years patient follow-up for repeat blood testing, a study researcher will administer the measure of alcohol consumption only.

The main cognitions assessed in the questionnaire will comprise tailored versions of those from the Illness Perception Questionnaire, based on Leventhal's self-regulation model of illness[51]. Items will explore perceptions relating to both the test result itself and perceived implications for patients' health. The main emotion assessed will be state anxiety[52]. The main behaviour assessed will be alcohol consumption which will be compared with data recorded by GPs (see 3.3.2). In addition, questions that emerge from analysis of the semi-structured interviews – such as the interpretation of results and planned behaviour changes, say in the response to ultrasound – will be incorporated into the structured questionnaire.

Table 4: Data collection on key psychological measures over study duration

Time point / Measures	T1 *(after LFT result, before scan appointment)*	T2 *(1 week after scan appointment*	T3 *(3 months after scan appointment*	T4 *(2 years after scan appointment)* Administered verbally
Behaviour				
- alcohol consumption	√	√	√	√
Cognition				
- perceptions of test results	√	√	√	√
- perceptions of health	√	√	√	√
Emotion				
- state anxiety	√	√	√	√

3.7.1.1 Psychology follow-up

All patients who are not under ongoing hospital supervision will be invited for 2-year follow up. They will be asked to complete a final questionnaire (T4). T4 questionnaires (Appendix 10.10.d)) will be posted to patients, in a pack, along with information concerning their study appointment from the study team. Patients will be asked to complete the questionnaire and to bring it with them to their study appointment, if they have not moved.

A further supply of T4 questionnaires will be available at GP surgeries so that patients who have not completed a questionnaire upon arrival at the clinic will be offered another copy and given a further chance to complete it, with assistance if available.

A record will be made by the study team of the following:

a) Patients invited back for a two-year follow-up appointment.
b) Patients sent a T4 prior to appointment.
c) T4 questionnaires:

- Completed prior to appointment
- Completed after appointment
- Reasons for non-completion.

3.7.2 Sample size for the questionnaire survey. State anxiety will be assessed using the short form of the Spielberger State Trait Anxiety Inventory[52]. It has a standard deviation 12 and the population mean value is 35. We would like to be able to detect a 0.3 SD difference in the change in anxiety before and after testing between people with abnormal and normal ultrasound results. This means we need the SD of the change in score over time, and we assume this to be 2. Assuming the baseline change in those with normal ultrasound results is 0.0, and that the proportion of all patients with normal and abnormal ultrasound results is the same, 233 would be needed with normal and abnormal ultrasounds at 90% power. Assuming a 30% questionnaire non-completion rate 334 are needed in each group. Even if ultrasound is abnormal in only 20% of people we will still have 89% power if we send the questionnaires to 1,000 people following completion of the qualitative phase. By asking this number of patients to complete questionnaires we shall have enough power to detect changes in anxiety, have a good chance of identifying other significant patient reactions to the testing process as well as having sufficient numbers for sub-group analyses of patient socio-demographic groups.

3.8 Health Economics. Our study will calibrate the risk of various diseases and groups of diseases according to clinical features and test results. Along with the costs of tests this will enable us to model the marginal losses of predictive accuracy from dropping tests from the comprehensive study protocol against the marginal savings from doing so. Failure to treat Table 1 diseases expeditiously is expensive involving treatment such as transplantation. The LFT panel is generally inexpensive. The private London Clinic price is £21 [53] and $23 from the US 'Health Test Direct'. The cost could be lower in a non-profit making organisation. Tests for Table 1 diagnoses are also not expensive. Ultrasound is more expensive and its role more ambiguous than the specific tests for Table 1 diseases. We will therefore measure the cost of ultrasound and add this to those of other tests, so that the marginal costs of ultrasound is explicit and can be offset against diagnostic discrimination and any psychological gain to patients (e.g. by 'newsworthiness', or apparent effect on motivation). The cost of ultrasound will be based on the level of staff typically deployed and the time taken for the ultrasound scan to be completed, which will be measured in a randomly selected subset of patients. However, bearing in mind that people at different levels of experience perform ultrasound, a sensitivity analysis will be carried out, assuming slightly different durations of testing among different grades of staff.

We will compute cost effectiveness of different testing strategies (from the measured positive predictive probabilities) and best estimate available of the natural history of the disease in its pre-clinical stage and from effectiveness of treatment at these early stages. However, we will need to include sensitivity analysis because these estimates are poorly calibrated in many cases. We will also model how these results vary by clinical and demographic features. A provocative possible finding is that there are groups where the prevalence in the tested

population of, say hepatitis C crosses the threshold where it would be cost-effective to include such a specific test in the baseline LFT panel.

3.9 Consensus Development. We will use formal consensus development processes for purposes, which we describe below on 2 separate occasions – at the start of the study and near the end (phase 1 and phase 2). Phase 1 consists of 3 components (described in sections 3.9.1 to 3.9.3) and will precede data collection and analysis and will inform study process, while phase 2 (described in section 3.9.4) will occur at the end of the study and will inform routine clinical practice.

3.9.1 Devising guidelines for the investigation of patients further to the standard diagnostic work-up described in section 3.3. GPs in the study will have clinical responsibility for their patients and will therefore need to respond to abnormal results. We will standardise this follow up as much as possible (see section 3.3.6) by developing consensus guidelines for further investigations.

3.9.2 Devising a general algorithm to determine how patients will be allocated to the final (reference standard) diagnostic groups at the end of the study. Again this must be harmonised across the study and made as transparent as possible.

3.9.3 Defining the hierarchy of variables to be used in analysis phase (as described in sections 3.5 and 3.6) and producing sensible groupings for rare diseases.

3.9.4 To develop some specimen algorithms on completion of the study (Phase 2).

The commissioning brief states that primary research is required to 'contribute over time to the development of a diagnostic algorithm'. There are three options:

1. No algorithm development work at the end of the study
2. To try to develop definitive algorithms
3. The middle ground of producing interim or indicative algorithms

We will not do (2) because there are many authoritative bodies, such as NICE, whose role this may usurp. On the other hand, we believe that some algorithm development is necessary to ensure that the statistical analyses produce the type of information needed to produce the algorithms which will enable GPs to give consistent rational care. These groups will also provide a convenient way to collect expert advice on the hierarchy of variables to use in logistic regressions as explained in section 3.5.2.

We have shown how we will produce primary data for this activity of three types:

a. Epidemiological data based on discriminant analysis to show how parsimonious testing strategies impact on diagnostic precision.
b. The understanding and preferences of patients which impact on welfare and behaviour.
c. The cost, particularly of imaging.

We have also shown from our literature review that may guidelines exist, but that they are not explicitly based on the above types of data. Hence they are often rather vague about what exactly a doctor should do in various situations.

3.9.5 Consensus development method. Formal consensus processes such as Delphi or nominal groups offer ways of synthesising judgements that are structured, transparent, offer the stimulus of feedback and give an explicit indication of the breadth of support for any conclusions. The nominal group technique (NGT), which usually involves about ten members and a meeting at which they can discuss and explore areas of disagreement. This reduces the risk of misunderstandings and exposes the reasons for differences of opinion[54]. The main alternative is the Delphi survey which involves two or more rounds of postal questionnaires. While this allows more geographically dispersed participants and avoids the risk of some individuals exercising undue influence[55], the opportunities for clarification and resolution of differences of opinion are more limited. A hybrid method (either called the modified nominal group, or the modified Delphi) used by RAND and others combines features of both processes, using a postal questionnaire for the first round of ratings followed by a meeting where the second round of ratings occurs[56]. An MRC funded four year research programme examining the methodological basis of formal consensus processes[57] has shown that although modified nominal groups produce closer consensus than Delphi groups, their judgements are less reliable. There was little to choose between the processes in terms of concordance of their judgements with the research evidence. The results suggest that the modified Delphi should be the preferred technique for producing algorithms for general use (phase 2) but that the NGT will be fit for the purpose of standardising study processes (phase 1).

A questionnaire will be developed for the consensus process based on data from the literature and from technical experts for phase 1 and including study data for phase 2. NGT (phase 1) participants will initially complete the questionnaire by post. They will then meet for a facilitated meeting which will follow a written protocol. At the meeting each participant will receive a new copy of the questionnaire with a reminder of their own initial ratings and the distribution of ratings for the group as a whole. Each item will be discussed in turn and reasons for any differences explored, after which participants will privately re-rate the questionnaire. Delphi group participants (phase 2) will comprise a large number of people who complete the entire process by postal questionnaire, but nominal groups will then meet to produce algorithms acceptable to more people, based on their results. The potential loss to follow-up will be minimised by telephoning repeat reminders. This method has been shown to result in response rates of over 90% (Raine, personal correspondence). Participants at the nominal group

meetings will be 2 patient representatives (one with an interest in general practice and the other with an interest in liver disease), and the clinical applicants. For the Delphi method (phase 2 only) a wider panel (of about 100 respondents) will be assembled from hepatologists, radiologists, chemical pathologists and general practitioners by sampling from the Royal College (specialist) and Department of Health (GPs) lists.

4. Project Timetable and Milestones

Target Date	Timeline
July 2005	Project start
December 2005	Recruitment starts
April 2006	
October 2006	
April 2007	
December 2007	Lambeth recruitment ends
February 2008	Follow-up begins
June 2008	Birmingham recruitment ends
June 2009	
October 2009	
November 2009	Lambeth follow-up ends
January 2010	
May 2010	Birmingham follow-up ends Statistical Analysis
June 2010	Draft Final Report Due

Milestone:	To be completed by:
Recruitment of study co-ordinators	July 2005
Purchase of portable ultrasound machines, recruitment of radiographers, completion of database, 1st consensus development meeting Recruitment to start	October 2005
500 patients recruited (6 months)	April 2006
1,500 patients recruited (18 months total recruitment time)	June 2008
Follow-up to begin	February 2008
500 patients followed up (6 months), outcome committee meetings start	April 2008
1,500 patients followed up (18 months total follow-up), statistician recruited, statistical analysis to begin, 2nd consensus development meeting	May2010
Draft report ready	June 2010

5. Add on Studies

5.1 Cryogenic Blood storage and Later Testing

We have found a high prevalence of fatty livers (40%) in our cohort of patients with mildly abnormal liver function tests in general practice and we wish to characterise this population biochemically. We have been successful in obtaining funding that will enable us to freeze the blood, taken from these patients. Patients without ultrasound evidence of fatty liver (60%) will act as controls. Patients will not need an additional venepuncture since we will freeze an aliquot of the blood taken under the existing protocol. Serum and cells will be stored separately at -70 and -20°C. The serum will be contained in 4 aliquots since it cannot be reused once thawed.

This study will be conducted at the Birmingham site only and we will recruit at least 1000 patients.

The blood components available for further testing are as follows:

1. Lipoprotein profile (the profile of 'good' and 'bad' fat in the blood)
2. C-reactive protein, procollagen peptide 3 and hyaluronate (markers of inflammation and of fibrosis)
3. Genes that may be associated with the development of liver disease

From those tests alone we can identify certain interesting hypotheses:

1. The profile of blood lipids is different and more abnormal in the ultrasound abnormal group.

2. The pattern of abnormal results is skewed towards abnormal triglycerides rather than cholesterol when compared with, say, a hypertensive population.

3. An abnormal liver ultrasound predicts abnormal liver profiles independently of BMI and alcohol intake (we have detailed alcohol histories from this well-characterised population).

4. Fat intake (obtained from dietary history) does/does not predict fatty liver (and its severity) net of LFT result.

5. The fasting glucose (samples are currently taken in this state) will/will not be higher in the abnormal group (we have not included HB1c because this requires collection of a separate sample and we are trying to make this study as simple as possible to minimise disincentive to participate).

6. Patients with high levels of serum markers for fibrosis are more likely to:
    a. have a higher degree of fatty infiltration in the liver as detected in ultrasound.
    b. will be more likely to progress in the proposed follow up study (for which funding will be sorted outside the HTA) to which patients have consented.

7. The rate of deterioration in liver appearance and function will be greatest when a fatty liver is associated with Hepatitis B and C.

Blood will be prepared and stored at the Institute for Biomedical Research, Medical School, University of Birmingham, Vincent Drive, Edgbaston, under the supervision of David Adams, Professor of Hepatology.

This add-on study will enable us to construct a unique cohort of immense scientific interest and practical importance.

5.2 A Qualitative Investigation into Liver Function Test Ordering Behaviour of General Practitioners Involved in the BALLETS Study

5.2.1 Aims

We intend to gain a greater understanding, of the type and range of non-clinical reasons and motives, behind the decision of a general practitioner (GP) to order a liver function test (LFT). A better understanding of these motives may lead to interventions that reduce the number of unnecessary tests.

5.2.2. Introduction

The UK national budget for pathology amounts to some £2.5 billion per annum and demand for diagnostic tests continues to rise rapidly. The number of tests requested has increased by over 10% per annum for the last three years alone. A large proportion of this increase has resulted from a growing demand for tests from general practitioners (GP). This was evident in the findings of a recent survey of NHS clinical biochemistry consultants who reported their workload from GPs had risen by over 80% since 2000.[1]

This increase in the number of ordered tests can be explained by a number of medical and diagnostic factors. A combination of these factors, including the growing range of available tests, guidelines that frequently promote the use of multiple-tests, financial incentives and our aging population, all go some way to explaining the increase. Certainly increased testing produces more false positive results, which in turn leads to knock-on investigations, which add further to the volume of tests ordered.

Aside from the considerations of the clinical management of a patient, research has shown that non-diagnostic motivations behind blood tests are commonly viewed as relevant by GPs, particularly when used to reassure the patient or doctor.[2] This has seen tests become ordered almost by rote by the GP, and led to an over-estimation of their diagnostic capability by patients. The net result is an over reliance on testing by both parties. Allied to this, testing due to the fear of litigation, known as defensive medicine, long a factor in the United States, is now also on the increase in the UK[3] aided by the ease with which tests can be ordered. All of which has impacted on the huge increase in testing witnessed over the last decade.

This study will be examining in more detail the non-diagnostic reasons behind the decision of a GP to order a test; whether it is based on maintaining the doctor-patient relationship, following the advice of colleagues, reducing the chance of unnecessary and expensive referrals or simply allaying their own fear of misdiagnosis.

A full list of references is shown in Appendix 10.12.c.

5.2.3 Objectives

1. To determine and assess the non-clinical reasons underlying a general practitioner's decision to order Liver Function Tests.
2. Use the information in conjunction with the findings of the primary BALLETS study to inform GP decision making and reduce the number of unnecessary tests.

5.2.4 Method

We will be using qualitative methods as we are working with a relatively small number of cases, so allowing a more detailed investigation. In addition we will adopt a constructionist approach[22] as we are primarily concerned with the behaviour of the General Practitioners and investigating what it is they do and their interaction with the patient.

A questionnaire has been produced that will be used to conduct a semi-structured interview which is flexible enough to allow new questions to be brought up during the interview as a result of what the interviewee says. We have a framework of themes that were identified from the existing literature concerning the ordering behaviour of GPs which have formed the rationale behind the key questions. These themes are described below.

5.2.4.1 Rationale

The primary subject areas under investigation are listed below. The question number in parenthesis refers to the questions in the semi-structured interview listed in Table 1.

Social Influence – Interaction between the GP and patients, colleagues and specialists

Quantitative studies have illustrated that GPs are more likely to test an individual dependent upon the patient's age and gender.[15,23] Anecdotal evidence has also reported that factors including racial, ethnic and socio-economic group can influence the decision to order. Examples have been reported where GPs are more likely to test if a patient is assertive and actively asks for a test. Similarly if a patient is concerned or worried then a GP will again order a test as a means of reassuring a patient. This is increasingly the case, as many patients see a blood test as the most reliable diagnostic tool at the GP's disposal.[10,24 25]

In addition previous interaction with the patient can impact on the GP's decision to order. Those patients that have a low medical consumption are more likely to be taken seriously than those who attend more frequently and are more consistently concerned about their health.[11] Anecdotal evidence has suggested tests can be used to incentivise a patient to improve their health. For example they can be used to impress on a patient that is drinking too much the damage that can be done to the liver. Further more, the long-term, ongoing nature of the relationship between a GP and their patient means that a GP is perhaps more aware of the consequences of not providing adequate care or of maintaining the GP/patient relationship than perhaps a consultant in a busy central hospital.[21]

Colleagues and specialists can also apply pressure on an individual GP to order a test. A specialist may recommend an individual to return to their GP and request a test as they have found no ill effects in the specific investigation. Colleagues may provide their own pressure to increase or reduce rates of test ordering following analysis of test ordering regimes and comparison with colleagues that shows them to test more, or less, frequently than their colleagues.[11]

Time and Cost

Previous studies have indicated that GPs have said that constraints on time within the consultation can lead to the decision to order a test. It has been reported that if they had more time during a consultation they may be able to take a full history of a patient reducing the need to order a blood test.[11] In addition, the lengthier period of time over which additional symptoms can manifest has meant some GPs have reported feeling under pressure to provide an answer before this point, again leading to the decision to order a test. Pressures of time have meant that GPs have also reported using the decision to order a test as a way of non-verbally communicating to a patient that the consultation period is over.[12]

The cost of a liver test panel is up to £2.69 for all eight tests. We also know that currently there is less financial pressure on investigation than prescribing and referral. We are interested to learn whether the cost of the test plays a role in determining whether the GP places an order, and if so to what extent it can influence that decision.

Defensive Medicine

It has been reported that some 98% of GPs are making changes to their practice because of the possibility of a patient complaining. Negative defensive practice includes prescription of unnecessary drugs, increases in follow-up referral rate and diagnostic testing. In addition certain treatments will be avoided and some practitioners have removed a patient with a problematic disorder from their list.[6]

Evidence has shown that the area in which a practice is based can impact on the prevalence of negative defensive decisions; for instance rural practices order fewer tests than those situated in urban areas[15], as they can spend longer with a patient, know them better and are more confident in their diagnosis. Anecdotal evidence also indicates that a practice whose demographic consists primarily of white-collar professional patients could lead to an increased fear of litigation than in a practice whose patients are predominantly blue-collar or of a racial or ethnic minority.

The GPs expectation of their efficacy

Previous experience of the GP plays a role in the decision making process. This can refer to first-hand experience or that of their colleagues. If they, or a colleague, have misdiagnosed an illness or condition, then they are far more likely to order a test when presented with a vague complaint. Similarly, after they or a colleague have reported uncovering a serious illness following a blood test, then the GP is more likely to order a test in future. Even when this is not the case they don't always have complete confidence in their ability to identify a condition using physical examinations and medical history. In addition, some GPs have said that their physical fitness affects their test ordering behaviour, and that they are more likely to order a test if tired or unwell, thereby reducing the length of the consultation. Also anecdotal evidence suggests that GPs are more likely to order an LFT if they have personal experience of a liver related complaint.

The process of ordering

It has been shown that the design of laboratory request forms can influence the decision to order a test[19] and changing the format and structure of request forms has been shown to reduce unwanted tests and requests.[20] Black takes the view that in relation to the increased testing in recent years the ease of access and the use of auto-analysers may be the most important factors in influencing the test ordering decision.[26] Research has shown that delaying the decision to order would ultimately reduce the number of unnecessary tests ordered. For example, ask a patient to report back in one week and if symptoms remain the same then a blood test will be ordered.[27]

A full list of references is shown in Appendix 10.12.c.

5.2.4.2 Study population (participating GPs)

The aim is to elicit as much information on relevant factors as possible from our study group therefore the 60 GPs at the eight participating practices will be asked to contribute. The GPs at each of these practices will be approached by the practice manager and/or the author, supplied with copies of the information sheet and asked to participate in the study. For a list of the practices involved please see Appendix 10.12.d, a copy of the information sheet is contained in Appendix 10.12.b. All GPs that agree to join the study must complete a consent form, contained in Appendix 10.12.a.

5.2.4.3 Interviews

The interviews will be semi-structured and will be conducted over the telephone or in person, at the GPs discretion. Previous studies of GPs utilising postal questionnaires had a poor response rate.[28] In all cases the interview will be preceded by a brief description of the study and the context of the interview. Please see Table 5 for a full list of the questions.

Table 5 Questions for semi-structured interview

1) Do you work full-time or part-time? If part time then what %age of full time (50%, 25% etc)
2) What are your thoughts on the use of liver function tests in general?
3) Do you believe that on the whole GPs order the right amount of LFTs?
4) Other than the generally recognised medical symptoms that might indicate liver disease are there any other signs or symptoms that could lead you to order a liver function test.
5) Would you agree with those GPs that say a blood test (LFT) is a way of signalling to the patient that you are taking a complaint seriously?
6) What role do you think defensive medicine plays in the decision to order an LFT?
7) Is there anything else you would like to say on the subject of LFTs?

5.2.4.4Analysis

The interviews will be either audio-taped or digitally recorded dependent upon whether it is conducted in person or by telephone. They will then be transcribed and analysed using computer-assisted analysis of qualitative data (CAQDAS) utilising nVIVO software. This will allow us to handle potentially large volumes of data, demonstrate the rigour of the analysis, including the production of counts of phenomena and aid in the searching of deviant cases. If necessary it can also help with the development of consistent coding schemes. We will combine information from all interviews and analyse according to the best principles of conversation analysis by

identifying sequences of related talk, examine how speakers take on certain roles and finally, look for particular outcomes in the talk.

Throughout the analysis we will seek to maintain validity by ensuring that the account accurately represents the phenomena we are referring to and secondly to ensure reliability, reflecting the degree of consistency with which instances are assigned to the same category, either by different observers or by the same observer on different occasions.

5.3 Follow-up of Abnormal Test Results

In the course of the study, a number of people tested positive for some specific liver diseases. All patients who have tested positive for Hepatitis B and C were referred for specific treatment.

Table 6 Specific diseases tested for as part of the BALLETS screening algorithm.

Test	Disease
Antimitochondrial Ab	Primary Biliary Cirrhosis (PBC)
Caeruloplasmin	Wilson's disease - copper excess in tissue
Iron & transferrin saturation	Haemochromatosis – iron accumulation in tissue
Alpha 1-antitrypsin	Alpha 1- antitrypsin deficiency – liver/lung damage
Smooth muscle Ab	Auto-immune hepatitis

During the course of our follow-up it has come to light that the prescribed follow-up, based on the algorithm shown in Appendix 10.9, has not been carried out for some patients testing positive for the conditions in Table 6.

This was undesirable from both the clinical and the scientific point of view and this sub-protocol describes the steps that will be taken towards remedying this situation.

5.3.1 Follow-up of abnormal results from the first BALLETS testing process

5.3.1.1 Primary biliary cirrhosis

Thirteen patients had a positive mitochondrial antibody test with a titre>1:40. We now know two things from the recent literature. Firstly, the majority of these patients will be true positives – they will have primary biliary cirrhosis. However, the course of the disease will be mild in the majority of cases. In these milder cases, the liver disease usually progresses at such a slow rate, that the person dies of independent causes without the liver disease causing troublesome symptoms. In other words many people have the disease but it follows a benign course, and they may never become symptomatic.

The more aggressive cases manifest earlier in life. If a patient has reached the age of, say, 70 and is asymptomatic and does not have a clinical abnormality, then we feel that no further follow-up is usually necessary. Two of the 9 cases fall into this category.

However, we think that the patient's general practitioner should be informed of the likelihood of primary biliary cirrhosis in the remaining cases (n=9) unless a referral has already been made. It should be noted that treatment with the bile acid, ursodeoxycholic acid, is licensed for this indication and may slow disease progression.

A copy of a letter to the general practitioner from the study hepatologist is attached to this document as Appendix 10.13.a.

5.3.1.2 Wilson's disease

This is a disease where the liver is affected by the accumulation of copper in the body. The copper accumulates because the protein in the blood which binds copper is insufficient. This, in turn, is the result of a genetic condition (autosomal recessive). The diagnosis is often difficult to make and a useful screening test is to measure the protein caeruloplasmin in the blood but the specificity of this test is low, i.e. a high proportion of positives are 'false positives'. However, the diagnosis is important since early treatment may prevent development of life-threatening complications and, where appropriate, family members who are asymptomatic should be offered screening.

There are ten patients who cross the recently defined threshold of abnormality. However, Wilson's disease is so rare, that the probability that the disease exists is low even after an abnormal test result; even the post-test probability of the disease is low. We propose to follow-up only those patients who have not already been referred and who are under 55 years of age and/or who have ultrasound signs of impending cirrhosis. The rationale for this lies partly in the low post-test probability of this disease and partly in the knowledge that Wilson's disease usually manifests before the 6th decade of life. Moreover, with one exception, the extent of the test abnormality is very mild in the BALLETS cases. The exception is a young person, and she is one of two cases that fulfil our criteria. Both these patients have moved out of the area but we intend to trace the patients (through NHS numbers) and inform their general practitioner about the possibility of Wilson's disease using a letter (Appendix 10.13.b – see attached).

5.3.1.3 Iron storage disease – haemochromatosis

This is also a metal storage disease; in this case caused by excessive absorption of iron and again it has a genetic basis. Unlike Wilson's disease, haemochromatosis is quite common. However, like primary biliary cirrhosis, it often follows a very mild clinical course and never comes to light. That said it can also cause cirrhosis and liver cell cancer, as well as affecting other organs such as heart, brain and pancreas. All these sequelae can be prevented if the condition is diagnosed and treated early. Moreover, the combination of haemochromatosis and moderate alcohol use is particularly noxious for the liver.

One of our difficulties is that a large number of patients cross the laboratory threshold of abnormality (an iron saturation of more than 50%). In fact, 27 patients are classified as abnormal on this basis. Only a small minority of these are likely to have the disease – perhaps one or two.

We therefore think it is necessary to be more discriminating, and it is important to understand that the laboratory definition of abnormality is based on a statistical threshold, not on a cost-benefit analysis of the trade-off between sensitivity and specificity. The recent literature shows that there is a very low probability of the disease unless the iron saturation exceeds 80%. We therefore propose to alert the general practitioner to the need for follow-up only in those cases that cross this threshold – there are 7 such cases.

We propose to send a letter (Appendix 10.13.c) to the GPs from these cases.

5.3.1.4 Alpha 1-Antitrypsin deficiency

We do have the necessary further tests in the great majority of cases that screened positive for this disease. However, in none of these did further testing (so-called "phenotype testing") confirm the presence of the disease. However, in a small number of cases (ten), this result is missing.

We propose offering this test to these ten patients when they come for their second liver function test visit.

We think that it is very unlikely that any will test positive (this is also an extremely rare disease) but we thought we should complete testing in these cases.

5.3.1.5 Autoimmune hepatitis

Four patients screened positive for smooth muscle antibodies – the screening test for this extremely rare disease. However, this test, unlike the anti-mitochondrial antibodies is not specific. One of the four patients has a significantly raised AST and this patient also has abnormal liver texture on ultrasound. We propose to follow this patient up using a letter (Appendix 10.13.d). In our opinion no further action is indicated in the three remaining cases.

5.4 A Qualitative Investigation into Patients' Experience of Taking Part in the BALLETS Study and the Impact of the Finding of a Fatty Liver on Ultrasound

During the recruitment phase 40% of study patients were found to have fatty liver on ultrasound. The recognised primary treatments for fatty liver are diet and regular exercise.

The literature on the subject suggests that the "working alliance" between care-provider and patient is important in adopting health behaviours. This alliance is defined by the mutual agreement on goals and objectives and the extent of the emotional bond (liking or trust) between patient and provider. In addition, care-providers use a number of verbal compliance strategies where the subtle use of language can help influence a patient's behaviour. Earlier this year evidence emerged that moderate and low-level lifestyle counselling interventions in patients with fatty livers are a practical and effective method for improving health.

Birmingham research nurses have had many positive anecdotal reports from patients returning to follow-up clinics, regarding improved drinking, eating and exercise habits, following the initial battery of BALLETS tests and especially if the scan was abnormal.

These results are supported by analysis of the results from the first 277 patients who have had follow up scans – Table 7. From the table it can be seen that a higher proportion of patients had lost weight, when the scan showed fatty liver and when it did not. This supports the idea that an abnormal scan may be particularly motivating to patients. This is further supported by the results of the second ultrasound scan where an abnormal scan reverted to normal in 22 cases while a normal scan became abnormal in only 8% of cases.

Table 7. Patients with and without fatty liver and weight loss at the first (baseline) and second (2 years) ultrasound scan.

	Total	Weight loss
Fatty liver T1	114 (41.1%)	56 (49.1%)
Non fatty T1	163 (58.8%)	65 (39.8%)
Fatty liver T2	85 (30.6%)	31 (36.4%)
Non fatty T2	192 (69.3%)	93 (48.4%)

5.4.1 Methodology

We would like to conduct a qualitative investigation, further exploring the anecdotal and preliminary evidence that events associated with participation in the BALLETS study were motivational. In particular we will investigate the finding of a fatty liver on the decision making process. This will enable a better understanding of successful aspects of initial encounters with the health service. In particular, we are interested in the patient's perception of the results of the initial scan, how the information was imparted and the factors that led to their adopting or failing to adopt, lifestyle changes beneficial to health.

To achieve this we will conduct a series of semi-structured interviews with patients whose fatty liver 'improved', as well as those patients whose liver showed no improvement, allowing some comparison between groups. In addition, we will interview all Birmingham sonographers to determine their opinions on the consultation process and the methods they used to impart the result of the scan and its possible implications.

5.4.1.1 Participants

To meet the aim of our qualitative sub-study, we propose to interview 40 patients who participated in the first and follow-up study clinics. As explained above, the particular groups of interest are people who have made lifestyle changes versus people who have not and people who had an initial abnormal scan versus people who have not. Table 8 illustrates the recruitment figures.

Table 8. Recruitment figures for Qualitative sub-study of BALLETS

	Fatty Liver	Non-fatty Liver
BMI unchanged	10	10
BMI reduced by ≥5%	10	10

Although we have imposed inclusion criteria for this study, the 40 patients will be selected at random within each of the four categories using their BALLETS study participant ID number. This will be done across all GP practices that took part in the study. It is estimated that not all people will agree to take part in this sub-study and therefore, other patients will be randomly selected until the final sample of 40 is reached.

All Birmingham sonographers will also be invited to take part and will be interviewed to determine their opinion on the consultation process, the methods they used to impart the results of the scan, and possible implications.

5.4.1.2 Process

Once eligible patients have been identified, they will be contacted by telephone to determine initial interest in taking part in the sub-study. Patients will receive an initial telephone call rather than a letter as anecdotal evidence reported to the research nurses conducting follow-up clinics indicates that patients prefer direct telephone contact from the researchers to discuss the study process. Once a patient has expressed an initial interest, information sheets (Appendix 10.14.a) will be sent. Another follow-up phone call will be made after 3 days to check that the information has been received and that the patient is willing to attend. An appointment time will then be offered and the patient will be able to choose if they would like to be seen at their home, at their GP practice, or by telephone (after receiving the consent form by post).

During the main visit, patients will be asked to provide informed consent (Appendix 10.14.b). The actual interview will be semi-structured in nature, with patients being asked specific questions about their experience of taking part in the BALLETS study. Although there is some structure to the interview to enable smooth flow of conversation (see semi-structured interview questions - Appendix 14c) the interview will allow for flexibility to ensure the patient's point of view is heard. The interview will be 30-60 minutes long. These interviews are audio-recorded with the permission of the patient.

A similar process will be employed for approaching and interviewing the sonographers (Appendix 10.14.d, e, f)

5.4.2 Data analysis

Once all the interviews are complete, they will be transcribed. Audio-recordings will be kept in a locked cabinet at the BALLETS study office. The transcripts will be anonymised and reference to anything that may identify a particular patient will be altered. Each transcript will be analysed using a qualitative data analysis method of interpretative analysis (Smith, 1995, 1996, 1999; Smith and Osborn, 2003) as an attempt to unravel the meanings contained in the transcripts (Smith, 1996). This method recognises that the meanings people ascribe to events are the product of interactions between people in the social world (Willig, 2001). In the analysis we will explore the participants' view of the world and adopt, as far as possible, an "insider-perspective" (Conrad, 1987) of the

phenomenon under study (Smith, 1996), which fits well with our research question, "What is the patient's experience of participating in the BALLETS study?" The method acknowledges dependence on the researcher's own view-point (Smith, 1995) and is in accordance with Elliott, Fischer and Rennie (1999) guidelines for good qualitative research where owning up to one's perspective and assumptions helps readers to interpret and understand the researcher's data. Hence, this sub-study follows existing guidelines for conducting good qualitative research (Parker, 2004).

6. Contribution: Expertise

Doug Altman	Oversee statistical analysis: Medical statistician with interest in diagnostic and predictive testing
David Armstrong	Deputy Chief Investigator (London): Medical sociologist with interest in sociological dimensions of disease labels
Bob Cramb	Laboratory tests (Birmingham): Chemical pathologist and lipidologist
Paramjit Gill	Oversee Primary care (Birmingham):Academic general practitioner with interest in ethnicity and migration
Alan Girling	Statistical analysis, mathematical modelling: mathematician with interest in modelling
Judith Harris	Initial organisation and co-ordination between Birmingham and London: Research fellow
Roger Jones	Oversee Primary Care (London): Academic GP with interest in gastroenterology/hepatology
Khalid Khan	Evaluation of diagnostic accuracy: Production of clinical algorithms for diagnostic test: Gynaecologist with interest (HTA supported) in diagnostic testing.
Richard Lilford	Chief Investigator (Birmingham): Clinical Epidemiologist with interest in modelling
Theresa Marteau	Psychology – administration of questionnaires, interpretation : Health Psychologist interested in responses to risk information
Karel Moons	Outcome committees. Expert in diagnostic evaluation – supported by Dutch Govt
James Neuberger	Hepatologist (Birmingham) supervise liver services: Hepatologist Birmingham
Simon Olliff	Supervisor in Quality Control, Ultrasound Birmingham: Consultant Radiologist
James Raftery	Health economic analysis: Economist with wide experience of cost effectiveness
Rosalind Raine	Facilitate consensus development meetings, designing questionnaires and advice on analysis
Giles Rottenberg	Ultrasound (London): Radiologist
Ramasamyiyer Swaminathan	Laboratory tests (London): Chemical Pathologist
Mark Wilkinson	Supervise liver services: Hepatologist (London)

We have also appointed a steering committee, which will consist of Prof Richard Lilford, Dr David Armstrong, Prof Doug Altman, Ms Pat Moseley (consumer) and Dr Alison Rogers (consumer). The committee will have an independent chair – Prof Andy Haines Dean of (LSHTM).

7. Consumers

We have explained in some detail our methods for including the consumer perspective in our study, with particular reference to patients being tested under advice from the general practitioner. Also, we have included the consumer perspective, in our algorithm development work, where consumers will contribute to the nominal group technique. We also wish to include the consumer perspective in the overall governance of the programme. To do this, we will ensure that the steering committee has two consumers. One consumer is selected from a society dealing with liver disease specifically – Alison Rogers (British Liver Trust). However, many of these patients will not have liver disease and therefore we are also to include a consumer with a particular perspective and interest in general practice (Ms Pat Moseley).

8. Justification of support required

Staff. Relatively senior co-ordinators are required on each site, to oversee the study. They will be responsible for ensuring that GP's who wish to participate are maximally empowered to do so. They will be responsible for making sure that patients are offered appointments and for ensuring that data forms/questionnaires are all retrieved so data entry can take place here in Birmingham (it is because of this increased role of the Birmingham centre that we have requested a slightly more senior co-ordinator than London). They will also be responsible for liaising with the ultrasonographers and for arranging for long-term follow-up. They will be most heavily involved in the start-up period (six months), the 18 months of active recruitment and for a further six months as patients complete the initial assessments. Co-ordination will need to continue for the subsequent two years of the study, to ensure that the long term follow-up takes place, that meetings take place, that the materials are collected and meetings take place for the reference standard diagnosis and to oversee and co-ordinate further data entry. It is also necessary to allow for programming time, to train the database to receive the data and to hire a half time data entry clerk to enter the information.

Support is required for a half time psychologist for four years to conduct the interviews necessary to help develop and pilot design the questionnaire and to organise the collection of data on three occasions for 1,500 participants. They will also conduct the analysis of the data.

The statistical and modelling components of this study are considerable and we therefore request a half time senior research fellow in statistics for the last two years of the study. We also require two sessions of health economics time for the final year of the study. The psychologist should be able to contribute to the algorithm production.

Travel/meetings. This proposal involves a number of meetings (e.g. consensus development, outcome committees), and these have been costed appropriately. We will pay our consumer representatives the standard rate for attending such meetings (£138 per day).

Portable Ultrasound. We have been advised that it would be cheaper to purchase portable ultrasound machines than lease them for a period of 18 months. The costs quoted are for purchase, plus a service contract.

NHS Service Support Costs 'Support for Science' NHS R & D Support Funding will be crucial to enable the general practices to host this study. This will be sought and administered by the 2 primary care research networks involved, (MidRec and STARnet), working with the lead primary care investigators (PG and RG). Calculations are based on current rates applied to existing studies in the Midlands, and cover GP *time* (informed consent, completion of test request form and short questionnaire) and *administration time* (checking and retrieving test results/records) TOTAL =£26.98 per patient.

Excess treatment costs. Extra blood tests and ultrasound sessions to ensure comprehensive package of testing

We have discussed these costs with Stella Barclay (NHS R&D Finance).

9. Reference List

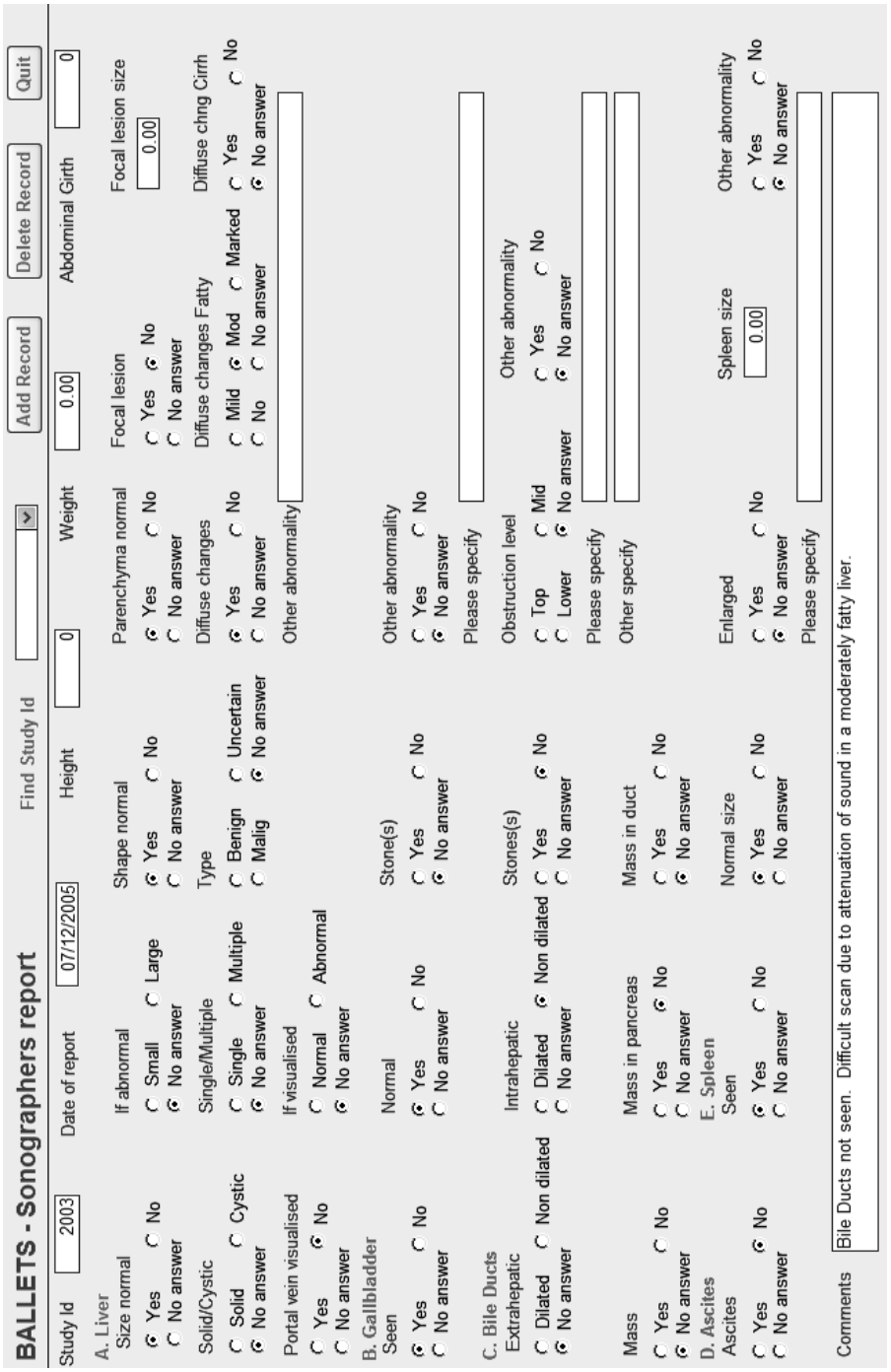
1. Sherwood P, Lyburn I, Brown S, Ryder S. How are abnormal results for liver function tests dealt with in primary care? Audit of yield and impact. BMJ 2001;322:276-8.
2. Witte K,.Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ.Behav. 2000;27:591-615.
3. Petrie KJ, Cameron LD, Ellis CJ, Buick D, Weinman J. Changing illness perceptions after myocardial infarction: an early intervention randomized controlled trial. Psychosom.Med. 2002;64:580-6.
4. Milne S, Sheeran P, Orbell S. Prediction and intervention in health related behaviour: a meta-analytic review of protection motivation theory. Journal of Applied Social Psychology 2000;30:106-43.
5. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. CMAJ. 1986;134:587-94.
6. Fryback DG,.Thornbury JR. The efficacy of diagnostic imaging. Med.Decis.Making 1991;11:88-94.
7. The STARD Group. The STARD Initiative -- Towards Complete and Accurate Reporting of Studies on Diagnostic Accuracy . http://www.consort-statement.org/stardstatement.htm . 2003.
8. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC.Med.Res.Methodol. 2003;3:25.
9. Dufour DR, Lott JA, Nolte FS, Gretch DR, Koff RS, Seeff LB. Diagnosis and monitoring of hepatic injury. I. Performance characteristics of laboratory tests. Clin.Chem. 2000;46:2027-49.
10. Dufour DR, Lott JA, Nolte FS, Gretch DR, Koff RS, Seeff LB. Diagnosis and monitoring of hepatic injury. II. Recommendations for use of laboratory tests in screening, diagnosis, and monitoring. Clin.Chem. 2000;46:2050-68.
11. Bonacini M. The value of hepatic iron index in cirrhosis. Gastroenterology 1997;113:1052-3.

12. Whitehead MW, Hawkes ND, Hainsworth I, Kingham JG. A prospective study of the causes of notably raised aspartate aminotransferase of liver origin. Gut 1999;45:129-33.
13. Hultcrantz R, Glaumann H, Lindberg G, Nilsson LH. Liver investigation in 149 asymptomatic patients with moderately elevated activities of serum aminotransferases. Scand.J.Gastroenterol. 1986;21:109-13.
14. Hay JE, Czaja AJ, Rakela J, Ludwig J. The nature of unexplained chronic aminotransferase elevations of a mild to moderate degree in asymptomatic patients. Hepatology 1989;9:193-7.
15. Mathiesen UL, Franzen LE, Fryden A, Foberg U, Bodemar G. The clinical significance of slightly to moderately increased liver transaminase values in asymptomatic patients. Scand.J.Gastroenterol. 1999;34:85-91.
16. Daniel S, Ben Menachem T, Vasudevan G, Ma CK, Blumenkehl M. Prospective evaluation of unexplained chronic liver transaminase abnormalities in asymptomatic and symptomatic patients. Am.J.Gastroenterol. 1999;94:3010-4.
17. Skelly MM, James PD, Ryder SD. Findings on liver biopsy to investigate abnormal liver function tests in the absence of diagnostic serology. J.Hepatol. 2001;35:195-9.
18. Sorbi D, McGill DB, Thistle JL, Therneau TM, Henry J, Lindor KD. An assessment of the role of liver biopsies in asymptomatic patients with chronic liver test abnormalities. Am.J.Gastroenterol. 2000;95:3206-10.
19. Kim HC, Nam CM, Jee SH, Han KH, Oh DK, Suh I. Normal serum aminotransferase concentration and risk of mortality from liver diseases: prospective cohort study. BMJ 2004;328:983.
20. Green RM,.Flamm S. AGA technical review on the evaluation of liver chemistry tests. Gastroenterology 2002;123:1367-84.
21. Rochling FA. Evaluation of abnormal liver tests. Clin.Cornerstone. 2001;3:1-12.
22. Chatwin T. Diagnosing liver disease in asymptomatic patients. JAAPA. 2001;14:39-47.
23. Pratt DS,.Kaplan MM. Evaluation of abnormal liver-enzyme results in asymptomatic patients. N.Engl.J.Med. 2000;342:1266-71.
24. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J et al. Consensus development methods, and their use in clinical guideline development. Health Technol.Assess. 1998;2:i-88.
25. Younossi ZM. Evaluating asymptomatic patients with mildly elevated liver enzymes. Cleve.Clin.J.Med. 1998;65:150-8.
26. Moseley RH. Evaluation of abnormal liver function tests. Med.Clin.North Am. 1996;80:887-906.
27. Health Protection Agency. Hepatitis C Information. http://www.hpa.org.uk/infections/topics_az/hepatitis_c/phlsgen_info.htm . 30-4-2003.
28. Health Protection Agency. Hepatitis B Information. http://www.hpa.org.uk/infections/topics_az/hepatitis_b/gen_info.htm . 30-4-2003.
29. Hemochromatosis. http://www.book-spot.co.uk/index.php/Hemochromatosis . 2004.
30. Primary Biliary Cirrhosis Foundation. Primary Biliary Cirrhosis. http://www.pbcfoundation.org.uk/ . 2004.
31. Autoimmune Hepatitis. http://www.autoimmunehepatitis.co.uk/ . 2004.

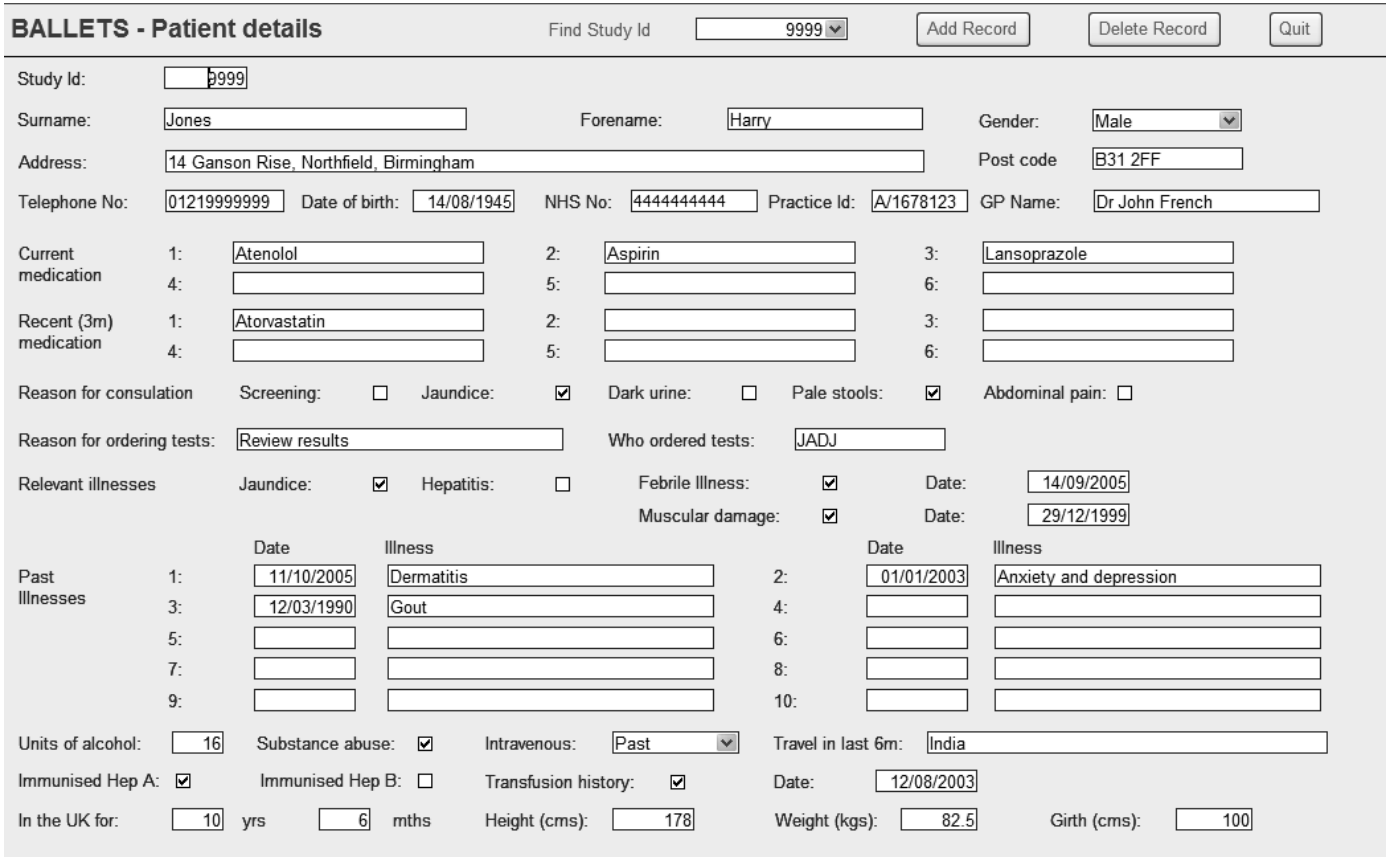
32. de Serres FJ. Alpha-1 antitrypsin deficiency is not a rare disease but a disease that is rarely diagnosed. Environ.Health Perspect. 2003;111:1851-4.
33. The British Liver Trust. Gilbert's Syndrome. http://www.britishlivertrust.org.uk/content/diseases/gilberts_syndrome.asp . 2004.
34. Jones D,.Gill P. Breaking down language barriers. The NHS needs to provide accessible interpreting services for all. BMJ 1998;316:1476.
35. Gill P, Kai J, Bhopal R, Wild S. Health Care Needs Assessment: Black and Minority Ethnic Groups. In Raftery J, Stevens A, Mant J, eds. Health Care Needs Assessment. The epidemiologically based needs assessment reviews., Abingdon: Radcliffe Medical Press Ltd., 2004.
36. Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests. http://www.fda.gov/cdrh/osb/guidance/1428.pdf . 2004. US Food and Drug Administration.
37. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol.Methods 2001;6:330-51.
38. Enders CK. The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. Psychol.Methods 2001;6:352-70.
39. Raghunathan TE. What do we do with missing data? Some options for analysis of incomplete data. Annu.Rev.Public Health 2004;25:99-117.
40. Schafer JL,.Graham JW. Missing data: our view of the state of the art. Psychol.Methods 2002;7:147-77.
41. Sinharay S, Stern HS, Russell D. The use of multiple imputation for the analysis of missing data. Psychol.Methods 2001;6:317-29.
42. Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. Stat.Med. 1984;3:143-52.
43. Steyerberg EW, Eijkemans MJ, Van Houwelingen JC, Lee KL, Habbema JD. Prognostic models based on literature and individual patient data in logistic regression analysis. Stat.Med. 2000;19:141-60.
44. Freedman L, Pee D, Midthune D. The problem of underestimating the residual error variance in forward stepwise regression. The Statistician 1992;41:405-12.
45. Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. Cancer Treat.Rep. 1985;69:1071-7.
46. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J.Clin.Epidemiol. 1995;48:1503-10.
47. Feinstein AR. Multivariable Analysis: an Introduction. New Haven: Yale University Press, 1996.
48. Schumacher M, Hollander N, Schwarzer G, Sauerbrei W. Prognostic factor studies. In Crowley J, ed. Handbook of Statistics in Clinical Oncology, pp 321-78. New York: Marcel Dekker, 1999.
49. Moyer LA, Mast EE, Alter MJ. Hepatitis C: Part I. Routine serologic testing and diagnosis. Am.Fam.Physician 1999;59:79-2.
50. Girling, A. Power Calculations for Screening Tests. http://www.publichealth.bham.ac.uk/nccrm/ballets.htm . 2004.

51. Moss-Morris R,.Chalder T. Illness perceptions and levels of disability in patients with chronic fatigue syndrome and rheumatoid arthritis. J.Psychosom.Res. 2003;55:305-8.
52. Marteau TM,.Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br.J.Clin.Psychol. 1992;31 (Pt 3):301-6.
53. London Clinic. www.biolab.co.uk/otherlab.html . 2004.
54. Delbecq AL,.Van de Ven AH. A group process model for problem identification and program planning. J Appl Behav Sci 1971;7:466-92.
55. Bayley EW, Richmond T, Noroian EL, Allen LR. A Delphi study on research priorities for trauma nursing. Am.J.Crit Care 1994;3:208-16.
56. Brook RH. The RAND/UCLA appropriateness method. In McCormick KA, Moore SR, Seigel RA, eds. Methodology Perspectives, pp 59-70. Rockville, MD: Public Health Service, US Dept of Health and Human Services, 1994.
57. Raine, R, Sanderson, C, Hutchings, A, Carter, S, Larkin, K, and Black, N. An experimental study of determinant of group judgement in clinical guideline development. The Lancet 2004.

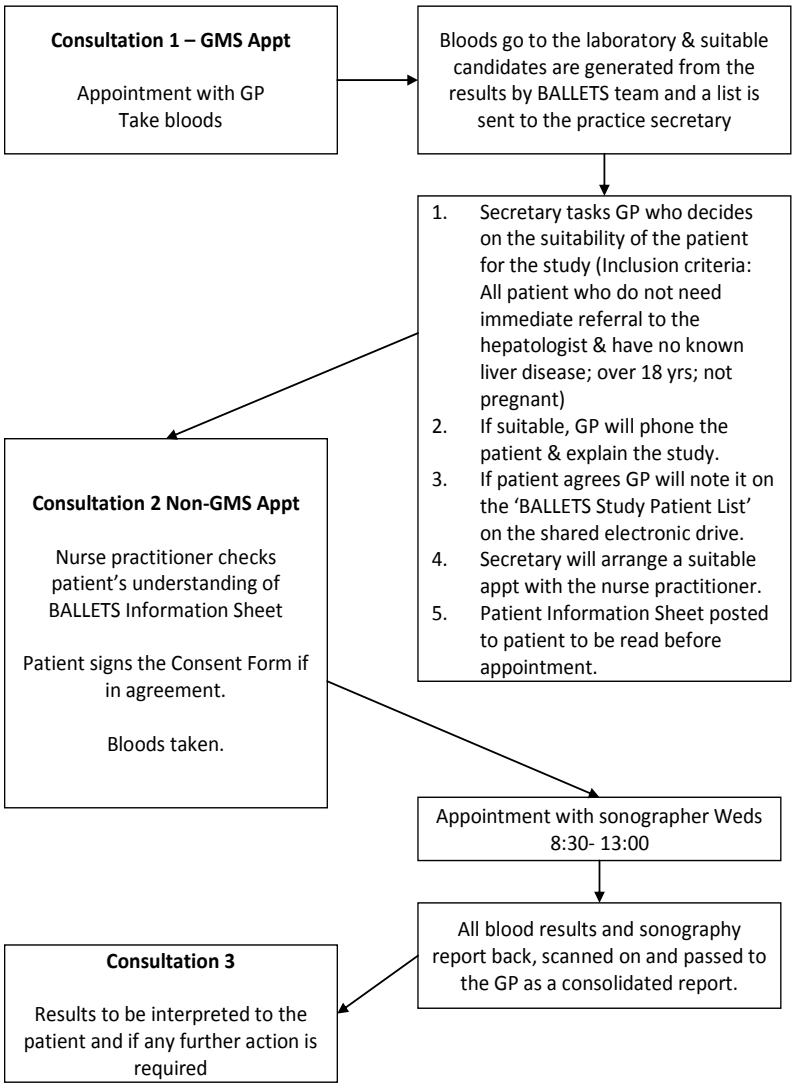
Appendix 10.1.a



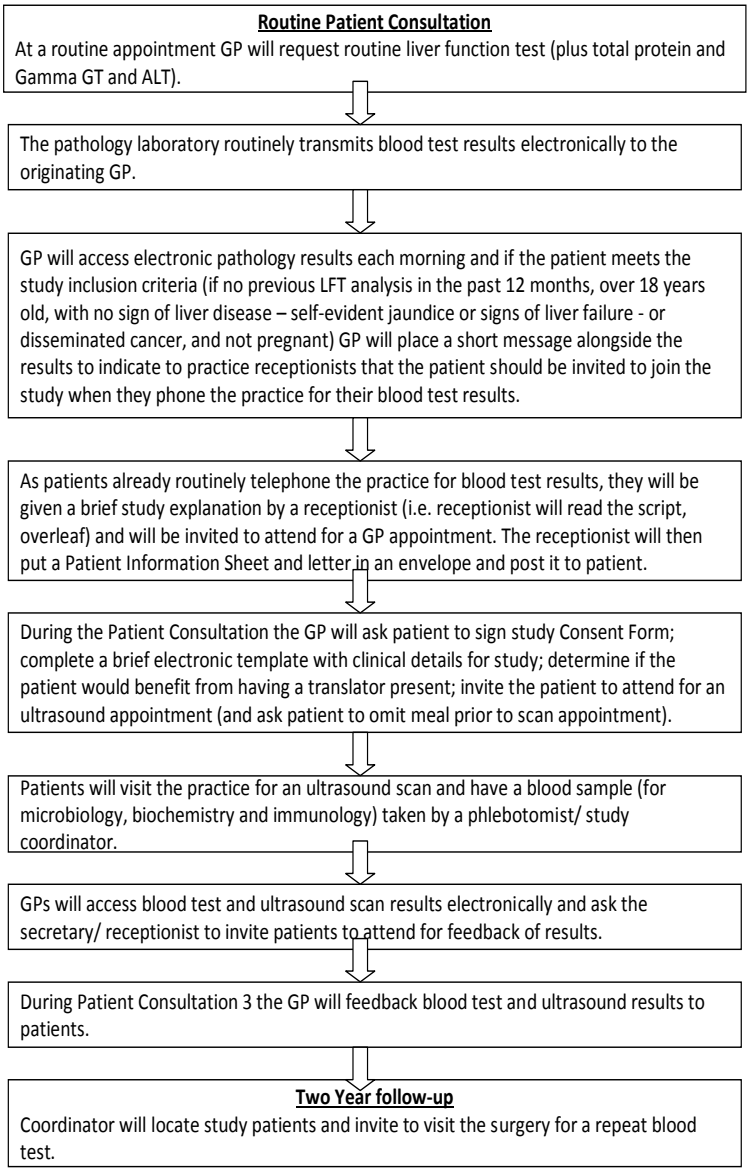
Appendix 10.1.b



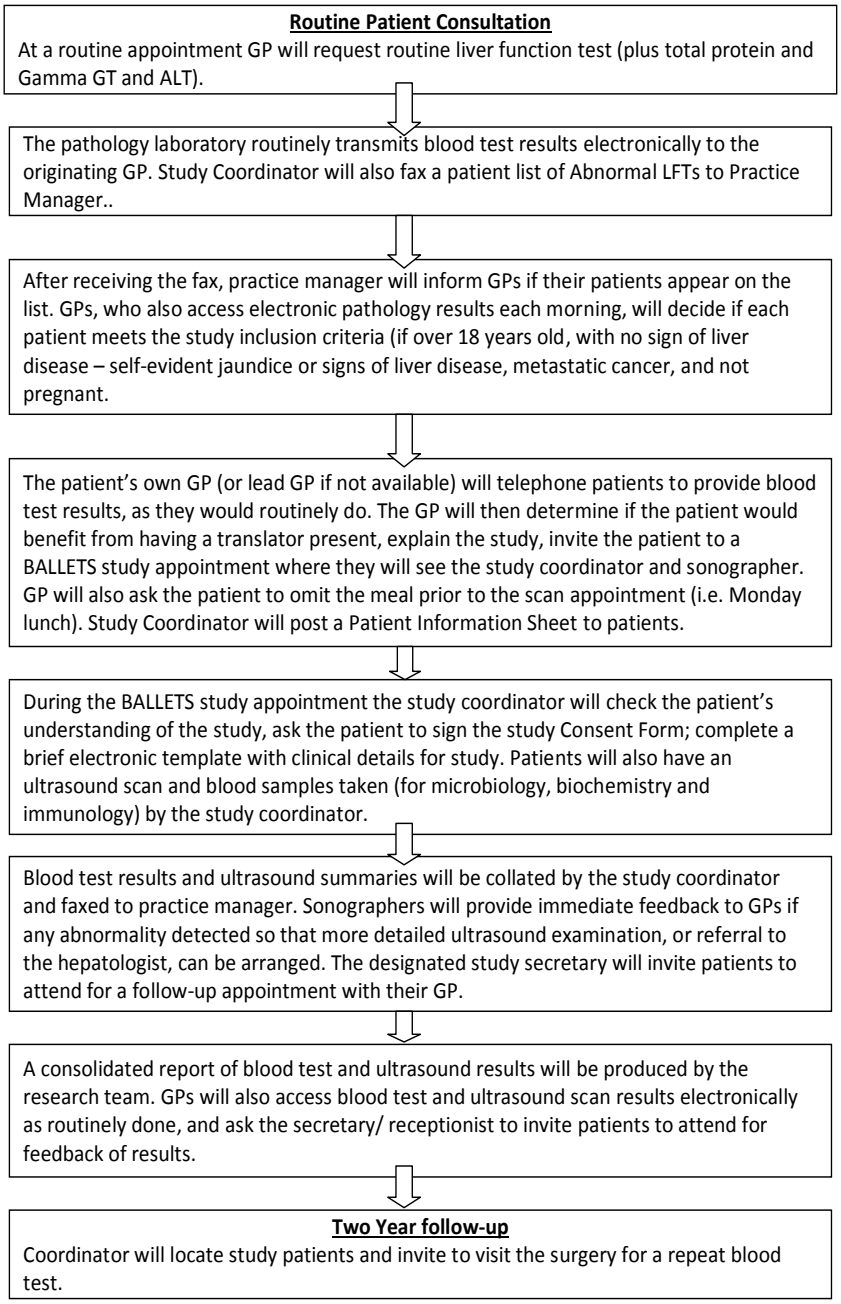
Appendix 10.2.a Hall Green Health BALLETS study patient process



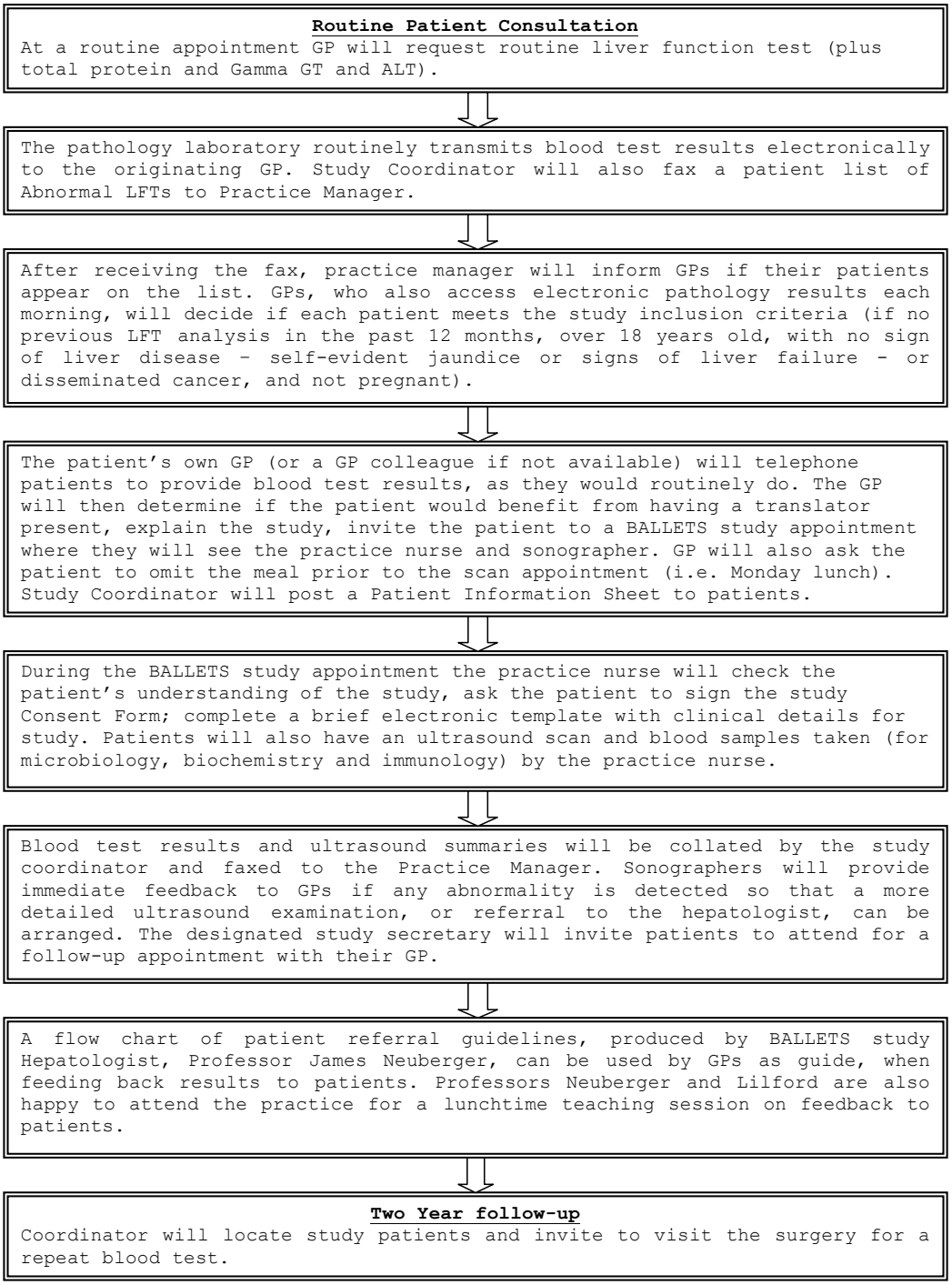
Appendix 10.2.b Lordswood House & Shenley Green Surgery BALLETS study patient process



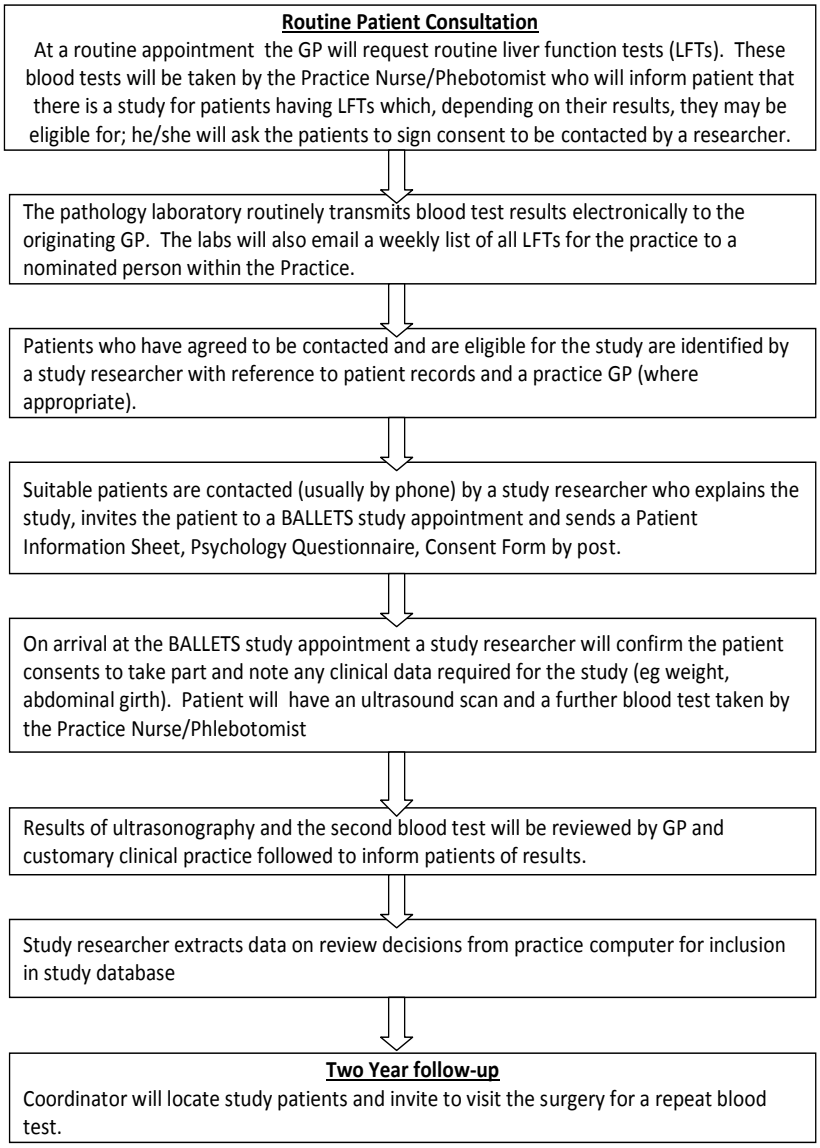
Appendix 10.2.c Greenridge Surgery BALLETS study patient process



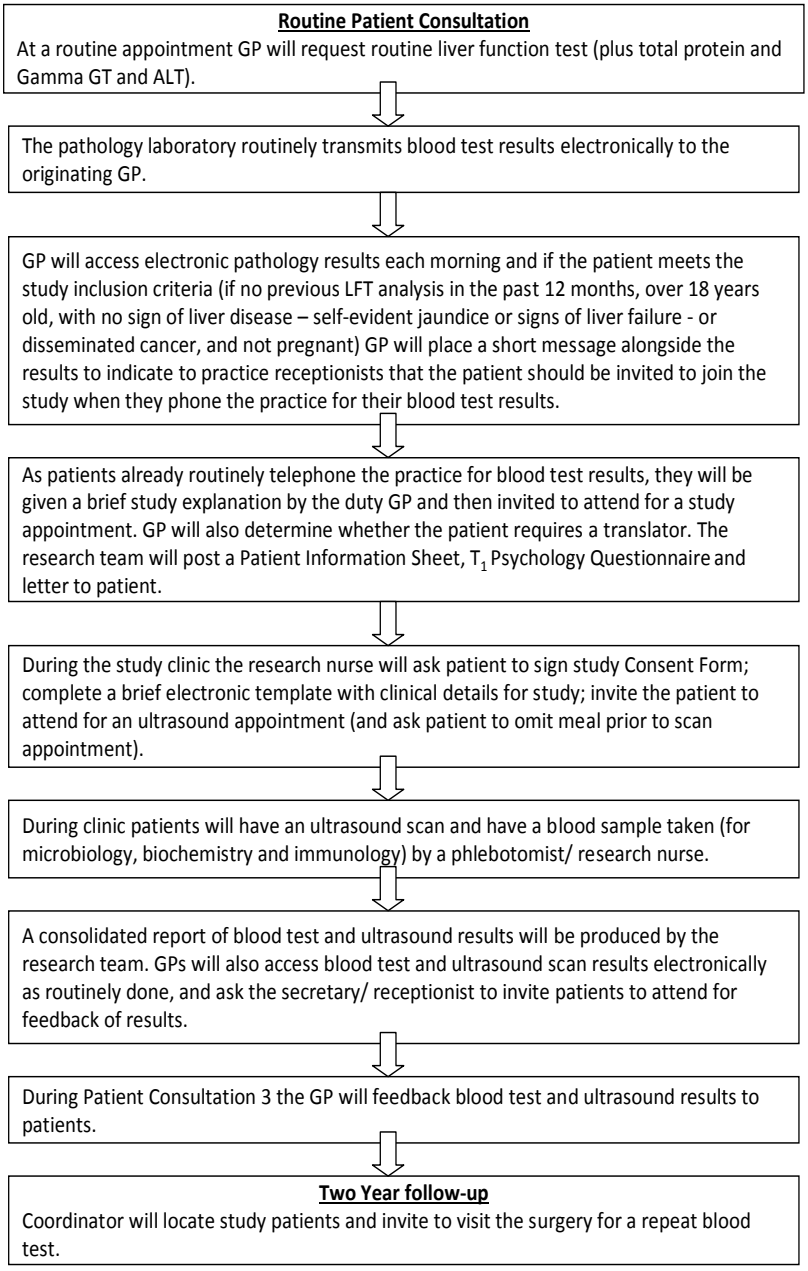
Appendix 10.2.d Yardley Wood Health Centre BALLETS study patient process



Appendix 10.2.e Lambeth BALLETS study patient process



Appendix 10.2.f Woodland Road Surgery, Wand and Cofton Medical Centres BALLETS study patient process



Appendix 10.3.a Hall Green Health Patient Information Sheet

UNIVERSITY OF BIRMINGHAM

PATIENT INFORMATION SHEET

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)
A study to evaluate the value of Liver Function Tests (LFTs)

You are being invited to take part in a research study. Before you decide whether or not you would like to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Consumers for Ethics in Research (CERES) publish a leaflet entitled 'Medical Research and You'. This leaflet gives more information about medical research and looks at some questions potential recruits may want to ask. You may obtain copies from CERES, PO Box 1365, London N16 0BW.

What is the purpose of the study?
Many general practitioners order blood tests when patients complain of symptoms such as tiredness and weight loss. These tests include liver function tests (LFTs) – these are a panel of tests that indicate how well the liver is working. The results may be clearly normal or abnormal, but sometimes only one or two of these tests is outside the normal range (ie higher or lower than occurs in most people) and so it is unclear whether or not it is something that needs further investigation. This is because these slightly high/low results may just be the normal level for a particular individual, or a temporary change that will return to normal by itself, as well as the possibility it is a sign of serious disease. We want to do this study to try to understand what these borderline results mean, so we can provide guidance to GPs to understand them better, and what further treatment/test are necessary (if any).

Awaiting tests results can be a stressful time, and so we would also like to look at the anxiety levels of patients, to see if these change over time, e.g. before and after they are given their test results.

Why have I been chosen?
You have been chosen as you recently saw your GP and he requested that LFTs be carried out. At least one of these has been reported as being outside the normal range for most people, and your GP does not know exactly why this is (as stated above, these could be *your* normal levels, or a temporary change that will resolve itself).

All adult patients who have these results have been asked to take part in the study and we aim to recruit 1,500 patients from practices in Birmingham and London.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?

After telephoning you to provide your blood test results, your GP will have explained the BALLETS study and invited you to BALLETS study appointments for an ultrasound scan and blood tests. Your GP will have also determined if you would like to have a translator present.

When you visit the BALLETS Nurse Practitioner clinic you will be asked to confirm that you are happy to be involved in the study, and if you are to sign a study Consent Form. The nurse practitioner will take a blood

sample, some clinical details will be recorded and you will be invited to attend the practice for an ultrasound appointment at the surgery. You will also be given you a short psychology questionnaire to complete.

At the ultrasound appointment a sonographer will do an ultrasound scan. The ultrasound machine is the same machine used when scanning the unborn child, and the examination involves lying on a couch, and exposing your stomach (you will not be asked to undress for this). A jelly will then be placed on the upper part of your stomach and a probe will be moved over your stomach to locate the liver and determine its characteristics. This should not take more than 30 minutes. It would be best if you do not eat for 4 hours before the ultrasound (but you can drink non-milky drinks). You should continue to take any medication, if any, as usual.

Your GP will be able to see your electronic blood test and ultrasound scan results as they are returned and will ask the secretary/ receptionist to invite you back to attend for feedback of the test results.

After 2 years the study coordinator will get in touch with you to invite you to visit the surgery for another blood test so we can look at your LFT results again. We will also ask you to fill in another psychology questionnaire.

What do I have to do?
If you are willing to take part in the study, you will be asked to sign a consent form. As outlined above in *"What will happen to me if I take part?*", we will then contact you to make the appointment for the ultrasound, and results of the repeat blood tests, and ask you to fill in a questionnaire to bring to the ultrasound appointment, and another questionnaire a week later. After 2 years, will be asked to give another blood sample, and fill out a third questionnaire.

There is also a possibility that you will be asked if you are willing to be interviewed by a researcher about your questionnaire responses. This is to help us make sure that the questionnaire that we are using does not miss out any important questions that we should be asking. We will only be asking a small number of participants to help us with this part of the study – if you are one of these participants, the study researcher will contact you directly to discuss it with you, and of course you can decline to be interviewed without it affecting either your standard clinical care, or the care you will receive as a participant in the main study.

What are the side effects of any examination received when taking part?
Ultrasound examinations are done routinely and no discomfort or side effects have been reported.

What are the possible disadvantages and risks of taking part?

The additional blood tests that will be done are those that your GP may have requested at a later date, if they felt that your liver needed further investigation. This is also true of the ultrasound, although far fewer people do go on to have an ultrasound as part of their routine clinical care, so there are no disadvantages/risks in this part of the study.

However, taking part in the study will mean that you will asked to give up some of your time, as you will be asked to complete 3 anxiety questionnaires (and possibly an interview about your questionnaire responses), and attend an additional follow-up appointment after 2 years, including a blood test, which will be in addition to your routine care.

What are the possible benefits of taking part?
You will have your abnormal blood tests investigated thoroughly, possibly more thoroughly than your GP would be expected to do in normal circumstances.

If we find a condition that you were unaware of then we will inform your GP and ask you to see them (all information about you will only be given to those involved in this study). As with many other diseases, liver diseases may be treated more effectively if detected earlier.

At present we do not know what these abnormal LFTs mean for patient's health and so hope to provide fuller guidance to GPs on their management.

What happens at the end of the study?
Your GP will continue to manage you according to his/her usual practice. The results of the study will be published in a report and in a medical journal and you will not be identified in any reports.

We will also your permission to keep in touch with you after the study has ended, as we are hoping that we will be able to follow-up the participants of this study for a number of years, eg to find out if any participants have developed liver disease since the study has ended.

What if something goes wrong?
This study does not involve any new treatment or invasive investigations, and so we do not envisage any risks to you. Of course, if at any stage during this study there is any risk to you or any of the other patients taking part you will be informed immediately and will not be expected to carry on in the study. The study is insured by The University of Birmingham.

Will my taking part in this study be kept confidential?
All information, which is collected about you during the course of the research will be kept strictly confidential. Any information about you that leaves the hospital/surgery will have your name and address removed so that you cannot be recognised from it.

As we have said above, we will be asking you to help us by allowing us to contact you in the future. This means that we will need to have your NHS number to trace you in case you move.

Who is organising and funding the research?
The NHS Heath Technology Assessment Programme is sponsoring this study, which is being carried out by researchers at Birmingham and London Universities, in association with local GPs.
Your GP's surgery will be paid for the staff time involved with taking part in this study.

Contact for Further Information
Your GP can answer any questions you have about the study. Also, Prof Richard Lilford, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 2226
Email: r.j.lilford@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and your GP will keep a copy in your notes.

Thank you for your help with this study.

Appendix 10.3.b Lordswood House Medical Practice Patient Information Sheet

UNIVERSITY OF BIRMINGHAM

PATIENT INFORMATION SHEET

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)
A study to evaluate the value of Liver Function Tests (LFTs)

You are being invited to take part in a research study. Before you decide whether or not you would like to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?
Many general practitioners order blood tests when patients complain of symptoms such as tiredness and weight loss. These tests include liver function tests (LFTs) – these are a panel of tests that indicate how well the liver is working. The results may be clearly normal or abnormal, but sometimes only one or two of these tests is outside the normal range (i.e. higher or lower than occurs in most people) and so it is unclear whether or not it is something that needs further investigation. This is because these slightly high/low results may just be the normal level for a particular individual, or a temporary change that will return to normal by itself, as well as the possibility it is a sign of serious disease. We want to do this study to try to understand what these borderline results mean, so we can provide guidance to GPs to understand them better, and what further treatment/test are necessary (if any).

Awaiting tests results can be a stressful time, and so we would also like to look at the anxiety levels of patients, to see if these change over time, e.g. before and after they are given their test results.

Why have I been chosen?
You have been chosen as you recently saw your GP and he requested that LFTs be carried out. At least one of these has been reported as being outside the normal range for most people, and your GP does not know exactly why this is (as stated above, these could be *your* normal levels, or a temporary change that will resolve itself).

All adult patients who have these results have been asked to take part in the study and we aim to recruit 1,500 patients from practices in Birmingham and London.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. *A decision to withdraw at any time, or a decision not to take part,* will not affect the standard of care you receive.

What will happen to me if I take part?

1. When you telephoned the practice for your blood test results, the receptionist will have given a brief explanation of the study and invited you to attend for a GP appointment. A letter from the GPs at the practice will have been sent to you with this information sheet.

2. When you visit your GP you will be asked to confirm that you are happy to be involved in the study, and if you are, to sign a study Consent Form. Some clinical details will be recorded and you will be invited to attend the practice for an ultrasound appointment and blood test. Your GP will also determine if you would like to have a translator present.

3. When you attend for the ultrasound scan at the practice, you will also have a blood sample taken by a phlebotomist or the study coordinator. The ultrasound machine is the same machine used when scanning the unborn child, and the examination involves lying on a couch, and exposing your stomach (you will not be asked to undress for this). A jelly will then be placed on the upper part of your stomach and a probe will be moved over your stomach to locate the liver and determine its characteristics. This should not take more than 30 minutes. It would be best if you do not eat for 4 hours before the ultrasound (but you can drink non-milky drinks). You should continue to take any medication, if any, as usual. At this appointment the study coordinator will ask you some clinical questions and give you a short psychology questionnaire to complete.

4. Your GP will be able to see your electronic blood test and ultrasound scan results as they are returned and will ask the secretary/ receptionist to invite you back to attend for feedback of the test results.

5. After 2 years the coordinator will get in touch with you to invite you to visit the surgery for another blood test so we can look at your LFT results again. We will also ask you to fill in another psychology questionnaire.

What do I have to do?
If you are willing to take part in the study, you will be asked to sign a consent form. As outlined above in *"What will happen to me if I take part?"*, the GP practice will then contact you to make the appointment for the ultrasound, and results of the repeat blood tests, and ask you to fill in a questionnaire to bring to the ultrasound appointment. After 2 years, will be asked to give another blood sample, and fill out a third questionnaire.

There is also a possibility that you will be asked if you are willing to be interviewed by a researcher about your questionnaire responses. This is to help us make sure that the questionnaire that we are using does not miss out any important questions that we should be asking. We will only be asking a small number of participants to help us with this part of the study – if you are one of these participants, the study researcher will contact you directly to discuss it with you, and of course you can decline to be interviewed without it affecting either your standard clinical care, or the care you will receive as a participant in the main study.

What are the side effects of any examination received when taking part?
Ultrasound examinations are done routinely and no discomfort or side effects have been reported.

What are the possible disadvantages and risks of taking part?

The additional blood tests that will be done are those that your GP may have requested at a later date, if they felt that your liver needed further investigation. This is also true of the ultrasound, although far fewer people do go on to have an ultrasound as part of their routine clinical care, so there are no disadvantages/risks in this part of the study.

However, taking part in the study will mean that you will asked to give up some of your time, as you will be asked to complete 3 anxiety questionnaires (and possibly an interview about your questionnaire responses), and attend an additional follow-up appointment after 2 years, including a blood test, which will be in addition to your routine care.

What are the possible benefits of taking part?
You will have your abnormal blood tests investigated thoroughly, possibly more thoroughly than your GP would be expected to do in normal circumstances.

If we find a condition that you were unaware of then we will inform your GP and ask you to see them (all information about you will only be given to those involved in this study). As with many other diseases, liver diseases may be treated more effectively if detected earlier.

At present we do not know what these abnormal LFTs mean for patient's health and so hope to provide fuller guidance to GPs on their management.

What happens at the end of the study?
Your GP will continue to manage you according to his/her usual practice. The results of the study will be published in a report and in a medical journal and you will not be identified in any reports.

We will also your permission to keep in touch with you after the study has ended, as we are hoping that we will be able to follow-up the participants of this study for a number of years, e.g. to find out if any participants have developed liver disease since the study has ended.

What if something goes wrong?
This study does not involve any new treatment or invasive investigations, and so we do not envisage any risks to you. Of course, if at any stage during this study there is any risk to you or any of the other patients taking part you will be informed immediately and will not be expected to carry on in the study. The study is insured by The University of Birmingham.

Will my taking part in this study be kept confidential?
All information, which is collected about you during the course of the research will be kept strictly confidential. Any information about you that leaves the hospital/surgery will have your name and address removed so that you cannot be recognised from it.

As we have said above, we will be asking you to help us by allowing us to contact you in the future. This means that we will need to have your NHS number to trace you in case you move.

Who is organising and funding the research?
The NHS Heath Technology Assessment Programme is sponsoring this study, which is being carried out by researchers at Birmingham and London Universities, in association with local GPs.
Your GP's surgery will be paid for the staff time involved with taking part in this study.

Contact for Further Information
Your GP can answer any questions you have about the study. Also, Prof Richard Lilford, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 2226
Email: r.j.lilford@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and your GP will keep a copy in your notes.

Thank you for your help with this study.

Appendix 10.3.c Greenridge Surgery Patient Information Sheet

UNIVERSITY OF BIRMINGHAM

PATIENT INFORMATION SHEET

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)
A study to evaluate the value of Liver Function Tests (LFTs)

You are being invited to take part in a research study. Before you decide whether or not you would like to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?
Many general practitioners order blood tests when patients complain of symptoms such as tiredness and weight loss. These tests include liver function tests (LFTs) – these are a panel of tests that indicate how well the liver is working. The results may be clearly normal or abnormal, but sometimes only one or two of these tests is outside the normal range (ie higher or lower than occurs in most people) and so it is unclear whether or not it is something that needs further investigation. This is because these slightly high/low results may just be the normal level for a particular individual, or a temporary change that will return to normal by itself, as well as the possibility it is a sign of serious disease. We want to do this study to try to understand what these borderline results mean, so we can provide guidance to GPs to understand them better, and what further treatment/test are necessary (if any).

Awaiting tests results can be a stressful time, and so we would also like to look at the anxiety levels of patients, to see if these change over time, e.g. before and after they are given their test results.

Why have I been chosen?
You have been chosen as you recently saw your GP and he requested that LFTs be carried out. At least one of these has been reported as being outside the normal range for most people, and your GP does not know exactly why this is (as stated above, these could be *your* normal levels, or a temporary change that will resolve itself).

All adult patients who have these results have been asked to take part in the study and we aim to recruit 1,500 patients from practices in Birmingham and London.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?

1. Your GP (or Dr Jhass if your GP was not available) will have telephoned you to provide your blood test results. Your GP will have explained the study, invited you to a BALLETS study appointment for an ultrasound scan and blood test, and determined if you would like to have a translator present.
2. After this phone call the BALLETS Study Coordinator sent you this Patient Information Sheet, which contains further details about the study and a telephone number to call if you would like to ask more questions.

3. When you visit the clinic the Study Coordinator will ask you to confirm that you are happy to be involved in the study, and if you are, to sign a study Consent Form. The study coordinator will take a blood sample and a sonographer will do an ultrasound scan. At this appointment the study coordinator will also ask you some clinical questions and give you a short psychology questionnaire to complete.
4. The ultrasound machine is the same machine used when scanning the unborn child, and the examination involves lying on a couch, and exposing your stomach (you will not be asked to undress for this). A jelly will then be placed on the upper part of your stomach and a probe will be moved over your stomach to locate the liver and determine its characteristics. This should not take more than 30 minutes. It would be best if you do not eat for 4 hours before the ultrasound (but you can drink non-milky drinks). You should continue to take any medication, if any, as usual.
5. Your GP will be able to see your electronic blood test and ultrasound scan results as they are returned and will ask the secretary/ receptionist to invite you back to attend for feedback of the test results.
6. After 2 years the study coordinator will get in touch with you to invite you to visit the surgery for another blood test so we can look at your LFT results again. We will also ask you to fill in another psychology questionnaire.

What do I have to do?
If you are willing to take part in the study, you will be asked to sign a consent form. As outlined above in *"What will happen to me if I take part?*", we will then contact you to make the appointment for the ultrasound, and results of the repeat blood tests, and ask you to fill in a questionnaire to bring to the ultrasound appointment, and another questionnaire a week later. After 2 years, will be asked to give another blood sample, and fill out a third questionnaire.

There is also a possibility that you will be asked if you are willing to be interviewed by a researcher about your questionnaire responses. This is to help us make sure that the questionnaire that we are using does not miss out any important questions that we should be asking. We will only be asking a small number of participants to help us with this part of the study – if you are one of these participants, the study researcher will contact you directly to discuss it with you, and of course you can decline to be interviewed without it affecting either your standard clinical care, or the care you will receive as a participant in the main study.

What are the side effects of any examination received when taking part?
Ultrasound examinations are done routinely and no discomfort or side effects have been reported.

What are the possible disadvantages and risks of taking part?

The additional blood tests that will be done are those that your GP may have requested at a later date, if they felt that your liver needed further investigation. This is also true of the ultrasound, although far fewer people do go on to have an ultrasound as part of their routine clinical care, so there are no disadvantages/risks in this part of the study.

However, taking part in the study will mean that you will asked to give up some of your time, as you will be asked to complete 3 anxiety questionnaires (and possibly an interview about your questionnaire responses), and attend an additional follow-up appointment after 2 years, including a blood test, which will be in addition to your routine care.

What are the possible benefits of taking part?
You will have your abnormal blood tests investigated thoroughly, possibly more thoroughly than your GP would be expected to do in normal circumstances.

If we find a condition that you were unaware of then we will inform your GP and ask you to see them (all information about you will only be given to those involved in this study). As with many other diseases, liver diseases may be treated more effectively if detected earlier.

At present we do not know what these abnormal LFTs mean for patient's health and so hope to provide fuller guidance to GPs on their management.

What happens at the end of the study?
Your GP will continue to manage you according to his/her usual practice. The results of the study will be published in a report and in a medical journal and you will not be identified in any reports.

We will also your permission to keep in touch with you after the study has ended, as we are hoping that we will be able to follow-up the participants of this study for a number of years, eg to find out if any participants have developed liver disease since the study has ended.

What if something goes wrong?
This study does not involve any new treatment or invasive investigations, and so we do not envisage any risks to you. Of course, if at any stage during this study there is any risk to you or any of the other patients taking part you will be informed immediately and will not be expected to carry on in the study. The study is insured by The University of Birmingham.

Will my taking part in this study be kept confidential?
All information, which is collected about you during the course of the research will be kept strictly confidential. Any information about you that leaves the hospital/surgery will have your name and address removed so that you cannot be recognised from it.

As we have said above, we will be asking you to help us by allowing us to contact you in the future. This means that we will need to have your NHS number to trace you in case you move.

Who is organising and funding the research?
The NHS Heath Technology Assessment Programme is sponsoring this study, which is being carried out by researchers at Birmingham and London Universities, in association with local GPs.
Your GP's surgery will be paid for the staff time involved with taking part in this study.

Contact for Further Information
Your GP can answer any questions you have about the study. Also, Prof Richard Lilford, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 2226
Email: r.j.lilford@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and your GP will keep a copy in your notes.

Thank you for your help with this study.

Appendix 10.3.d Yardley Wood Health Centre Patient Information Sheet

UNIVERSITY OF BIRMINGHAM

PATIENT INFORMATION SHEET

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)
A study to evaluate the value of Liver Function Tests (LFTs)

You are being invited to take part in a research study. Before you decide whether or not you would like to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?
Many general practitioners order blood tests when patients complain of symptoms such as tiredness and weight loss. These tests include liver function tests (LFTs) – these are a panel of tests that indicate how well the liver is working. The results may be clearly normal or abnormal, but sometimes only one or two of these tests is outside the normal range (ie higher or lower than occurs in most people) and so it is unclear whether or not it is something that needs further investigation. This is because these slightly high/low results may just be the normal level for a particular individual, or a temporary change that will return to normal by itself, as well as the possibility it is a sign of serious disease. We want to do this study to try to understand what these borderline results mean, so we can provide guidance to GPs to understand them better, and what further treatment/test are necessary (if any).

Awaiting tests results can be a stressful time, and so we would also like to look at the anxiety levels of patients, to see if these change over time, e.g. before and after they are given their test results.

Why have I been chosen?
You have been chosen as you recently saw your GP and he requested that LFTs be carried out. At least one of these has been reported as being outside the normal range for most people, and your GP does not know exactly why this is (as stated above, these could be *your* normal levels, or a temporary change that will resolve itself).

All adult patients who have these results have been asked to take part in the study and we aim to recruit 1,500 patients from practices in Birmingham and London.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?

1. Your GP will have telephoned you to provide your blood test results. Your GP will have explained the study, invited you to a BALLETS study appointment for an ultrasound scan and blood test, and determined if you would like to have a translator present.
2. After this phone call the BALLETS study coordinator sent you this Patient Information Sheet, which contains further details about the study and a telephone number to call if you would like to ask more questions.
3. When you visit the clinic the study coordinator will ask you to confirm that you are happy to be involved in the study, and if you are, to sign a study Consent Form. The study coordinator will

take a blood sample and a sonographer will do an ultrasound scan. At this appointment the study coordinator will also ask you some clinical questions and give you a short psychology questionnaire to complete.

4. The ultrasound machine is the same machine used when scanning the unborn child, and the examination involves lying on a couch, and exposing your stomach (you will not be asked to undress for this). A jelly will then be placed on the upper part of your stomach and a probe will be moved over your stomach to locate the liver and determine its characteristics. This should not take more than 30 minutes. It would be best if you do not eat for 4 hours before the ultrasound (but you can drink non-milky drinks). You should continue to take any medication, if any, as usual.
5. Your GP will be able to see your electronic blood test and ultrasound scan results as they are returned and will ask the secretary/ receptionist to invite you back to attend for feedback of the test results.
6. After 2 years the study coordinator will get in touch with you to invite you to visit the surgery for another blood test so we can look at your LFT results again. We will also ask you to fill in another psychology questionnaire.

What do I have to do?
If you are willing to take part in the study, you will be asked to sign a consent form. As outlined above in *"What will happen to me if I take part?*", we will then contact you to make the appointment for the ultrasound, and results of the repeat blood tests, and ask you to fill in a questionnaire to bring to the ultrasound appointment, and another questionnaire a week later. After 2 years, will be asked to give another blood sample, and fill out a third questionnaire.

There is also a possibility that you will be asked if you are willing to be interviewed by a researcher about your questionnaire responses. This is to help us make sure that the questionnaire that we are using does not miss out any important questions that we should be asking. We will only be asking a small number of participants to help us with this part of the study – if you are one of these participants, the study researcher will contact you directly to discuss it with you, and of course you can decline to be interviewed without it affecting either your standard clinical care, or the care you will receive as a participant in the main study.

What are the side effects of any examination received when taking part?
Ultrasound examinations are done routinely and no discomfort or side effects have been reported.

What are the possible disadvantages and risks of taking part?

The additional blood tests that will be done are those that your GP may have requested at a later date, if they felt that your liver needed further investigation. This is also true of the ultrasound, although far fewer people do go on to have an ultrasound as part of their routine clinical care, so there are no disadvantages/risks in this part of the study.

However, taking part in the study will mean that you will asked to give up some of your time, as you will be asked to complete 3 anxiety questionnaires (and possibly an interview about your questionnaire responses), and attend an additional follow-up appointment after 2 years, including a blood test, which will be in addition to your routine care.

What are the possible benefits of taking part?
You will have your abnormal blood tests investigated thoroughly, possibly more thoroughly than your GP would be expected to do in normal circumstances.

If we find a condition that you were unaware of then we will inform your GP and ask you to see them (all information about you will only be given to those involved in this study). As with many other diseases, liver diseases may be treated more effectively if detected earlier.

At present we do not know what these abnormal LFTs mean for patient's health and so hope to provide fuller guidance to GPs on their management.

What happens at the end of the study?
Your GP will continue to manage you according to his/her usual practice. The results of the study will be published in a report and in a medical journal and you will not be identified in any reports.

We will also your permission to keep in touch with you after the study has ended, as we are hoping that we will be able to follow-up the participants of this study for a number of years, e.g. to find out if any participants have developed liver disease since the study has ended.

What if something goes wrong?
This study does not involve any new treatment or invasive investigations, and so we do not envisage any risks to you. Of course, if at any stage during this study there is any risk to you or any of the other patients taking part you will be informed immediately and will not be expected to carry on in the study. The study is insured by The University of Birmingham.

Will my taking part in this study be kept confidential?
All information, which is collected about you during the course of the research will be kept strictly confidential. Any information about you that leaves the hospital/surgery will have your name and address removed so that you cannot be recognised from it.

As we have said above, we will be asking you to help us by allowing us to contact you in the future. This means that we will need to have your NHS number to trace you in case you move.

Who is organising and funding the research?
The NHS Heath Technology Assessment Programme is sponsoring this study, which is being carried out by researchers at Birmingham and London Universities, in association with local GPs.
Your GP's surgery will be paid for the staff time involved with taking part in this study.

Contact for Further Information
Your GP can answer any questions you have about the study. Also, Prof Richard Lilford, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 2226
Email: r.j.lilford@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and your GP will keep a copy in your notes.

Thank you for your help with this study.

Appendix 10.3.e Woodland Road Surgery Patient Information Sheet

UNIVERSITY OF BIRMINGHAM

PATIENT INFORMATION SHEET

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)
A study to evaluate the value of Liver Function Tests (LFTs)

You are being invited to take part in a research study. Before you decide whether or not you would like to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?
Many general practitioners order blood tests when patients complain of symptoms such as tiredness and weight loss. These tests include liver function tests (LFTs) – these are a panel of tests that indicate how well the liver is working. The results may be clearly normal or abnormal, but sometimes only one or two of these tests is outside the normal range (ie higher or lower than occurs in most people) and so it is unclear whether or not it is something that needs further investigation. This is because these slightly high/low results may just be the normal level for a particular individual, or a temporary change that will return to normal by itself, as well as the possibility it is a sign of serious disease. We want to do this study to try to understand what these borderline results mean, so we can provide guidance to GPs to understand them better, and what further treatment/test are necessary (if any).

Awaiting tests results can be a stressful time, and so we would also like to look at the anxiety levels of patients, to see if these change over time, e.g. before and after they are given their test results.

Why have I been chosen?
You have been chosen as you recently saw your GP and he requested that LFTs be carried out. At least one of these has been reported as being outside the normal range for most people, and your GP does not know exactly why this is (as stated above, these could be *your* normal levels, or a temporary change that will resolve itself).

All adult patients who have these results have been asked to take part in the study and we aim to recruit 1,500 patients from practices in Birmingham and London.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?
1. When you telephoned the practice for your blood test results, the receptionist will have given a brief explanation of the study and invited you to attend a BALLETS study appointment. A letter from the GPs at the practice, and a psychology questionnaire, will have been sent to you with this information sheet.

2. When you attend for a BALLETS clinic appointment at the surgery, the study coordinator or a research nurse, will tell you some more about the study before asking you to sign a consent form. At the same appointment you will have some blood taken and an ultrasound scan. The ultrasound machine is the same

machine used when scanning the unborn child, and the examination involves lying on a couch, and exposing your stomach (you will not be asked to undress for this). A jelly will then be placed on the upper part of your stomach and a probe will be moved over your stomach to locate the liver and determine its characteristics. This should not take more than 30 minutes. It would be best if you do not eat for 4 hours before the ultrasound (but you can drink non-milky drinks). You should continue to take any medication, if any, as usual.

3. Your GP will be able to see your electronic blood test and ultrasound scan results as they are returned. Three weeks after the appointment, when all the results are available, you can make an appointment to see your GP to discuss the results.

4. BALLETS Study psychologists will send you a questionnaire to complete and return.

5. After 2 years the coordinator will get in touch with you to invite you to visit the surgery for another blood test so we can look at your LFT results again. We will also ask you to fill in another psychology questionnaire.

What do I have to do?
If you are willing to take part in the study, you will be asked to sign a consent form. As outlined above in *"What will happen to me if I take part?*", we will then contact you to make the appointment for the ultrasound, and results of the repeat blood tests, and ask you to fill in a questionnaire to bring to the ultrasound appointment, and another questionnaire a week later. After 2 years, will be asked to give another blood sample, and fill out a third questionnaire.

There is also a possibility that you will be asked if you are willing to be interviewed by a researcher about your questionnaire responses. This is to help us make sure that the questionnaire that we are using does not miss out any important questions that we should be asking. We will only be asking a small number of participants to help us with this part of the study – if you are one of these participants, the study researcher will contact you directly to discuss it with you, and of course you can decline to be interviewed without it affecting either your standard clinical care, or the care you will receive as a participant in the main study.

What are the side effects of any examination received when taking part?
Ultrasound examinations are done routinely and no discomfort or side effects have been reported.

What are the possible disadvantages and risks of taking part?

The additional blood tests that will be done are those that your GP may have requested at a later date, if they felt that your liver needed further investigation. This is also true of the ultrasound, although far fewer people do go on to have an ultrasound as part of their routine clinical care, so there are no disadvantages/risks in this part of the study.

However, taking part in the study will mean that you will asked to give up some of your time, as you will be asked to complete 3 anxiety questionnaires (and possibly an interview about your questionnaire responses), and attend an additional follow-up appointment after 2 years, including a blood test, which will be in addition to your routine care.

What are the possible benefits of taking part?
You will have your abnormal blood tests investigated thoroughly, possibly more thoroughly than your GP would be expected to do in normal circumstances.

If we find a condition that you were unaware of then we will inform your GP and ask you to see them (all information about you will only be given to those involved in this study). As with many other diseases, liver diseases may be treated more effectively if detected earlier.

At present we do not know what these abnormal LFTs mean for patient's health and so hope to provide fuller guidance to GPs on their management.

What happens at the end of the study?
Your GP will continue to manage you according to his/her usual practice. The results of the study will be published in a report and in a medical journal and you will not be identified in any reports.

We will also your permission to keep in touch with you after the study has ended, as we are hoping that we will be able to follow-up the participants of this study for a number of years, eg to find out if any participants have developed liver disease since the study has ended.

What if something goes wrong?
This study does not involve any new treatment or invasive investigations, and so we do not envisage any risks to you. Of course, if at any stage during this study there is any risk to you or any of the other patients taking part you will be informed immediately and will not be expected to carry on in the study. The study is insured by The University of Birmingham.

Will my taking part in this study be kept confidential?
All information, which is collected about you during the course of the research will be kept strictly confidential. Any information about you that leaves the hospital/surgery will have your name and address removed so that you cannot be recognised from it.

As we have said above, we will be asking you to help us by allowing us to contact you in the future. This means that we will need to have your NHS number to trace you in case you move.

Who is organising and funding the research?
The NHS Heath Technology Assessment Programme is sponsoring this study, which is being carried out by researchers at Birmingham and London Universities, in association with local GPs.
Your GP's surgery will be paid for the staff time involved with taking part in this study.

Contact for Further Information
Your GP can answer any questions you have about the study. Also, Prof Richard Lilford, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 2226
Email: r.j.lilford@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and your GP will keep a copy in your notes.

Thank you for your help with this study.

Appendix 10.3.f Cofton Medical Centre Patient Information Sheet

UNIVERSITY OF BIRMINGHAM

PATIENT INFORMATION SHEET

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)
A study to evaluate the value of Liver Function Tests (LFTs)

You are being invited to take part in a research study. Before you decide whether or not you would like to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?
Many general practitioners order blood tests when patients complain of symptoms such as tiredness and weight loss. These tests include liver function tests (LFTs) – these are a panel of tests that indicate how well the liver is working. The results may be clearly normal or abnormal, but sometimes only one or two of these tests is outside the normal range (ie higher or lower than occurs in most people) and so it is unclear whether or not it is something that needs further investigation. This is because these slightly high/low results may just be the normal level for a particular individual, or a temporary change that will return to normal by itself, as well as the possibility it is a sign of serious disease. We want to do this study to try to understand what these borderline results mean, so we can provide guidance to GPs to understand them better, and what further treatment/test are necessary (if any).

Awaiting tests results can be a stressful time, and so we would also like to look at the anxiety levels of patients, to see if these change over time, e.g. before and after they are given their test results.

Why have I been chosen?
You have been chosen as you recently saw your GP and he requested that LFTs be carried out. At least one of these has been reported as being outside the normal range for most people, and your GP does not know exactly why this is (as stated above, these could be *your* normal levels, or a temporary change that will resolve itself).

All adult patients who have these results have been asked to take part in the study and we aim to recruit 1,500 patients from practices in Birmingham and London.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?
1. When you telephoned the practice for your blood test results, the triage GP will have given a brief explanation of the study and invited you to attend a BALLETS study appointment. A letter from the GPs at the practice, and a psychology questionnaire, will have been sent to you with this information sheet.

2. When you attend for a BALLETS clinic appointment at the surgery, the study coordinator or a research nurse, will tell you some more about the study before asking you to sign a consent form. At the same appointment you will have some blood taken and an ultrasound scan. The ultrasound machine is the same machine used when scanning the unborn child, and the examination involves lying on a couch, and exposing your stomach (you will not be asked to undress for this). A jelly will then be placed on the upper part of your stomach and a probe will be moved over your stomach to locate the liver and determine its characteristics. This should not take more than 30 minutes**. It would be best if you do not eat for 4 hours before the ultrasound (but you can drink non-milky drinks).** You should continue to take any medication, if any, as usual.

3. Your GP will be able to see your electronic blood test and ultrasound scan results as they are returned. Three weeks after the appointment, when all the results are available, you can phone the triage GP to discuss the results.
4. BALLETS Study psychologists will send you a questionnaire to complete and return.
5. After 2 years the coordinator will get in touch with you to invite you to visit the surgery for another blood test so we can look at your LFT results again. We will also ask you to fill in another psychology questionnaire.

What do I have to do?
If you are willing to take part in the study, you will be asked to sign a consent form. As outlined above in *"What will happen to me if I take part?"*, we will then contact you to make the appointment for the ultrasound, and results of the repeat blood tests, and ask you to fill in a questionnaire to bring to the ultrasound appointment, and another questionnaire a week later. After 2 years, will be asked to give another blood sample, and fill out a third questionnaire.

There is also a possibility that you will be asked if you are willing to be interviewed by a researcher about your questionnaire responses. This is to help us make sure that the questionnaire that we are using does not miss out any important questions that we should be asking. We will only be asking a small number of participants to help us with this part of the study – if you are one of these participants, the study researcher will contact you directly to discuss it with you, and of course you can decline to be interviewed without it affecting either your standard clinical care, or the care you will receive as a participant in the main study.

What are the side effects of any examination received when taking part?
Ultrasound examinations are done routinely and no discomfort or side effects have been reported.

What are the possible disadvantages and risks of taking part?

The additional blood tests that will be done are those that your GP may have requested at a later date, if they felt that your liver needed further investigation. This is also true of the ultrasound, although far fewer people do go on to have an ultrasound as part of their routine clinical care, so there are no disadvantages/risks in this part of the study.

However, taking part in the study will mean that you will asked to give up some of your time, as you will be asked to complete 3 anxiety questionnaires (and possibly an interview about your questionnaire responses), and attend an additional follow-up appointment after 2 years, including a blood test, which will be in addition to your routine care.

What are the possible benefits of taking part?
You will have your abnormal blood tests investigated thoroughly, possibly more thoroughly than your GP would be expected to do in normal circumstances.

If we find a condition that you were unaware of then we will inform your GP and ask you to see them (all information about you will only be given to those involved in this study). As with many other diseases, liver diseases may be treated more effectively if detected earlier.

At present we do not know what these abnormal LFTs mean for patient's health and so hope to provide fuller guidance to GPs on their management.

What happens at the end of the study?
Your GP will continue to manage you according to his/her usual practice. The results of the study will be published in a report and in a medical journal and you will not be identified in any reports.

We will also your permission to keep in touch with you after the study has ended, as we are hoping that we will be able to follow-up the participants of this study for a number of years, eg to find out if any participants have developed liver disease since the study has ended.

What if something goes wrong?
This study does not involve any new treatment or invasive investigations, and so we do not envisage any risks to you. Of course, if at any stage during this study there is any risk to you or any of the other patients taking part you will be informed immediately and will not be expected to carry on in the study. The study is insured by The University of Birmingham.

Will my taking part in this study be kept confidential?
All information, which is collected about you during the course of the research will be kept strictly confidential. Any information about you that leaves the hospital/surgery will have your name and address removed so that you cannot be recognised from it.

As we have said above, we will be asking you to help us by allowing us to contact you in the future. This means that we will need to have your NHS number to trace you in case you move.

Who is organising and funding the research?
The NHS Heath Technology Assessment Programme is sponsoring this study, which is being carried out by researchers at Birmingham and London Universities, in association with local GPs.
Your GP's surgery will be paid for the staff time involved with taking part in this study.

Contact for Further Information
Your GP can answer any questions you have about the study. Also, Prof Richard Lilford, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 2226
Email: r.j.lilford@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and your GP will keep a copy in your notes.

Thank you for your help with this study.

Appendix 10.3.g Wand Medical Centre Patient Information Sheet

UNIVERSITY OF BIRMINGHAM

PATIENT INFORMATION SHEET

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)
A study to evaluate the value of Liver Function Tests (LFTs)

You are being invited to take part in a research study. Before you decide whether or not you would like to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?
Many general practitioners order blood tests when patients complain of symptoms such as tiredness and weight loss. These tests include liver function tests (LFTs) – these are a panel of tests that indicate how well the liver is working. The results may be clearly normal or abnormal, but sometimes only one or two of these tests is outside the normal range (ie higher or lower than occurs in most people) and so it is unclear whether or not it is something that needs further investigation. This is because these slightly high/low results may just be the normal level for a particular individual, or a temporary change that will return to normal by itself, as well as the possibility it is a sign of serious disease. We want to do this study to try to understand what these borderline results mean, so we can provide guidance to GPs to understand them better, and what further treatment/test are necessary (if any).

Awaiting tests results can be a stressful time, and so we would also like to look at the anxiety levels of patients, to see if these change over time, e.g. before and after they are given their test results.

Why have I been chosen?
You have been chosen as you recently saw your GP and he requested that LFTs be carried out. At least one of these has been reported as being outside the normal range for most people, and your GP does not know exactly why this is (as stated above, these could be *your* normal levels, or a temporary change that will resolve itself).

All adult patients who have these results have been asked to take part in the study and we aim to recruit 1,500 patients from practices in Birmingham and London.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?
1. When you telephoned the practice for your blood test results, the duty doctor will have given a brief explanation of the study and invited you to attend a BALLETS study appointment. A letter from the GPs at the practice, and a psychology questionnaire, will have been sent to you with this information sheet.

2. When you attend for a BALLETS clinic appointment at the surgery, the study coordinator or a research nurse, will tell you some more about the study before asking you to sign a consent form. At the same appointment you will have some blood taken and an ultrasound scan. The ultrasound machine is the same machine used when scanning the unborn child, and the examination involves lying on a couch, and exposing your stomach (you will not be asked to undress for this). A jelly will then be placed on the upper

part of your stomach and a probe will be moved over your stomach to locate the liver and determine its characteristics. This should not take more than 30 minutes**. It would be best if you do not eat for 4 hours before the ultrasound (but you can drink non-milky drinks).** You should continue to take any medication, if any, as usual.

3. Your GP will be able to see your electronic blood test and ultrasound scan results as they are returned. Three weeks after the appointment you can phone the duty doctor to discuss your results.

4. BALLETS Study psychologists will send you another questionnaire to complete and return.

5. After 2 years the coordinator will get in touch with you to invite you to visit the surgery for another blood test so we can look at your LFT results again. We will also ask you to fill in another psychology questionnaire.

What do I have to do?
If you are willing to take part in the study, you will be asked to sign a consent form. As outlined above in *"What will happen to me if I take part?*", we will then contact you to make the appointment for the ultrasound, and results of the repeat blood tests, and ask you to fill in a questionnaire to bring to the ultrasound appointment, and another questionnaire a week later. After 2 years, will be asked to give another blood sample, and fill out a third questionnaire.

There is also a possibility that you will be asked if you are willing to be interviewed by a researcher about your questionnaire responses. This is to help us make sure that the questionnaire that we are using does not miss out any important questions that we should be asking. We will only be asking a small number of participants to help us with this part of the study – if you are one of these participants, the study researcher will contact you directly to discuss it with you, and of course you can decline to be interviewed without it affecting either your standard clinical care, or the care you will receive as a participant in the main study.

What are the side effects of any examination received when taking part?
Ultrasound examinations are done routinely and no discomfort or side effects have been reported.

What are the possible disadvantages and risks of taking part?

The additional blood tests that will be done are those that your GP may have requested at a later date, if they felt that your liver needed further investigation. This is also true of the ultrasound, although far fewer people do go on to have an ultrasound as part of their routine clinical care, so there are no disadvantages/risks in this part of the study.

However, taking part in the study will mean that you will asked to give up some of your time, as you will be asked to complete 3 anxiety questionnaires (and possibly an interview about your questionnaire responses), and attend an additional follow-up appointment after 2 years, including a blood test, which will be in addition to your routine care.

What are the possible benefits of taking part?
You will have your abnormal blood tests investigated thoroughly, possibly more thoroughly than your GP would be expected to do in normal circumstances.

If we find a condition that you were unaware of then we will inform your GP and ask you to see them (all information about you will only be given to those involved in this study). As with many other diseases, liver diseases may be treated more effectively if detected earlier.

At present we do not know what these abnormal LFTs mean for patient's health and so hope to provide fuller guidance to GPs on their management.

What happens at the end of the study?
Your GP will continue to manage you according to his/her usual practice. The results of the study will be published in a report and in a medical journal and you will not be identified in any reports.

We will also your permission to keep in touch with you after the study has ended, as we are hoping that we will be able to follow-up the participants of this study for a number of years, eg to find out if any participants have developed liver disease since the study has ended.

What if something goes wrong?
This study does not involve any new treatment or invasive investigations, and so we do not envisage any risks to you. Of course, if at any stage during this study there is any risk to you or any of the other patients taking part you will be informed immediately and will not be expected to carry on in the study. The study is insured by The University of Birmingham.

Will my taking part in this study be kept confidential?
All information, which is collected about you during the course of the research will be kept strictly confidential. Any information about you that leaves the hospital/surgery will have your name and address removed so that you cannot be recognised from it.

As we have said above, we will be asking you to help us by allowing us to contact you in the future. This means that we will need to have your NHS number to trace you in case you move.

Who is organising and funding the research?
The NHS Heath Technology Assessment Programme is sponsoring this study, which is being carried out by researchers at Birmingham and London Universities, in association with local GPs.
Your GP's surgery will be paid for the staff time involved with taking part in this study.

Contact for Further Information
Your GP can answer any questions you have about the study. Also, Prof Richard Lilford, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 2226
Email: r.j.lilford@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and your GP will keep a copy in your notes.

Thank you for your help with this study.

Appendix 10.3.h Lambeth Patient Information Sheet

BALLETS

Birmingham and Lambeth Liver Evaluation and Testing Strategies

UNIVERSITY OF BIRMINGHAM

KING'S College LONDON
University of London

PATIENT INFORMATION SHEET

A study to evaluate the value of Liver Function Tests

Summary

You are being invited to take part in a research study. Before you decide whether or not you want to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and talk about it with other people if you wish.

Liver Function Tests (LFTs) are a panel of tests that indicate how well the liver is working. The results may be clearly normal or abnormal, but sometimes fall only slightly outside the normal range (ie slightly higher or lower than occurs in most people). A result like this which is slightly higher or lower than normal means that it is unclear if something might be wrong and so the doctor may not be sure what needs to be done. We want to do this study to try to understand what these types of results mean, so we can provide guidance to GPs to understand them better and know which further treatments or tests may be needed.

Taking part involves further visits to your GP surgery for further blood tests and an ultrasound scan, although it is likely you would have had these additional visits even without being involved in the study You will also be asked to fill in three short questionnaires.

If there is anything that is not clear or if you would like more information, please contact Susan Hoult Robinson who is organising the study for King's College London. She will be happy to answer any questions. Phone: (020) 7848 4149 or email: s.hoult-robinson@kcl.ac.uk

Consumers for Ethics in Research (CERES) publish a leaflet entitled 'Medical Research and You'. This leaflet gives more information about medical research and looks at some questions potential recruits may want to ask. You may obtain copies from CERES, PO Box 1365, London N16 0BW.

What is the purpose of the study?

- General practitioners often order blood tests when patients complain of problems such as tiredness, weight loss or feeling rundown. These tests include Liver Function Tests (LFTs) which show the doctor how well the liver is working.

- The results may be clearly normal or clearly abnormal, but sometimes fall only slightly outside the normal range (ie slightly higher or lower than occurs in most people). A result like this which is slightly higher or lower than normal means that it is unclear if something might be wrong and so the doctor may not be sure what needs to be done. This is because these types of results may just be the normal level for a particular individual, a temporary change that will return to normal by itself, or possibly be a sign of serious disease.

- We want to do this study to try to understand what these slightly high or low results mean. This will help GPs to understand them better in the future, and to know which further treatments or tests may be needed.

We would also like to know more about how patients understand these tests and the effects that getting test results can have.

Why have I been chosen?
You have been chosen as you recently saw your GP and he or she requested that LFTs be carried out. At least one of these have been reported as being outside the normal range for most people, and your GP does not know exactly why this is (as stated above, these could be your normal levels, or a temporary change that will resolve itself).

All adult patients who have these results have been asked to take part in the study. We aim to recruit 1,500 patients in Birmingham and London.

Do I have to take part?
It is up to you to decide whether or not to take part. If you decide to take part, you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?

1) We will ask you to have a repeat blood test. Even if you do not take part in this study, your doctor is likely to suggest this. Your blood sample will be tested for LFTs and other tests at the laboratory. Having these tests now may save you the need to have them done later.

2) You will have an appointment with a study radiographer who will give you an ultrasound examination of your liver and collect other information, such as height and weight, from you. The ultrasound machine is the same machine used when scanning a pregnant woman. The examination involves lying on a couch, and scanning your abdomen (tummy). You will not have to undress for this, just lift your top up. A jelly will then be placed on the upper part of your stomach to locate the liver and determine its characteristics. The scan should not take more than 30 minutes and is painless.

3) Your GP will contact you again to will discuss your blood test and ultrasound results with you.

4) We will ask you to complete 3 questionnaires over the course of the study: one before you have the ultrasound scan and repeat blood test, one after these procedures and the final questionnaire 3 months after you enter the study. These can be filled out by post, email or over the phone depending on what is easiest for you. The questionnaires will look at your thoughts about the tests you have.

5) After this, your care will depend on your test results. Your doctor will discuss this with you in the usual way.

6) After 2 years, we will contact you again, and ask you to give another blood sample, so we can look at your LFT results again.

You may also be asked if you are willing to be interviewed by a researcher. This is to help us understand in more details patients' experiences of LFTs. We will only be asking a small number of participants. If you are one of these participants, the study researcher will contact you directly to discuss it with you. You can decline to be interviewed without it affecting either your clinical care, or your involvement in the main study.

What are the side effects of any examination received when taking part?
Ultrasound examinations are done routinely and no discomfort or side effects have been reported.

What are the possible disadvantages and risks of taking part?
The additional blood tests that will be done are those that your GP may have requested at a later date, if they thought that your liver needed further investigation. This is also true of the ultrasound, although fewer people do go on to have an ultrasound as part of their normal clinical care. Therefore there are no disadvantages or risk in this part of the study.

Taking part in the study will mean that you will be asked to give up some of your time to come for a blood test and scan, to complete 3 questionnaires (and possibly an interview about your experiences of LFTs), and attend an additional follow-up appointment after 2 years, including a blood test. This will be in addition to your normal care.

What are the possible benefits of taking part?
You will have your blood tests investigated more thoroughly than your GP would be expected to do in normal circumstances.

If we find a condition that you were unaware of then we will inform your GP and ask you to see them (all information about you will only be given to those involved in this study). As with many other diseases, liver diseases may be treated more effectively if detected earlier.

At present we do not know what these abnormal LFTs mean for patients' health. By contributing to this research you will be helping us to provide further help to GPs to understand them better, and know which further treatments or tests may be needed.

What happens at the end of the study?
Your GP will continue to manage you according to his or her usual practice. The results of the study will be published in a report and in a medical journal. You will not be personally identified in any reports.

We will also ask for your permission to keep in touch with you after the study has ended. We are hoping that we will be able to follow-up the participants of this study for a number of years to collect information on their longer-term health.

What if something goes wrong?
This study does not involve any new treatment or invasive investigations, and so we do not see any risks to you. Of course, if at any stage during this study there is any risk to you or any of the other patients taking part you will be informed immediately and will not be expected to carry on in the study.

Will my taking part in this study be kept confidential?
All information which is collected about you during the course of the research will be kept strictly confidential. As we have said above, we will be asking you to help us by allowing us to contact you in the future. This means that we will need to have your NHS number to trace you in case you move.

Who is organising and funding the research?
The NHS Health Technology Assessment Programme is funding this study, which is being carried out by researchers at Birmingham and London Universities, in association with local GPs. Your GP's surgery will be paid for the staff time involved with taking part in this study.

Contact for Further Information
Susan Hoult Robinson, who is organising the study for King's College London, will be happy to speak to you if you have any questions.
Phone: (020) 7848 4149
Email: s.hoult-robinson@kcl.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and your GP will keep a copy in your notes.

Thank you for your help with this study.

Appendix 10.4.a Birmingham Participants Consent Form

UNIVERSITY OF BIRMINGHAM

Practice ID:	
Patient Number:	

CONSENT FORM

BALLETS: Birmingham and Lambeth Liver Evaluation Testing Strategies

A study to evaluate the value of liver function tests

Lead Researcher: Professor Richard Lilford

Please initial boxes below

1. I confirm that I have read and understand the information sheet dated 23 August 2007, (version 4.0) for the above study and have had the opportunity to ask questions ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected ☐

3. I understand that sections of any of my medical notes may be looked at by individuals from the Research Team or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my records ☐

4. I agree to take part in the above study. ☐

Name of Patient	Date	Signature
Name of Person taking Consent (if different from Researcher)	Date	Signature
Researcher	Date	Signature

1 for patient; 1 for researcher; 1 to be kept with GP notes

BALLETS:Birmingham and Lambeth Liver Evaluation Testing Strategies

Contact details

I confirm that I have consented to participate in the above study, and that I am happy for the

Research Team to telephone me using the numbers I have given below.

Phone: _____

Mobile: _____

My preferred method of contact is: Landline / Mobile *

My preferred time of contact is: morning (9am – 1pm)

afternoon (1pm – 5pm)

evening (5pm – 7pm)*

I understand that the research team will be posting questionnaires to me, as indicated in the Patient Information Sheet. I would prefer to receive them via normal postal services/email * (where possible), at the following address(es):

Address: _____

_____

_____

_____

Email: _____

* please delete as applicable

_____ _____ _____

Name of Patient Date Signature

Appendix 10.4.b Lambeth Participants Consent Form

BALLETS
Birmingham and Lambeth Liver Evaluation and Testing Strategies

UNIVERSITY OF BIRMINGHAM
KING'S College LONDON
University of London

Practice ID:	
Patient Number:	

CONSENT FORM

BALLETS: Birmingham and Lambeth Liver Evaluation Testing Strategies

A study to evaluate the value of liver function tests

Lead Researcher: Professor Richard Lilford

Please initial boxes below

1. I confirm that I have read and understand the information sheet dated 14th June (version 2) for the above study and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐
3. I understand that sections of any of my medical notes may be looked at by study researchers of members of regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my records. ☐
4. I agree to take part in the above study. ☐

_____	_____	_____
Name of Patient	Date	Signature
_____	_____	_____
Name of Person taking consent (if different from researcher)	Date	Signature
_____	_____	_____
Researcher	Date	Signature

1 for patient; 1 for researcher; 1 to be kept with GP notes

BALLETS:Birmingham and Lambeth Liver Evaluation Testing Strategies

Contact details

I confirm that I have consented to participate in the above study, and that I am happy for a study researcher to telephone me using the numbers I have given below.

Phone: _____

Mobile: _____

My preferred method of contact is: Landline / Mobile *

My preferred time of contact is: morning (9am – 1pm)

afternoon (1pm – 5pm)

evening (5pm – 7pm)*

I understand that the research team will be asking me to fill in questionnaires, as indicated in the Patient Information Sheet. Please indicate how you would prefer to receive these questionnaires: via normal postal services: ☐ , via email: ☐ a researcher can contact me and ask me the questions over the phone: ☐ (please tick a box).

If applicable, please write your preferred address or email below.

Address: _____

_____

_____

_____

Email: _____

_____ _____ _____

Name of Patient Date Signature

Appendix 10.5.a Lordswood House Receptionist Script

UNIVERSITY OF BIRMINGHAM

BALLETS Study (Birmingham and Lambeth Liver Evaluation and Testing Strategies)

Lordswood House Medical Practice Receptionist Script (22/03/2006)

"Your liver function test was slightly abnormal and Dr ..... would like to discuss this with you.

Our practice is taking part in a Department of Health funded study looking into what is the best thing to do if someone has a slightly abnormal liver function test.

If liver function tests are found to be very abnormal patients are usually referred to a specialist, but it is unclear as to whether this is necessary when the results are only slightly abnormal, as in your case.

If you are interested in taking part in this study I will send you an information leaflet and you can discuss this with Dr ..... when you come here for your appointment. Basically the study would involve you having a further blood test and ultrasound scan here at the practice."

For further information about the study please contact:
Louise Bentham
BALLETS Study Coordinator,
Department of Public Health and Epidemiology,
The University of Birmingham
Edgbaston
B15 2TT
0121 414 6805

Appendix 10.5.b Lordswood House Letter to Patients

LORDSWOOD HOUSE

54, LORDSWOOD ROAD, HARBORNE, BIRMINGHAM B17 9DB

Dr M G Edward
Dr W van Marle
Dr M Simpson
Dr E Hamnett
Dr G R D Ralston
Dr R H A Jordan
Dr J T Whiteley
Dr T M Mulcahy

Telephone: 0121 426 2030

Fax Number: 0121 428 2658

Dear

Re: BALLETS (Birmingham and Lambeth Liver Evaluation and Testing Strategies) Study.

Our practice is taking part in a Department of Health funded study looking into what is the best thing to do if someone has a slightly abnormal liver function test.

If liver function tests are found to be very abnormal patients are usually referred to a specialist, but it is unclear as to whether this is necessary when the results are only slightly abnormal, as in your case.

We would like to invite you to take part in the BALLETS study, as we feel it offers a good, additional service to our patients and basically involves a further blood test and ultrasound scan at the practice.

The enclosed information leaflet describes the study and your GP would be happy to discuss it with you when you come here for your appointment.

The doctors at Lordswood would like to stress that it is up to you to decide whether or not to take part. If you do take part you are still free to withdraw at any time and without giving a reason. We feel that it is also important for you to know that a decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

Yours sincerely,

THE DOCTORS AT LORDSWOOD HOUSE

Appendix 10.6 Lambeth Letter to Patients

BALLETS
Birmingham and Lambeth Liver Evaluation and Testing Strategies

UNIVERSITY OF BIRMINGHAM
KING'S College LONDON
University of London

Patient ID:
Patient Number:

Department of General Practice and Primary Care
Kings College London
5 Lambeth Walk
LONDON SE11 6SP
Tel +44 207 848 4149 (direct)
Tel +44 207 848 4100 (dept)
Fax +44 207 848 4102

CONSENT FORM

Your GP Practice is participating in a study being conducted by the Department of General Practice at King's College London School of Medicine and the University of Birmingham to evaluate the value of liver function tests. The Chief Investigators are Professor Richard Lilford of the University of Birmingham and Dr David Armstrong of King's College London.

Your GP has asked you to have a liver function blood test. If the results of this test show that you would be a suitable candidate for our study and you give your consent, you will be contacted by a researcher. The researcher will be happy to answer any questions you have about the study and will send you a detailed information sheet. Briefly, taking part in the study would involve you returning to your GP Practice for an additional blood test and ultrasound scan of your abdomen (tummy). Your participation will be completely voluntary and you will be free to withdraw from the study at any time without giving a reason.

In order to see if you would be suitable for this study, we need your permission for study researchers to look at your medical notes and for a researcher to contact you directly. You details will be kept completely confidential.

1. I confirm that I am willing for a study researcher to contact me on the telephone numbers given below. ☐

2. I understand that sections of any of my medical notes may be looked at by study researchers or members of regulatory authorities where it is relevant to the study. I give permission for these individuals to have access to my records. ☐

3. I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason and without my medical care of legal rights being affected. ☐

Name of Patient	Signature	Date
Name of Person taking consent	Signature	Date

CONTACT DETAILS

I confirm that, if the results of this blood test indicate that I am a suitable participant for the study, I am happy for a study researcher to telephone me using the numbers I have given below.

Phone _____

Mobile _____

My preferred method of contact is: Landline / Mobile (please delete as applicable)

My preferred time of contact is: Morning (9am – 1pm)

Afternoon (1pm – 5pm)

Evening (5pm – 8pm)

If you require any further information or have any questions at this stage, please contact the Study Co-ordinator at King's College London on 020 7848 4149 who will be happy to help.

Appendix 10.7 Ultrasound Examination Data Collection Form

Study ID: …… Patient's Name: ……..
Date of report …. / …. / ……

A. Liver

Size normal		☐ Yes	☐ No	
If abnormal is it	☐ Small	☐ Large		
Shape probably normal	☐ Yes	☐ No		
Parenchyma normal		☐ Yes	☐ No	
If abnormal Focal lesion(s)		☐ Yes	☐ No	Size ……..
Are they		☐ Solid	☐ Cystic	
Are they		☐ Single	☐ Multiple	
Type		☐ Most likely benign	☐ Uncertain, possibly sinister	
		☐ Most likely malignant		

Diffuse changes	☐ Yes	☐ No		
Diffuse changes - fatty	☐ Mild	☐ Moderate	☐ Marked	
	☐ No			
Diffuse changes - cirrhosis		☐ Yes	☐ No	
Portal vein visualised		☐ Yes	☐ No	
If visualised was it		☐ Normal	☐ Abnormal	
Any other abnormality		……		

B. Gallbladder

Seen	☐ Yes	☐ No
Normal	☐ Yes	☐ No
If abnormal, Stone(s)	☐ Yes	☐ No
Other abnormalities	☐ Yes	☐ No
Please specify	……	

C. Bile Ducts

Extrahepatic	☐ Dilated	☐ Non dilated	
Intrahepatic	☐ Dilated	☐ Non dilated	
Stone(s)	☐ Yes	☐ No	
Obstruction level	☐ Top duct	☐ Middle duct	☐ Lower duct
Other abnormalities	☐ Yes	☐ No	
Please specify	……		
Mass	☐ Yes	☐ No	
Mass in pancreas	☐ Yes	☐ No	
Mass in duct	☐ Yes	☐ No	
Other, specify	……		

D. Ascites

Ascites	☐ Yes	☐ No

E. Spleen

Seen	☐ Yes	☐ No	
Normal size	☐ Yes	☐ No	
Enlarged	☐ Yes	☐ No	Size ……..
Other abnormalities	☐ Yes	☐ No	
Please specify	……		

Overall comments if not covered above
……
……

Appendix 10.8 Consolidated Report Form

BALLETS REPORT	**Patient Study ID**	
1. Practice Patient ID		
2. GP Name		
3. Patient Family Name		
4. Other names		
5. DOB		
6. Gender		
7. NHS Number		
8. Current and recent medication (including vitamins and herbal remedies)		
9. Reason for consultation?	a) b) Signs: □ Jaundice □ Dark urine □ Pale stools c) Symptoms: □ Abdominal pain	
10. Reason for ordering LFTs		
11. Who ordered LFTs?		
12. Past Illnesses		
13. Recent Febrile Illness		
14. Alcohol Consumption	Units per week over past 6 months	
15. Substance Abuse	□ Intravenous □ Past □ Current	
16. Recent Travel History	Over last 6 months?	Where?
17. Immunisation history	Hep A □ Hep B □	
18. Transfusion history	□ No □Yes	Date:
19. How long have you been resident in the UK?		
20. Height (cms)		
21. Weight (kgs)		
22. Abdominal Girth		

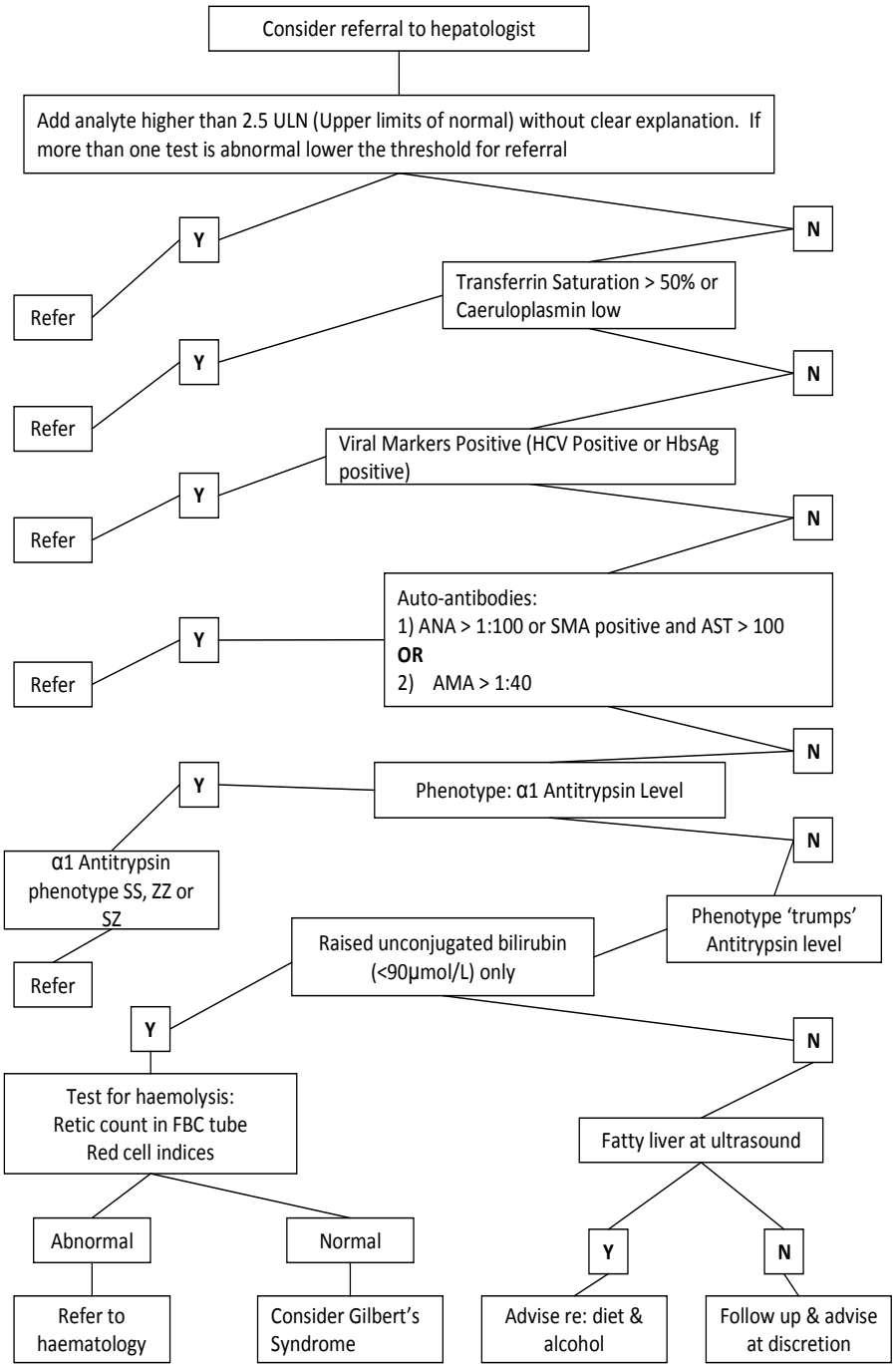
	Index Test Results	**Repeat LFTs**
	DATE:	**DATE:**
Albumin (34 - 51 g/L)		
Bilirubin (1 - 22 µmol/L)		
Globulin (21-37 g/L)		
AST (3 - 43 U/L)		
ALT (5 - 41 U/L)		
ALP (70 – 330 U/L)		
γ-GT (9 – 50 U/L male) (9 – 40 U/L female)		
Total Protein (60 - 80 µmol/L)		
Further Disease Testing		
Hepatitis B Viral Markers (HBV Surface Ag)		
Hepatitis C Virus Antibody (HCV Ab)		
Alpha-1 antitrypsin		Phenotyping to
Caeruloplasmin		
Iron & Transferrin		
Smooth muscle Ab		
Antimitochondrial Ab		

Ultrasound Examination

DATE:

Appendix 10.9 Hepatology Referral Guidelines

Disclaimer: These are guidelines for abnormal LFTs not rules and do not override clinical judgement.



Appendix 10.10.a Patient Interview Information Sheet

BALLETS: Birmingham and Lambeth Liver Evaluation

Testing Strategies

KING'S College LONDON
University of London

PATIENT INTERVIEW INFORMATION SHEET (23/02/06)

What is the purpose of the interview?
Liver function tests are widely used by doctors. Many tests have an abnormal result. We would like to understand how patients view these results. This will help us to improve patient care.

Why me?

Your doctor's surgery has agreed to help us with this research. People who have received an abnormal liver function test from this surgery have been approached to see if they would be able to take part in a short interview.

Do I have to take part?
No. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the care you receive from your practice.

What happens if I take part?
You will be interviewed by a researcher. The interviewer will ask you what you think about the test result you have received. The interview will take between 15 minutes and ½ an hour depending on how much you want to say. The interview could take place at your home or at the medical practice, whichever you would prefer.

What do I have to do to take part?
Complete and sign the consent form that you will be given.

Additional Points
Because we don't want to miss any information, we would like to tape record the interview.

Is it confidential?
All information collected during this research will be kept strictly confidential. The tapes will be locked in a cupboard and listened to only by the person who transcribes them.

Who is organising and funding the research?
This study is funded by the National Health Service Health Technology Assessment Programme.

General Information about Research
Consumers for Ethics in Research (CERES) publish a leaflet entitled 'Medical Research and You'. This leaflet gives more information about medical research and looks at some questions you may want to ask. Please ask us for a copy, or if you wish, a copy may be obtained from CERES, PO Box 1365, London N16 0BW.

If you have any questions about the study please contact Gareth Hollands, researcher, who would be happy to discuss this with you. Phone: 020 7188 2606, email: g.hollands@iop.kcl.ac.uk or write to: Psychology Department (at Guy's), Institute of Psychiatry, 5th Floor Thomas Guy House, Guy's Campus, London SE1 9RT.

Appendix 10.10.b Patient Interview Consent Form

KING'S College LONDON
University of London

BALLETS: Birmingham and Lambeth Liver Evaluation

Testing Strategies

PATIENT INTERVIEW CONSENT FORM

Please initial boxes below

I confirm that I have read and understand the interview information sheet (dated 23/02/06) and have had the opportunity to ask questions. ☐

I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

I agree to take part in the interview. ☐

I agree to have the interview tape-recorded. ☐

_____	_____	_____
Name of patient	Signature	Date
_____	_____	_____
Name of person taking consent	Signature	Date
_____	_____	_____
Researcher	Signature	Date

1 for patient; 1 for researcher; 1 to be kept with GP notes

Appendix 10.10.c Patient Views Questionnaire 1

BALLETS

Birmingham and Lambeth Liver Evaluation and Testing Strategies

UNIVERSITY OF BIRMINGHAM

KING'S College LONDON

University of London

Liver Function Tests – Patient views

Thank you for agreeing to take part in this research. We would like you to fill in this questionnaire and take it with you to your study appointment.

Liver function tests are widely used by doctors and sometimes these tests have an abnormal result which falls only slightly outside the normal range. We would like to understand how patients view the results that they receive. This will help us to improve patient care.

Please remember there are no right or wrong answers. If you cannot answer a particular question, don't worry: just leave it and carry on to the next one.

All information collected in this questionnaire will be treated with the strictest confidence.

For more information on the study please contact Susan Hoult Robinson, who is organising the study for King's College London.

Phone: (020) 7848 4149

Email: s.hoult-robinson@kcl.ac.uk

Address: Department of General Practice and Primary Care

5 Lambeth Walk

London SE11 6SP

Study Number:

Answering the Questionnaire

Some questions in this questionnaire ask you to circle a number that most applies to you.

Example:

If you feel fairly happy, you would answer the following question like this:

Question: How do you feel today?

Not happy 1 2 3 4 5 6 7 Very happy

Thank you very much for your help with this research

Section 1: Your liver function test result

[1] What do you think your liver function test result means? (please tick one box)

I **definitely** have problems with my liver ☐

It is **likely** that I have problems with my liver ☐

I **might** have problems with my liver ☐

It is **unlikely** that I have problems with my liver ☐

I **definitely do not** have problems with my liver ☐

I **do not know** what this result means for the health of my liver ☐

[2] Please describe the thoughts that come to mind when you think about the results of your liver function test result (please write below)

[3] How well do you feel you understand what your liver function test result reveals about the health of your liver? (please circle one number)

Do not understand at all	1	2	3	4	5	6	7	Completely understand

[4] To what extent do you feel you have a clear picture of what your liver function test result tells you about the health of your liver? (please circle one number)

Not at all clear	1	2	3	4	5	6	7	Completely clear

[5] How well informed did you feel about your liver function test result? (please circle one number)

Not at all informed	1	2	3	4	5	6	7	Extremely well informed

[6] How satisfied were you with the amount of information you were given about your liver function test result? (please circle one number)

Not at all satisfied	1	2	3	4	5	6	7	Extremely satisfied

[7] How confusing was the information you were given about your liver function test result? (please circle one number)

Not at all confusing	1	2	3	4	5	6	7	Extremely confusing

[8] How clear was the information you were given about your liver function test result? (please circle one number)

Not at all clear	1	2	3	4	5	6	7	Extremely clear

Section 2: Your test result and the health of your liver

[9] How concerned are you about the health of your liver? (please circle one number)

Not at all concerned	1	2	3	4	5	6	7	Extremely concerned

[10] How worried are you about the health of your liver? (please circle one number)

Not at all worried	1	2	3	4	5	6	7	Extremely worried

[11] If you follow the medical advice that you are given, how likely is it that you will have problems with your liver in the future? (please circle one number)

Not at all likely	1	2	3	4	5	6	7	Extremely likely

[12] How serious do you think it would be if you were to have problems with your liver in the future? (please circle one number)

Not at all serious	1	2	3	4	5	6	7	Extremely serious

For each of the following statements, please circle the number that best corresponds to your views.

		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
13.	What I do can determine whether I have problems with my liver	1	2	3	4	5
14.	Problems with my liver will last for a long time	1	2	3	4	5
15.	Problems with my liver could cause difficulties for those who are close to me	1	2	3	4	5
16.	There is a lot I can do to control the health of my liver	1	2	3	4	5
17.	Problems with my liver will last for a short time	1	2	3	4	5
18.	Problems with my liver could have major consequences for my life	1	2	3	4	5

[19] Have you had any symptoms that you think might be related to problems with your liver?

Yes ☐ No ☐

If yes, please list them here:

_____

_____

_____

_____

_____

_____

How important do you think each of the following factors is in causing your liver function test result? Please circle a number for each factor.

		Not at all important						Extremely important
20.	My diet	1	2	3	4	5	6	7
21.	Smoking	1	2	3	4	5	6	7
22.	Stress	1	2	3	4	5	6	7
23.	My genes	1	2	3	4	5	6	7
24.	Lack of exercise	1	2	3	4	5	6	7
25.	Drinking alcohol	1	2	3	4	5	6	7
26.	A germ or virus	1	2	3	4	5	6	7
27.	Medication I take	1	2	3	4	5	6	7
28.	False / inaccurate test result	1	2	3	4	5	6	7

[29] We would like you to list three images that you immediately associate with a particular topic. These may be single words, or small phrases. It is important that you do this quickly— it is your immediate impressions that we are interested in.

Think for a moment about: **Problems with your liver**.

What are the first three images that come to your mind when you think about this condition?
Please list these images below:

1)

2)

3)

Now we would like you to rate how vivid (intense, clear) your images were. Please circle one number for each image using the following scale, ranging from **"no image at all (you only "know" that you are thinking of something)"** to **"perfectly clear and as vivid as normal vision."**

	No Image At All	Vague and Dim	Somewhat Vivid	Reasonably Vivid	Perfectly Clear and Vivid
Image 1	1	2	3	4	5
Image 2	1	2	3	4	5
Image 3	1	2	3	4	5

Section 3: Your current mood

A number of statements which people have used to describe themselves are given below. Read each statement and then circle the most appropriate number to the right of the statement to indicate how you feel **right now, at this moment**. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

Please circle a number for **each** statement.

		Not at all	Somewhat	Moderately	Very much
30.	I feel calm	1	2	3	4
31.	I am tense	1	2	3	4
32.	I feel upset	1	2	3	4
33.	I feel relaxed	1	2	3	4
34.	I feel content	1	2	3	4
35.	I am worried	1	2	3	4

Section 4: Your test result and health behaviour

[36] Since you received your liver function test result, have you changed the amount of fat in your diet? (please tick one box)

No, it is about the **same amount** as before ☐
Yes, I am eating **more** fat in my diet now ☐
Yes, I am eating **less** fat in my diet now ☐

[37] Since you received your liver function test result, have you changed the amount of physical activity you do? (please tick one box)

No, I do about the **same amount** as before ☐
Yes, I am doing **more** physical activity now ☐
Yes, I am doing **less** physical activity now ☐

[38] Since you received your liver function test result, have you changed the amount you smoke? (please tick one box)

- If you are not a smoker or were not a smoker before the test result, please tick the last option

No, I am smoking about the **same amount** as before ☐
Yes, I am smoking **more** now ☐
Yes, I am smoking **less** now ☐
Yes, I have **completely stopped** smoking now ☐
I am not a smoker / I was not a smoker before the test result ☐

Section 5: Your test result and alcohol

[39] Have you drunk any alcohol in the past 3 months?

Yes ☐ No ☐

- **If No, please ignore all the remaining questions (40-45)**

[40] "Consuming no alcohol or only a little alcohol is an effective way of keeping your liver healthy". Do you agree or disagree with this statement? (please circle one number)

Strongly disagree	1	2	3	4	5	6	7	Strongly agree

[41] "Reducing my alcohol consumption will help my liver to stay healthy". Do you agree or disagree with this statement? (please circle one number)

Strongly disagree	1	2	3	4	5	6	7	Strongly agree

[42] How much do you feel you understand how reducing your consumption of alcohol might help your liver remain healthy? (please circle one number)

Do not understand at all	1	2	3	4	5	6	7	Completely understand

[43] To what extent do you feel you have a clear picture of how reducing your alcohol consumption might help your liver remain healthy? (please circle one number)

Not at all clear	1	2	3	4	5	6	7	Completely clear

[44] Since you received your liver function test result, how much alcohol are you drinking in comparison to before the result? (please tick one box)

About the **same amount** as before ☐
Much more alcohol now ☐
Slightly more alcohol now ☐
Much less alcohol now ☐
Slightly less alcohol now ☐

[45] For each day of the past week please indicate

- **what types of alcoholic drinks you drank (type and size)**
- **how many of each drink you had**

➢ **If it helps you to remember, write the days of the week in the spaces provided**

➢ **If you did not have any given drink please indicate this with a zero (0)**

PLEASE ANSWER THIS QUESTION ON PAGES 9 AND 10

How many drinks did you have yesterday?
This day was _____

Type of alcoholic drink	Size of the drink	Number of drinks
Beer / lager	Half-pint	
	Pint	
	Can (500ml)	
Cider	Half-pint	
	Pint	
	Can (500ml)	
Wine	Glass	
	Bottle (75cl)	
	Half-bottle (37.5cl)	
Spirits	Single measure (25ml)	
	Double measure (50ml)	
Others (please write in)		

How many drinks did you have 2 days ago?
This day was _____

Type of alcoholic drink	Size of the drink	Number of drinks
Beer / lager	Half pint	
	Pint	
	Can (500ml)	
Cider	Half-pint	
	Pint	
	Can (500ml)	
Wine	Glass	
	Bottle (75cl)	
	Half-bottle (37.5cl)	
Spirits	Single measure (25ml)	
	Double measure (50ml)	
Others (please write in)		

How many drinks did you have 3 days ago?
This day was _____

Type of alcoholic drink	Size of the drink	Number of drinks
Beer / lager	Half-pint	
	Pint	
	Can (500ml)	
Cider	Half-pint	
	Pint	
	Can (500ml)	
Wine	Glass	
	Bottle (75cl)	
	Half-bottle (37.5cl)	
Spirits	Single measure (25ml)	
	Double measure (50ml)	
Others (please write in)		

How many drinks did you have 4 days ago?
This day was _____

Type of alcoholic drink	Size of the drink	Number of drinks
Beer / lager	Half pint	
	Pint	
	Can (500ml)	
Cider	Half-pint	
	Pint	
	Can (500ml)	
Wine	Glass	
	Bottle (75cl)	
	Half-bottle (37.5cl)	
Spirits	Single measure (25ml)	
	Double measure (50ml)	
Others (please write in)		

How many drinks did you have 5 days ago?
This day was _____

Type of alcoholic drink	Size of the drink	Number of drinks
 Beer / lager	Half-pint	
	Pint	
	Can (500ml)	
 Cider	Half-pint	
	Pint	
	Can (500ml)	
 Wine	Glass	
	Bottle (75cl)	
	Half-bottle (37.5cl)	
 Spirits	Single measure (25ml)	
	Double measure (50ml)	
 Others (please write in)		

How many drinks did you have 6 days ago?
This day was _____

Type of alcoholic drink	Size of the drink	Number of drinks
 Beer / lager	Half pint	
	Pint	
	Can (500ml)	
 Cider	Half-pint	
	Pint	
	Can (500ml)	
 Wine	Glass	
	Bottle (75cl)	
	Half-bottle (37.5cl)	
 Spirits	Single measure (25ml)	
	Double measure (50ml)	
 Others (please write in)		

How many drinks did you have a week ago?
This day was _____

Type of alcoholic drink	Size of the drink	Number of drinks
 Beer / lager	Half-pint	
	Pint	
	Can (500ml)	
 Cider	Half-pint	
	Pint	
	Can (500ml)	
 Wine	Glass	
	Bottle (75cl)	
	Half-bottle (37.5cl)	
 Spirits	Single measure (25ml)	
	Double measure (50ml)	
 Others (please write in)		

Many thanks for answering this questionnaire

Appendix 10.10.d Follow-up Psychology and Lifestyle Questionnaire

BALLETS

Birmingham and Lambeth Liver Evaluation and Testing Strategies

UNIVERSITY OF BIRMINGHAM

KING'S College LONDON

University of London

Liver Function Tests – Patient views

Thank you for agreeing to take part in this research. We would like you to fill in this questionnaire and take it with you to your study appointment. This is the final questionnaire we will ask you to fill out.

Please remember there are no right or wrong answers. If you cannot answer a question, don't worry: just leave it and carry on to the next one.

All information collected in this questionnaire will be treated with the strictest confidence.

Questionnaire Completed:

(Please tick one box).

Before appointment ☐

At appointment ☐

For more information, please contact Ruth Collins who is organising the study for King's College London.

Phone: (020) 7188 9558

Email: ruth.e.collins@iop.kcl.ac.uk

Address: Department of General Practice and Primary Care

King's College London

5 Lambeth Walk

SE11 6SP

Study Number:

Answering the Questionnaire

Some questions in this questionnaire ask you to circle a number that most applies to you.

Example:

If you feel fairly happy, you would answer the following question like this:

Question: How do you feel today?

Not happy 1 2 3 (4) 5 6 7 Very happy

Thank you very much for your help with this research

Section 1: Your health

1. In general, would you say that your health is:

- Excellent ☐
- Very Good ☐
- Good ☐
- Fair ☐
- Poor ☐

2. How TRUE or FALSE is each of the following statements for you? (please circle one number for each statement).

		Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
a.	My health is excellent	1	2	3	4	5
b.	I expect my health to get worse	1	2	3	4	5
c.	I am as healthy as anybody I know	1	2	3	4	5
d.	I seem to get sick a little easier than other people	1	2	3	4	5

3. How concerned are you about the health of your liver? (please circle one number)

Not at all concerned 1 2 3 4 5 6 7 Extremely concerned

4. How worried are you about the health of your liver? (please circle one number)

Not at all worried	1	2	3	4	5	6	7	Extremely worried

Section 2: Your current mood

A number of statements which people have used to describe themselves are given below. Read each statement and then circle the most appropriate number to the right of the statement to indicate how you feel **right now, at this moment**. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

Please circle one number for **each** statement.

		Not at all	Somewhat	Moderately	Very much
5.	**I feel calm**	1	2	3	4
6.	**I am tense**	1	2	3	4
7.	**I feel upset**	1	2	3	4
8.	**I feel relaxed**	1	2	3	4
9.	**I feel content**	1	2	3	4
10.	**I am worried**	1	2	3	4

Section 3: Your test result and health behaviour

The following section is concerned with your health related behaviour since you received your liver function test results about two years ago. Please answer the questions considering your behaviour **in comparison** to your behaviour prior to receiving these results. There are no right or wrong answers. Please tick one box.

11. Since you received your liver function test result, have you changed the amount of fat in your diet? (please circle one number)

No, it is about the **same amount** as before ☐
Yes, I am eating **more** fat in my diet now ☐
Yes, I am eating **less** fat in my diet now ☐

12. Since you received your liver function test result, have you changed the amount of physical activity you do? (please circle one number)

No, I do about the **same amount** as before ☐
Yes, I am doing **more** physical activity now ☐
Yes, I am doing **less** physical activity now ☐

13. Do you smoke?

Yes ☐ No ☐

If NO please go to question 15.

14. Since you received your liver function test result, have you changed the amount you smoke? (please circle one number)

No, I am smoking about the **same amount** as before	☐
Yes, I am smoking **more** now	☐
Yes, I am smoking **less** now	☐
Yes, I have **completely stopped** smoking now	☐

15. Have you drunk any alcohol in the past 3 months?

Yes ☐ No ☐

16. Since you received your liver function test result, how much alcohol are you drinking in comparison to before the results? (please circle one number)

About the **same amount** as before	☐
Much more alcohol now	☐
Slightly more alcohol now	☐
Much less alcohol now	☐
Slightly less alcohol now	☐

Please check that you have completed all the questions and bring the questionnaire to your study appointment. Please use the space below to write any comments on your experiences of tests on your liver.

If you would like to discuss anything about this please contact:

Ruth Collins

Phone: (020) 7188 9558

Email: ruth.e.collins@iop.kcl.ac.uk

Address: Department of General Practice and Primary Care

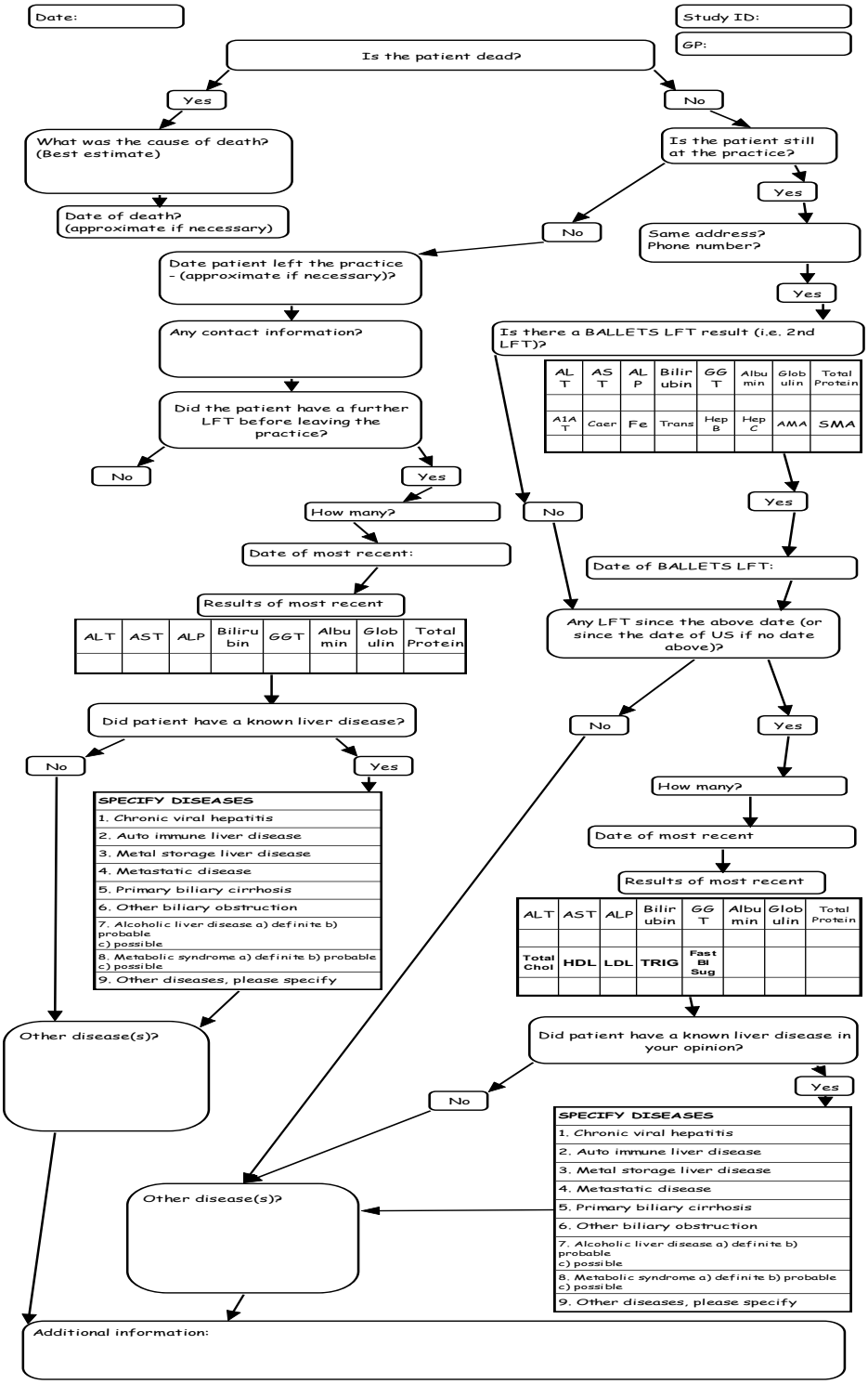
King's College London

5 Lambeth Walk

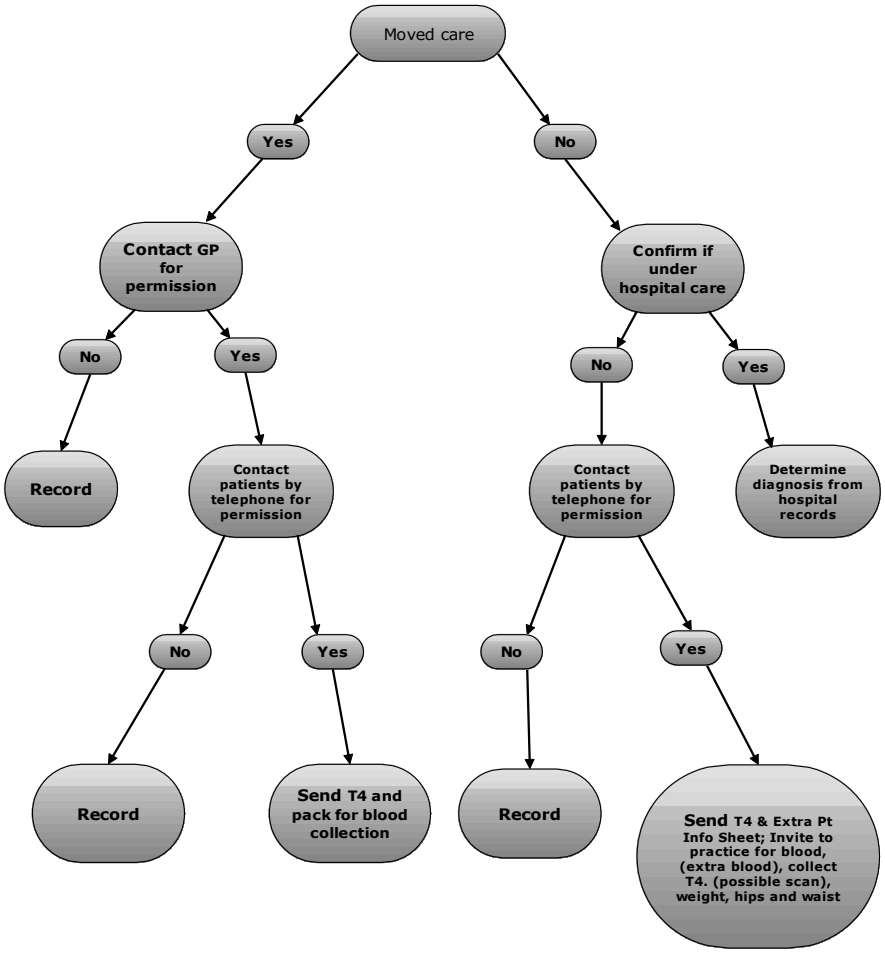
SE11 6SP

Thank you very much for your continued help with this research.

Appendix 10.11.a Data Collection Form



Appendix 10.11.b Long-term follow-up protocol for patients



Appendix 10.11.c Extra Patient Information Sheet

UNIVERSITY OF BIRMINGHAM

EXTRA PATIENT INFORMATION SHEET

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)

A study to evaluate the value of Liver Function Tests (LFTs)

Two years ago you took part in the BALLETS study. As you will recall, the study, sponsored by the NHS Heath Technology Assessment Programme, is designed to investigate the value of liver function tests. During the past 2 years, the study has recruited over 1100 patients at GP practices in Birmingham and Lambeth. As you know GPs refer patients to the study if they have a mildly abnormal liver function test and are considered clinically suitable (i.e. over 18 years old, not pregnant and with no liver disease).

At the study clinic, after being formally consented, patients have further blood tests, an ultrasound scan and a short interview. During this appointment, the research team asks if patients would be prepared to attend a two-year follow up appointment to have another liver function blood test.

Because of some interesting findings during the study, the research team would like to repeat the ultrasound scan and would also like to collect and freeze a second blood sample while we are taking blood for the follow-up liver function test. Before you decide whether or not you would like to give this additional blood sample, it is important for you to understand why the add-on study is being done. Patients will not need an additional venepuncture (i.e. needle to remove blood) since we will freeze a sample of the blood taken at the same time as the planned liver function test. Please take time to read the following information carefully and ask us if you would like more information.

What is the purpose of the add-on study?
Because of the high number of fatty livers (40%) detected in the ultrasound examinations of study patients, we would like to repeat the ultrasound scan and to examine more closely the blood samples of all patients, including those with and without a fatty liver. The study sponsor (HTA) has provided additional funding that will enable us to freeze a sample of blood, taken from study patients.

The blood sample will be prepared and stored at the Institute for Biomedical Research, Medical School, University of Birmingham, Vincent Drive, Edgbaston. The further testing will include a lipoprotein profile (the profile of 'good' and 'bad' fat in the blood); C-reactive protein, procollagen peptide 3 and hyaluronate (which are markers of inflammation and of fibrosis); and genes that may be associated with the development of liver disease.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?
When you come along to your follow-up appointment, you will be offered the opportunity to ask more questions about the add-on study and the follow-up appointment. You will also be asked to sign a consent form to take part in the add-on study, if you are happy to do so. The second sample of blood would be taken at the same time as the already agreed follow-up liver function test and would not involve a second venepuncture (i.e. needle to remove blood). An ultrasound scan will be done by the sonographer, in the same way as when you attended for your last BALLETS appointment.

What are the possible disadvantages and risks of taking part?
You will have to sign an additional consent form that will make the follow-up appointment around 15 minutes longer, however the additional blood sample will be taken at the same time as the two-year follow-up liver function test.

What are the possible benefits of taking part?
Although the follow up liver function test will be investigated thoroughly and reported back to you, we do not plan to feed the add-on study results back on an individual basis as testing will be anonymous.

At present we do not know what these additional tests mean for a patient's health and so hope to provide fuller guidance to GPs on their management.

What happens if you change your mind?
If in the future you decide you do not wish your samples to be stored – the samples will be destroyed.

What happens at the end of the study?
Your GP will continue to manage you according to his/her usual practice. The results of the study will be published in a report and in a medical journal and you will not be identified in any reports.

A blood sample you have gifted may be made available to researchers who may be in the UK or overseas. They may work in universities, hospitals or private commercial companies that do medical research. You will not receive any personal financial reward for making your gift. The University of Birmingham may ask researchers for fees to cover some of the costs it incurs but the samples you have gifted will never be sold for profit.

Will my taking part in this study be kept confidential?
The blood sample will be labelled with your study ID only. Any information about you that leaves the GP practice will have your name and address removed so that you cannot be recognised from it.

Contact for Further Information
Your GP can answer any questions you have about the study. Also, Prof Richard Lilford, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 2226
Email: r.j.lilford@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and your GP will keep a copy in your notes.

Thank you for your help with this study.

Appendix 10.11.d Extra Consent Form

UNIVERSITY OF BIRMINGHAM

Practice ID:	
Patient Number:	

EXTRA CONSENT FORM

BALLETS: Birmingham and Lambeth Liver Evaluation Testing Strategies

An additional study to collect and store blood for later testing in order to evaluate the value of liver function tests

Lead Researcher: Professor Richard Lilford

Please initial boxes below

1. I confirm that I have read and understand the extra information sheet dated 19th March 2008 (version 1.0) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐

3. I agree to give one additional sample of blood for research in this study. I understand how the sample will be collected, that giving a sample for this research is voluntary and that I am free to withdraw my approval for use of the sample at any time without giving any reason. ☐

4. I understand that the study using the sample I give will include genetic research aimed at understanding the genetic influences on liver disease, but the results of these investigations are unlikely to have any implications for me personally. ☐

5. I agree to take part in the follow up study. ☐

_____	_____	_____
Name of Patient	Date	Signature
_____	_____	_____
Researcher	Date	Signature

1 for patient; 1 for researcher; 1 to be kept with GP notes

Appendix 10.12.a GP Consent Form

UNIVERSITY OF BIRMINGHAM

Practice ID:	
Patient Number:	

GP CONSENT FORM

BALLETS: Birmingham and Lambeth Liver Evaluation Testing Strategies

Lead Researcher: Professor Richard Lilford

Sub study

Please initial boxes below

1. I confirm that I have read and understand the information sheet dated 12th November 2008 (version 1.0) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without my legal rights being affected. ☐

3. I understand that the sub study will involve either a short semi-structured interview to be conducted by telephone or a semi-structured interview to be conducted in person dependent upon my preference. The contents of the interview will be anonymised and will remain confidential. ☐

4. I agree to take part in the sub study. ☐

_____	_____	_____
Name of GP	Date	Signature
_____	_____	_____
Researcher	Date	Signature

1 for GP; 1 for researcher

Appendix 10.12.b Information Sheet for General Practitioners

UNIVERSITY OF BIRMINGHAM

INFORMATION SHEET FOR GENERAL PRACTIONERS

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)

A sub-study investigating liver function test ordering behaviour of general practitioners involved in the BALLETS study in Birmingham

Sub-study

As you will recall, the BALLETS study, sponsored by the NHS Heath Technology Assessment Programme, is designed to investigate the value of liver function tests. During the past 2 years, the study has recruited over 1100 patients at GP practices in Birmingham and Lambeth.

This sub-study has been commissioned to investigate the type and range of non-clinical reasons and motives, behind the decision of a general practitioner (GP) to order a liver function test (LFT). It is a qualitative study and all GPs at practices participating in the BALLET study will be asked to take part.

What is the purpose of the add-on study?
1. To determine and assess the non-clinical reasons and motives underlying a general practitioner's decision to order LFTs.

2. Use the information in conjunction with the findings of the primary BALLETS study to inform GP decision making and reduce the number of unnecessary tests.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason.

What will happen to me if I take part?
You will be asked to participate in a short semi-structured interview. The interview will be conducted in person or by telephone at your discretion. If the interview is conducted by telephone then it is expected to last no-longer than eight minutes. If you choose to be interviewed in person then it is possible that the interview could last longer. The interviews will be recorded to allow transcription and analysis but will remain confidential and only be seen by members of the study team.

What are the possible disadvantages and risks of taking part?
The length of time it takes to complete the interview.

What are the possible benefits of taking part?

The findings of this study will be used in future guidelines for test ordering and will therefore aid other GPs in their decision making process with the aim of reducing the number of unnecessary tests.

What happens if you change your mind?

If in the future you decide you do not wish to be included in the study then the data gained from your interview will be removed.

What happens at the end of the study?

The results of the study will be published in a report and in a medical journal. You will not be identified at any point.

Will my taking part in this study be kept confidential?

Any information about you will be anonymised, as will the identity of your practice.

Contact for Further Information

Prof Richard Lilford, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 2226
Email: r.j.lilford@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and you will keep the other.

Thank you for your help with this study.

Appendix 10.12.c References

REFERENCES

1 Beastall GH. The impact of the new General Medical Services contract: national evidence. *Bull R Coll Pathol* 2004; **128**:24.
2 Galloway MJ, Osgerby JC. An audit of the indications for the reporting of blood films: results from the National Pathology Benchmarking Study'. *J Clin Pathol* 2006; **59**:479-481.
3 Coral BD. Laboratory demand management'. *The bulletin of the royal college of pathologists* 2007; **139**.
4 Murphy MK, Sanderson CFB, Black NA, Askham J, Lamping DL, Marteau T, *et al*. Consensus development methods and their use in clinical guideline development. *Health Technology Assessment* 1998; **2**(3).
5 Washington DC: Office of Technology Assessment. Defensive medicine and medical malpractice. 1994. Report No.:
6 Summerton N. Positive and negative factors in defensive medicine: a questionnaire study of general practitioners. *BMJ* 1995; **310**:27-29.
7 Guthrie B. Why do general practitioners take blood. *European Journal of General Practice* 2001; **7**(138):142.
8 Lundberg GD. Perseverance of laboratory test ordering: a syndrome affecting clinicians. *JAMA* 1983; **249**:639.
9 Axt-Adam P, van der Wouden JC, van der Does E. Influencing behaviour of physicians ordering laboratory tests: A literature study. *Medical Care* 1993; **31**(9):784-794.
10 Wileman L, May C, Chwe-Graham CA. Medically unexplained symptoms and the problem of power in the primary care consultaion: A qualitative study. *Fam Pract* 2002; **19**(2):178-182.
11 van der WT, van Bokhoven MA, Dinant GJ, van Hasselt CM, Grol RP. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract* 2002; **52**(485):974-980.
12 van der WT, van VM, Dinant GJ, van Hasselt CM, Grol R. Unexplained complaints in general practice: prevalence, patients' expectations, and professionals' test-ordering behavior. *Med Decis Making* 2003; **23**(3):226-231.
13 Sheehan JD. The management of medically unexplained symptoms. *Sem Gastrointest Dis* 1999; **10**:30-36.
14 Reid S, Whooley D, Crayford T, Hotopf M. Medically unexplained symptoms - GPs' attitudes towards their cause and management. *Fam Pract* 2001; **18**(5):519-523.
15 Vinker S, Kvint I, Erez R, Elhayany A, Kahan E. Effect of the characteristics of family physicians on their utilisations of laboratory tests. *British Journal of General Practice* 2007; **57**(377):382.
16 Kristiansen IS, Hjortdahl P. The general practitioner and laboratory utilization: Why does it vary? *Family Practice* 1992; **9**(1):22-27.
17 Smellie WS, Clark G, McNulty CAM. Inequalities of primary care microbiology testing between hospital catchment areas. *J Clin Pathol* 2003; **56**:933-936.
18 Smellie WS, Galloway MJ, Chinn D, Gedling P. Is clinical practice variability the major reason for differences in pathology requesting patterns in general practice? *J Clin Pathol* 2002; **55**:312-314.
19 Zaat JO, van Eijk JT, Bonte HA. Laboratory test form design influences test ordering by general practitioners in The Netherlands. *Med Care* 1992; **30**(3):189-198.
20 Bailey J, Jennings A, Parapia L. Change of pathology request forms can reduce unwanted requests and tests. *J Clin Pathol* 2005; **58**(8):853-855.
21 van Walraven C, Cernat G, Austin PC. Effect of provider continuity on test repetition. *Clin Chem* 2006; **59**(12):5219-5228.
22 Denzin and Lincoln. Handbook of qualitative research. 2nd ed. Thousand Oaks, Ca; 2000.
23 Hartley RM, Charlton JR, Harris CM, Jarman B, Hartley RM, Charlton JR, *et al*. Influence of patient characteristics on test ordering in general practice. *British Medical Journal Clinical Research Ed* 1984; **. 289**(6447):735-738.

24 van der Weijden T, van Velsen M, Dinant G-J, van Hasselt M, Grol R. Unexplained complaints in general practice:prevalence, patients' expectations and professionals test-ordering behaviour. *Medical Decision Making* 2003; **23**:226-231.
25 van Bokhoven MA, Pleunis-van Empel MCH, Koch H, Grol R, Dinant G-J, van der Weijden T. Why do patients have their blood tested? A qualitative study of patient expectations in general practice. *Family Practice* 2006; **7**:75.
26 Black N. Medical litigation and the quality of care. *Lancet* 1990; **335**:35-37.
27 Dowell J, Pitkethley M, Bain J, Martin S. a randomised controlled trial of delayed antibiotic prescribing as a strategy for managing uncomplicated respiratory tract infection in primary care. *Br J Gen Pract* 2001; **51**:200-205.
28 Sibbald B, Addington-Hall J, Brenneman D, Freeling P. Telephone vs postal surveys of general practitioners: methodological considerations. *Br J Gen Pract* 1994; **44**:297-300.

Appendix 10.12.d List of Practices

List of practices

Practices to be approached for sub-study. Principal investigator and lead GP relates to the primary BALLETS study.

Practice	Principal Investigator	Practice Manager
Hall Green Health, 979 Stratford Road, B28 8BG GP telephone: 0121 325556	Dr Bill Strange	Chris Jenkins
Lordswood House, 54 Lordswood Road, B17 9DB Tele: 0121 426 2030	Dr Ewan Hamnett	Joyce Marriott
Greenridge Surgery, 713 Yardley Wood Road, B13 0PT Tele: 0121 4443597	Dr Richard McManus *Lead GP: Dr Lak Jhass*	Fay Staff
Yardley Wood Health Centre, 401 Highfield Road, B14 4DU Telephone: 0121 474 5186	Dr Peter Clarke	Angela Styring
Woodland Road Surgery 57 Woodland Road, Northfield B31 2HZ Tele: 0121 4751065	Dr David Taylor	Jenny Morgan
Cofton Medical Centre, 2 Robinsfield Drive, West Heath B31 4TU Tele: 0121 693 5777	Dr Victoria Lloyd *Lead GP: Kirstie Blackford*	Julie Walker
Wand Medical Centre, 15 Frank St, Highgate B12 0UF Tele: 0121 4401561	Dr Adam Fraser	Tracey Gardner
Shenley Green Surgery , 22 Shenley Green, Selly Oak, B29 4HH Tele: 0121 4757997	Dr Andres Puga	Vicky Chambers

Appendix 10.13.a – Primary Biliary Cirrhosis letter to GPs

Dear Dr

Re: BALLETS study - Elevated mitochondrial antibodies
Patient:

Firstly, thank you again for your participation in the BALLETS study; it is much appreciated.

As you will be aware, your patient (identified above) has been found to have an elevated anti-mitochondrial antibody. This suggests that your patient may have or may develop Primary Biliary Cirrhosis (PBC).

PBC is a chronic cholestatic liver disease, characterised by progressive, immune-mediated destruction of the middle-sized intra-hepatic ducts that may lead to cirrhosis and, rarely, liver failure. The condition typically affects middle-aged women and may be associated with other auto-immune diseases. The condition is often asymptomatic but is typically associated with lethargy and itching. Liver tests are typically abnormal (with a raised alkaline phosphatase) but, especially in the early stages, may be normal. The hall-mark of the condition is the AMA: this may be present before abnormal liver tests develop. The significance of an isolated AMA remains controversial but recent epidemiological data suggest that in those with AMA at a titre of 1:40 or more on two occasions will either have PBC or else develop PBC over the subsequent years. It may nevertheless follow a benign course, and only come to light as a result of medical investigations for other reasons.

A diagnosis of PBC carries many implications for the patient who is often asymptomatic at the time when the AMA are detected. Early detection may lead to treatments for the symptoms and surveillance for the possible complications and allow for early treatment which is safe and may improve the outcome, but it may transform a healthy person into a 'patient' so informing the patient and provision of appropriate information is important.

You may, therefore, wish to refer the person for further evaluation to a local liver unit or else wish to discuss this with your local hepatologist. BALLETS study hepatologist, Professor James Neuberger, would be delighted to discuss the course of action with you.

For further information:
Liver Unit, Queen Elizabeth Hospital, Birmingham B15 2TH
Tel: 0121 627 2414

Guidelines from the American Association for the Study of Liver Disease:
http://www.aasld.org/practiceguidelines/Documents/Practice%20Guidelines/biliarycirrhosis.pdf

Patient support:
PBC Foundation http://www.pbcfoundation.org.uk/
Tel: 0131 225 8586
British Liver Trust http://www.britishlivertrust.org.uk/home.aspx
Tel: 0800 652 7330

Appendix 10.13.b – Wilson's disease letter to GPs

Dear Dr

Re: BALLETS study - Abnormal caeruloplasmin
Patient:

Firstly, thank you again for your participation in the BALLETS study; it is much appreciated.

As you will be aware, your patient (identified above) has been found to have an abnormal caeruloplasmin level (≤ 1.4g/l). While there are several causes for this finding, it raises the possibility that your patient may have Wilson's disease.

Wilson's disease is an autosomal recessive genetic disorder in which copper accumulates in the tissues. The copper accumulates because the protein in the blood which binds copper is insufficient. The diagnosis is often difficult to make and a useful screening test is to measure the protein caeruloplasmin in the blood but the specificity of this test is low, i.e. a high proportion of positives are 'false positives'. However, the diagnosis is important since early treatment may prevent development of life-threatening complications and, where appropriate, family members who are asymptomatic should be offered screening.

Wilson's disease is so rare, that the probability that the disease exists is low even after an abnormal test result; even the post-test probability of the disease is low. We are following up only those patients who are under 55 years of age and/or who have ultrasound signs of impending cirrhosis. The rationale for this lies partly in the low post-test probability of this disease and partly in the knowledge that Wilson's disease usually manifests before the 6th decade of life.

While you may wish to take this further yourself, you may wish to refer for further evaluation. We suggest referral to a local Liver Unit is appropriate. If it is acceptable to you, could you please refer to the BALLETS study hepatologist, Professor James Neuberger?

For further information:

Liver Unit, Queen Elizabeth Hospital, Birmingham B15 2TH
Tel: 0121 627 2414

Guidelines from the American Association for the Study of Liver Disease:
http://www.aasld.org/practiceguidelines/Documents/WilsonDisease2008.pdf

Patient support
British Liver Trust: http://www.britishlivertrust.org.uk/home.aspx
Tel: 0800 652 7330

Appendix 10.13.c – Haemochromatosis letter to GPs

Dear Dr

Re: BALLETS study - Elevated iron saturation
Patient:

Firstly, thank you again for your participation in the BALLETS study; it is much appreciated.

As you will be aware, your patient (identified above) has been found to have an elevated iron saturation. While there are several causes for this finding, it raises the possibility that your patient may have genetic haemochromatosis.

Genetic haemochromatosis is a genetic disease that is characterised, in some people, by progressive iron overload that may lead to cirrhosis and liver cell cancer, diabetes, cardiomyopathy, arthritis and damage to other organs. While there is as yet no cure, treatment by venesection is effective in preventing any, or further, organ damage. Because many people are asymptomatic until end-organ damage has occurred, it is important to make an early diagnosis.

Most people with genetic haemochromatosis can be detected by genetic analysis for the two common mutations. Where a case is identified, it is important not only to counsel and, where indicated, treat the person but also provide advice for family screening.

While you may wish to take this further yourself, you may wish to refer for further evaluation. We suggest referral to a local Liver Unit is appropriate. If acceptable to you, could you please refer to the study hepatologist, Professor James Neuberger (just for Birmingham patients)?

For further information:

Liver Unit, Queen Elizabeth Hospital, Birmingham B15 2TH
Tel: 0121 627 2414

Guidelines from the American Association for the Study of Liver Disease:
http://www.aasld.org/practiceguidelines/Documents/Practice%20Guidelines/hemochratosis.pdf

Patient support
British Liver Trust: http://www.britishlivertrust.org.uk/home.aspx
Tel: 0800 652 7330

Haemochromatosis Society UK
http://www.haemochromatosis.org.uk/home.html
Tel: 0208 449 1363

Appendix 10.13.d – Autoimmune hepatitis letter to GPs

Dear Dr

Re: BALLETS study - Abnormal smooth muscle antibodies
Patient:

Firstly, thank you again for your participation in the BALLETS study; it is much appreciated.

As you will be aware, your patient (identified above) has been found to have an abnormal smooth muscle antibodies. While there are several causes for this finding, it raises the possibility that your patient may have autoimmune hepatitis.

Autoimmune hepatitis is a chronic hepatitis of unknown aetiology characterized by immunologic and autoimmunologic features, generally including the presence of circulating autoantibodies and a high total serum and or gamma globulin, often IgG concentration.

While you may wish to take this further yourself, you may wish to refer for further evaluation. We suggest referral to a local Liver Unit is appropriate. BALLETS study hepatologist, Professor James Neuberger, would be delighted to discuss the course of action with you.

For further information:

Liver Unit, Queen Elizabeth Hospital, Birmingham B15 2TH
Tel: 0121 627 2414

Guidelines from the American Association for the Study of Liver Disease:
http://www.aasld.org/practiceguidelines/Documents/Practice%20Guidelines/autoimmune_hepatitis.pdf

Patient support
British Liver Trust: http://www.britishlivertrust.org.uk/home.aspx
Tel: 0800 652 7330

Appendix 10.14.a Patient Information Sheet

UNIVERSITY OF BIRMINGHAM

PATIENT INFORMATION SHEET

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)

A sub-study investigating patients' experience of taking part in the BALLETS study

As you will recall, the BALLETS study, sponsored by the NHS Health Technology Assessment Programme, is designed to investigate the value of liver function tests. Over the first 2 years, the study recruited over 1100 patients at GP practices in Birmingham and Lambeth. As you know GPs referred patients to the study if they had a mildly abnormal liver function test and were considered clinically suitable (i.e. over 18 years old, not pregnant and with no liver disease). At the study clinic, after being formally consented, patients had blood tests, an ultrasound scan and a short interview. You recently attended a follow-up clinic for the BALLETS study where you had a repeat blood test, ultrasound scan and short interview.

This sub-study has been commissioned to explore patients' experiences of taking part in the BALLETS study.

What is the purpose of the add-on study?

To explore patients' experiences of taking part in the BALLETS study to inform patients' motivational and decision making approaches to adopting healthy behaviours.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and will be asked to sign a consent form. If you decide to take part, you are still free to withdraw up to 2 weeks after the interview and without giving a reason.

What will happen to me if I take part?
You will be asked to participate in a short semi-structured interview which will last between 30-60 minutes. The interview will be conducted in person or by telephone, at your discretion. The interview can take place at your own home or the GP practice and will be at a convenient time of day for you. The interviews will be audio-recorded to allow transcription and analysis but will remain confidential and only be seen by members of the study team. The contents of the digital audiotape used during your interview, will be remain confidential and will be stored at the BALLETS study office at the University of Birmingham.

What are the possible disadvantages and risks of taking part?
The interview will take 30 to 60 minutes to complete.

What are the possible benefits of taking part?
The findings of this study will be used in the future to inform clinical staff of the best ways of working with patients found to have a fatty liver to help them adopt healthy behaviours.

What happens if you change your mind?
After completing the study, you are allowed to withdraw your data from the study up to 2 weeks after completing the interview. This is due to the nature of qualitative data analysis where interviews from different people are merged to form a basis for analysis.

What happens at the end of the study?
The results of the study will be published in a report and in a scientific journal. You will not be identified at any point.

Will my taking part in this study be kept confidential?
Any information about you will be anonymised, as will the identity of your practice.

Contact for Further Information
Dr Ian Litchfield, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 6006
Email: litchfii@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and you will keep the other.

Thank you for your help with this study.

Appendix 10.14.b Patient Consent Form

UNIVERSITY OF BIRMINGHAM

Practice ID:	
Patient Number:	

PATIENT CONSENT FORM

BALLETS: Birmingham and Lambeth Liver Evaluation Testing Strategies

A sub-study investigating patients' experience of taking part in the BALLETS study

Lead Researcher: Professor Richard Lilford

Please initial boxes below

1. I confirm that I have read and understand the extra information sheet dated 10th March, 2010 (version 1.0) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw up to 2 weeks after the interview without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that the sub study will involve either a short semi-structured interview to be conducted by telephone or a semi-structured interview to be conducted in person dependent upon my preference. The contents of the interview will be anonymised and will remain confidential. ☐

4. I agree to take part in this study. ☐

Name of Patient	Date	Signature
Researcher	Date	Signature

1 for patient; 1 for researcher; 1 to be kept with GP notes

Appendix 10.14.c BALLETS Qualitative Sub-study: Semi-structured Patient Interview Questions

BALLETS QUALITATIVE SUB-STUDY: SEMI-STRUCTURED PATIENT INTERVIEW QUESTIONS

1) Can you tell me about how you took part in the BALLETS study? (Ice-breaker)

Prompts: who approached you initially about the study? Where did you hear about the study?

2) Can you tell me about the first BALLETS appointment?

Prompts: What happened in the consultation? Who was present? What questions were you asked?

3) How did you feel after you heard the results of the ultrasound/LFT from the BALLETS study?

Prompts: Were the results what you expected?

4) After knowing the results from the BALLETS consultation, did you initiate any lifestyle changes? (leading questions)

Prompts: did you change your diet? How about your alcohol consumption? Did you change your patterns of physical exercise?
If no changes, focus on why no changes were initiated.

5) How did you feel when you were approached for a follow-up BALLETS consultation?

Prompts: were you apprehensive/ happy to come back for a follow-up? What thoughts went through your mind when you received a phone call/letter?

6) How did you feel the follow-up consultation went?

Prompts: What happened in the consultation? Who was present? What questions were you asked?

7) Now that you have completed the BALLETS study, reflecting back, how was your personal experience of taking part in the study?

8) Is there anything you would like to feedback to the BALLETS team?

Prompts: anything that you would have liked to have been different? Did you have any positive or negative experiences?

Appendix 10.14.d Sonographer Information Sheet

UNIVERSITY OF BIRMINGHAM

SONOGRAPHER INFORMATION SHEET

BALLETS (Birmingham and Lambeth Liver Evaluation Testing Strategies)

A sub-study investigating patients' experience of taking part in the BALLETS study

As you will recall, the BALLETS study, sponsored by the NHS Health Technology Assessment Programme, is designed to investigate the value of liver function tests. Over the first 2 years, the study recruited over 1100 patients at GP practices in Birmingham and Lambeth. As you know GPs referred patients to the study if they had a mildly abnormal liver function test and were considered clinically suitable (i.e. over 18 years old, not pregnant and with no liver disease). At the study clinic, after being formally consented, patients had blood tests, an ultrasound scan and a short interview.

This sub-study has been commissioned to explore patients' experiences of taking part in the BALLETS study. As part of the study we are also interviewing sonographers who were employed by the study to determine their opinions on the consultation process and the methods they used to impart the results of the scan and possible implications.

What is the purpose of the add-on study?

1. To explore patients' experiences of taking part in the BALLETS study to inform patients motivational and decision making styles with regards to adopting healthy behaviours.
2. To merge that data with the data obtained from Sonographers to understand how patients motivate themselves dependant upon the results they receive from ultrasound scans which may or may not be abnormal.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and will be asked to sign a consent form. If you decide to take part, you are still free to withdraw up to 2 weeks after the interview and without giving a reason.

What will happen to me if I take part?

You will be asked to participate in a short semi-structured interview which will last between 30-60 minutes. The interview will be conducted in person or by telephone

at your discretion. The interview can take place at your own home or at the University of Birmingham and will be at a time of the day convenient for you. The interviews will be audio-recorded to allow transcription and analysis but will remain confidential and only be seen by members of the study team. The contents of the digital audiotape used during your interview, will be remain confidential and will be stored in the BALLETS study office at the University of Birmingham.

What are the possible disadvantages and risks of taking part?
The length of time it takes to complete the interview.

What are the possible benefits of taking part?
Your input will enable us to learn more about the consultation process and the methods used to impart the results of an ultrasound scan.

What happens if you change your mind?
After completing the study, you are allowed to withdraw all data from the study up to 2 weeks after completing your interview. This is due to the nature of qualitative data analysis where interviews from different people are merged to form a basis for analysis.

What happens at the end of the study?
The results of the study will be published in a report and in a scientific journal. You will not be identified at any point.

Will my taking part in this study be kept confidential?
Any information about you will be anonymised.

Contact for Further Information
Dr Ian Litchfield, who is organising the study from the University of Birmingham, will be happy to speak to you if you have any questions.

Phone: 0121 414 6006
Email: litchfii@bham.ac.uk

You will be given a copy of this Information Sheet and a signed consent form to keep.
The researchers will keep one copy and you will keep the other.

Thank you for your help with this study.

Appendix 10.14.e Sonographer Consent Form

UNIVERSITY OF BIRMINGHAM

SONOGRAPHER CONSENT FORM

BALLETS: Birmingham and Lambeth Liver Evaluation Testing Strategies

A sub-study investigating patients' experience of taking part in the BALLETS study

Lead Researcher: Professor Richard Lilford

Please initial boxes below

1. I confirm that I have read and understand the extra information sheet dated 26th February, 2010 (version 1.0) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw up to 2 weeks after the interview without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that the sub study will involve either a short semi-structured interview to be conducted by telephone or a semi-structured interview to be conducted in person dependent upon my preference. The contents of the interview will be anonymised and will remain confidential. ☐

4. I agree to take part in this study. ☐

Name of Patient	Date	Signature
Researcher	Date	Signature

1 for sonographer; 1 for researcher

Appendix 10.14.f BALLETS Qualitative Sub-study: Semi-structured Sonographer Interview Questions

BALLETS QUALITATIVE SUB-STUDY: SEMI-STRUCTURED SONOGRAPHER INTERVIEW QUESTIONS

1. During a routine hospital-based consultation how much information would you impart to a patient about findings?
2. How did BALLETS study consultations compare to a hospital-based consultation?
3. What did you consider your study role to be?
4. How do you feel about general feedback to patients? Prompt: how about your relationship to patients?
5. How important is it to you that you feedback results from the scan?
6. Did you feel that study patients took on board feedback about the scan?

Appendix 2

BALLETS study analysis

Liver function test results by laboratory

Distribution of liver function test data

It is clear that many of the analytes in the standard panel of LFTs are subject to distributional skewness that precludes direct application of statistical methods based on the normal distribution. This feature is present in the results generated by all three of the laboratories that contributed to the study, and is evident from the histograms in *Figure 24*.

A further complication arises because of the possible influence of differences in laboratory practice on the results. The presence of such differences may be inferred from an inspection of normal reference ranges for individual analytes, particularly in the case of ALP (see main report, *Table 4*). In our analyses, the issue of skewness was addressed by means of a logarithmic transformation applied to all LFTs prior to analysis. This device also enables multiplicative interlaboratory effects to be modelled conveniently as additive effects on the log-scale.

Histograms of log-transformed LFTs are displayed in *Figure 25*. The effect of the transformation has been substantially to remove the skewness associated with the first five analytes (ALT, AST, bilirubin, ALP, GGT), without noticeably disturbing the relative symmetry of the distribution of the protein measures (albumin, globulin, total protein).

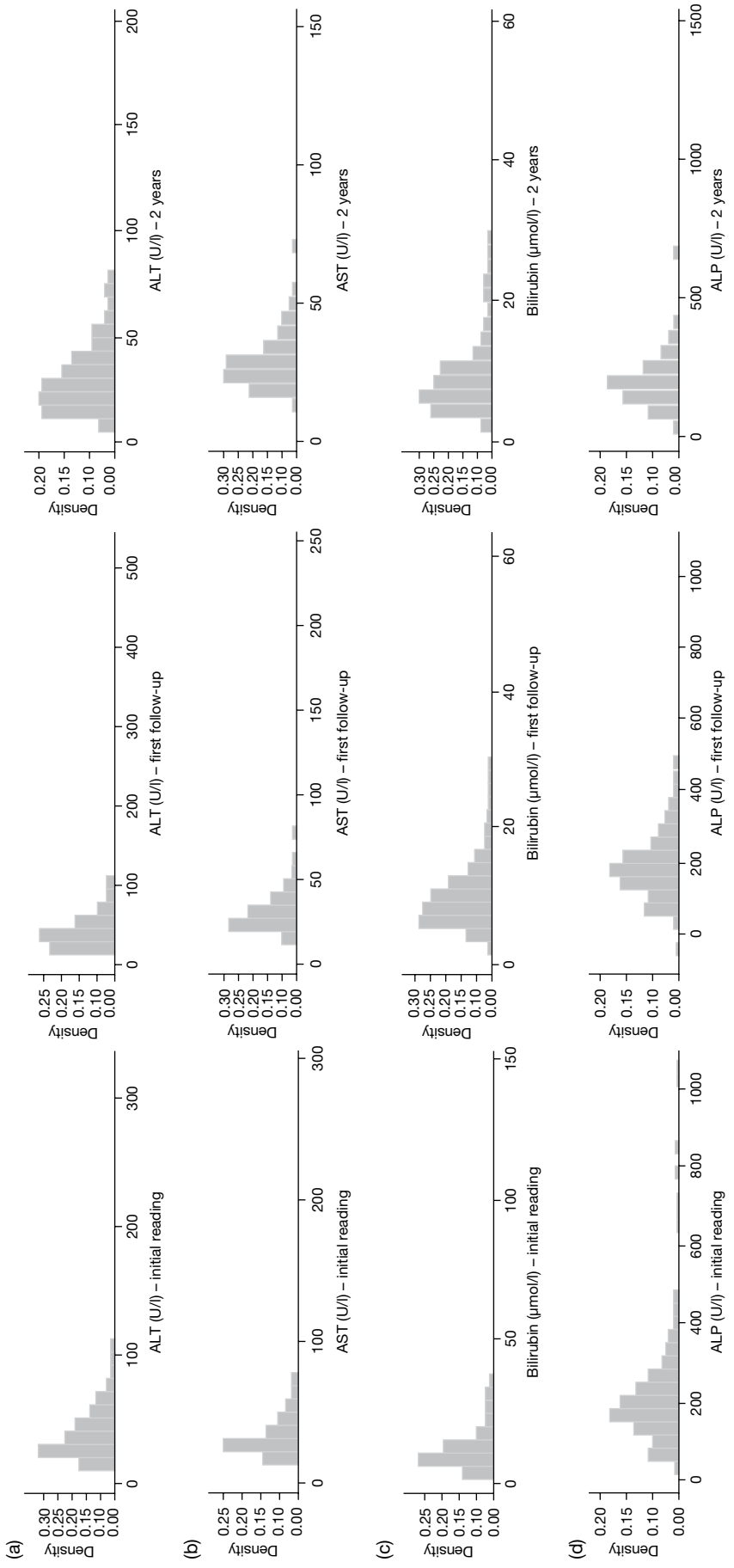
Laboratory effects

Apparent differences between results from different laboratories may be partly explained by differences in the patient population between different practices. Moreover, any genuine laboratory effects would be confounded by variation in testing policies between individual GPs and practices, leading to patient selection effects. However, these difficulties of interpretation do not preclude an informal analysis of results by laboratory.

A graphical check on distributional stability across laboratories is provided by the Q–Q plots in *Figures 26–28*. In these plots the straight lines correspond to distributional equivalence between laboratory 1 and laboratory 2 (top) and between laboratory 1 and laboratory 3 (bottom). Systematic departures are evident for ALP, and cannot be ruled out in some other cases (see bilirubin and GGT).

Quantile–quantile plots of LFT results generated by laboratory 2 against laboratory 1 and by laboratory 3 against laboratory are shown in *Figures 26–28*. LFT results have been pooled over baseline and both follow-ups (FU1 and FU2). The straight line on each plot represents distributional equivalence.

Similar Q–Q plots for log-transformed LFTs are shown in *Figures 29–31*. Once again, the straight lines represent distributional equivalence between laboratories. In these plots, a vertical shift in the reference line represents a multiplicative factor between results from different laboratories. Thus, it appears from the left-hand panels that a multiplicative adjustment of ALP would eliminate differences between laboratory 1 and laboratory 2, and *Figure 30* substantially reduce differences between laboratory 1 and laboratory 3. As noted above, such adjustments are readily


(a)
Density
ALT (U/l) – initial reading
ALT (U/l) – first follow-up
ALT (U/l) – 2 years
(b)
AST (U/l) – initial reading
AST (U/l) – first follow-up
AST (U/l) – 2 years
(c)
Bilirubin (μmol/l) – initial reading
Bilirubin (μmol/l) – first follow-up
Bilirubin (μmol/l) – 2 years
(d)
ALP (U/l) – initial reading
ALP (U/l) – first follow-up
ALP (U/l) – 2 years

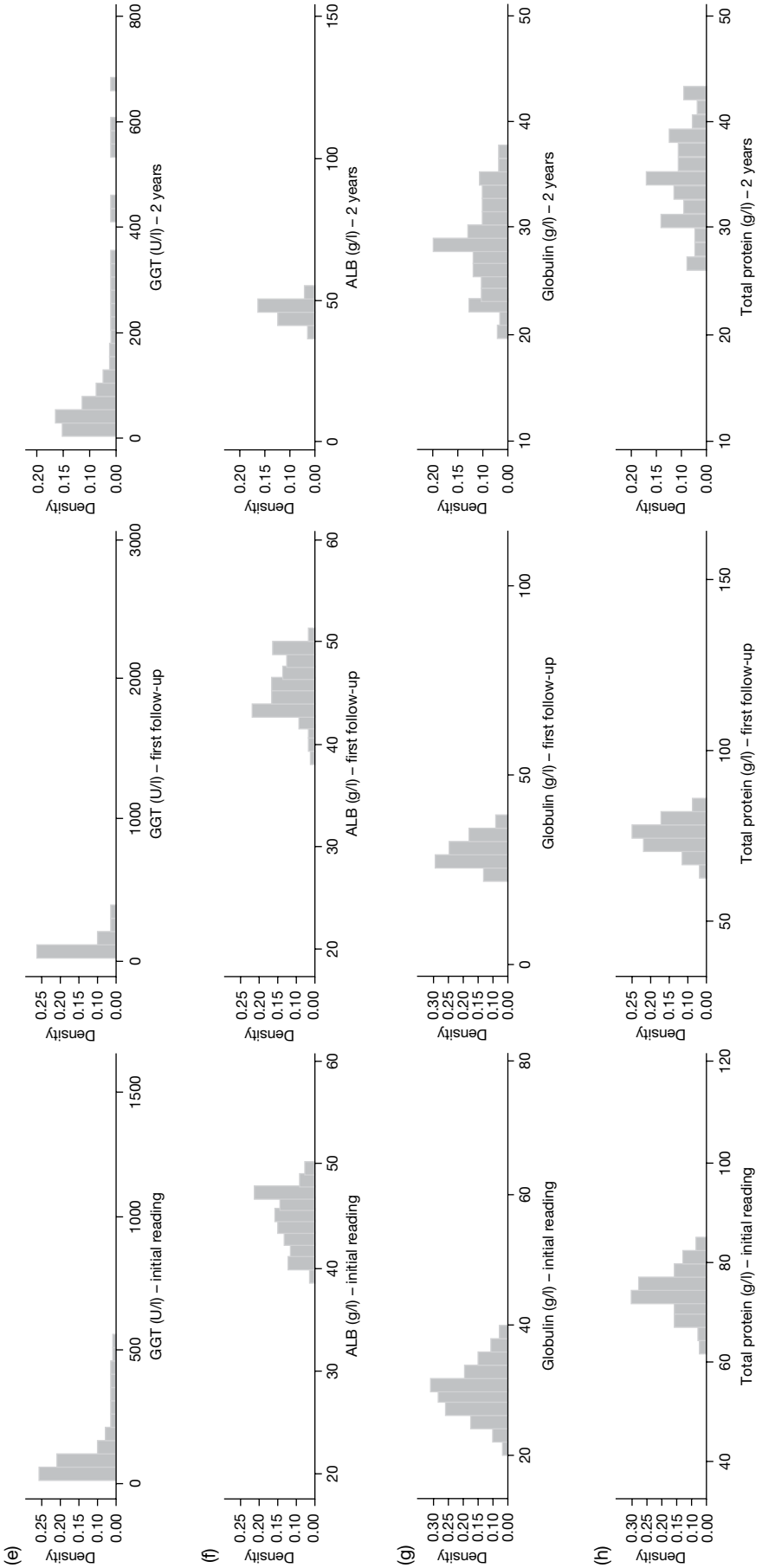
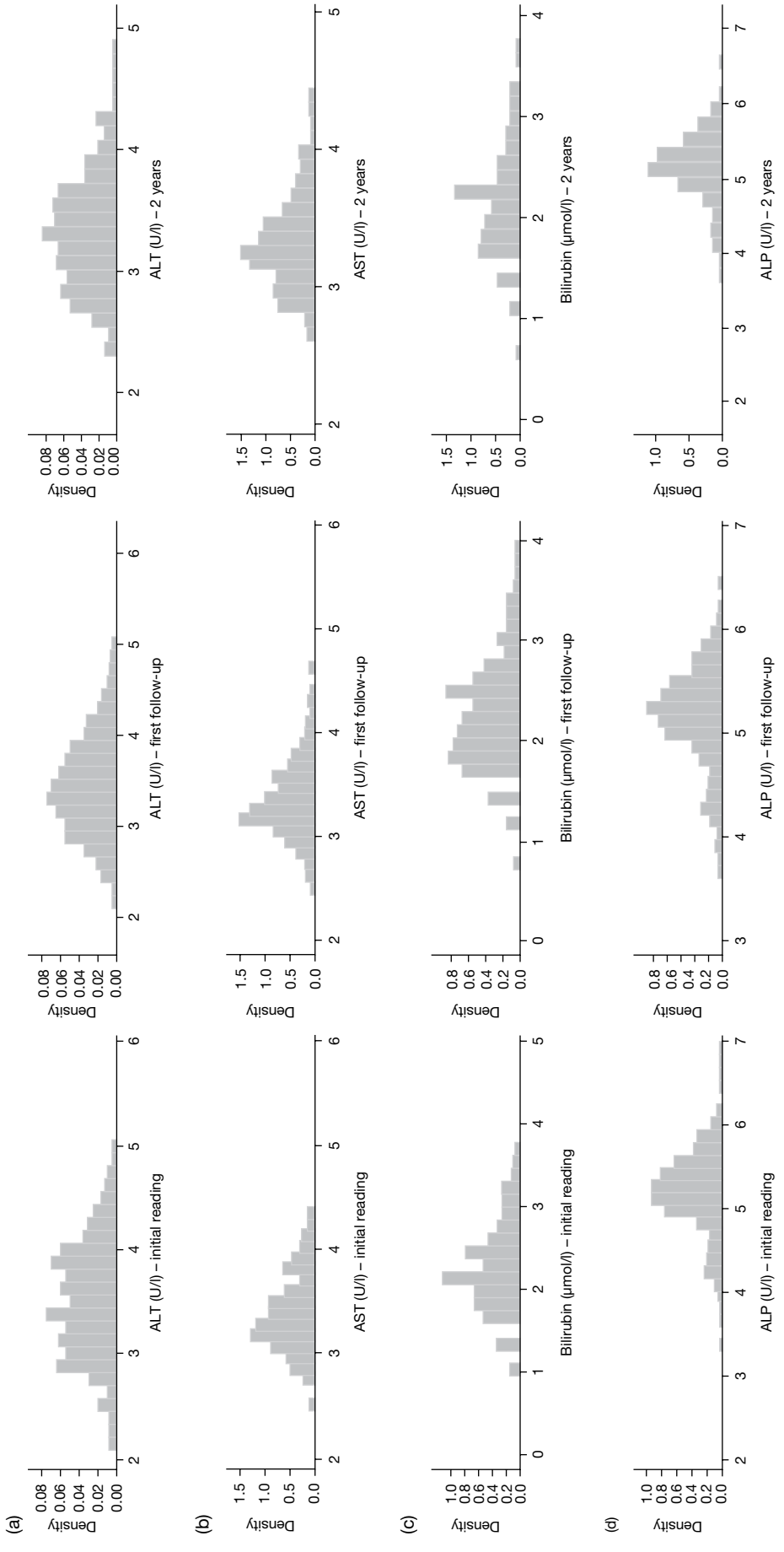


FIGURE 24 Histograms of LFT results (raw data) at baseline, FU1 and FU2.


(a)
(b)
(c)
(d)
Density
ALT (U/l) – initial reading
AST (U/l) – initial reading
Bilirubin (μmol/l) – initial reading
ALP (U/l) – initial reading
ALT (U/l) – first follow-up
AST (U/l) – first follow-up
Bilirubin (μmol/l) – first follow-up
ALP (U/l) – first follow-up
ALT (U/l) – 2 years
AST (U/l) – 2 years
Bilirubin (μmol/l) – 2 years
ALP (U/l) – 2 years

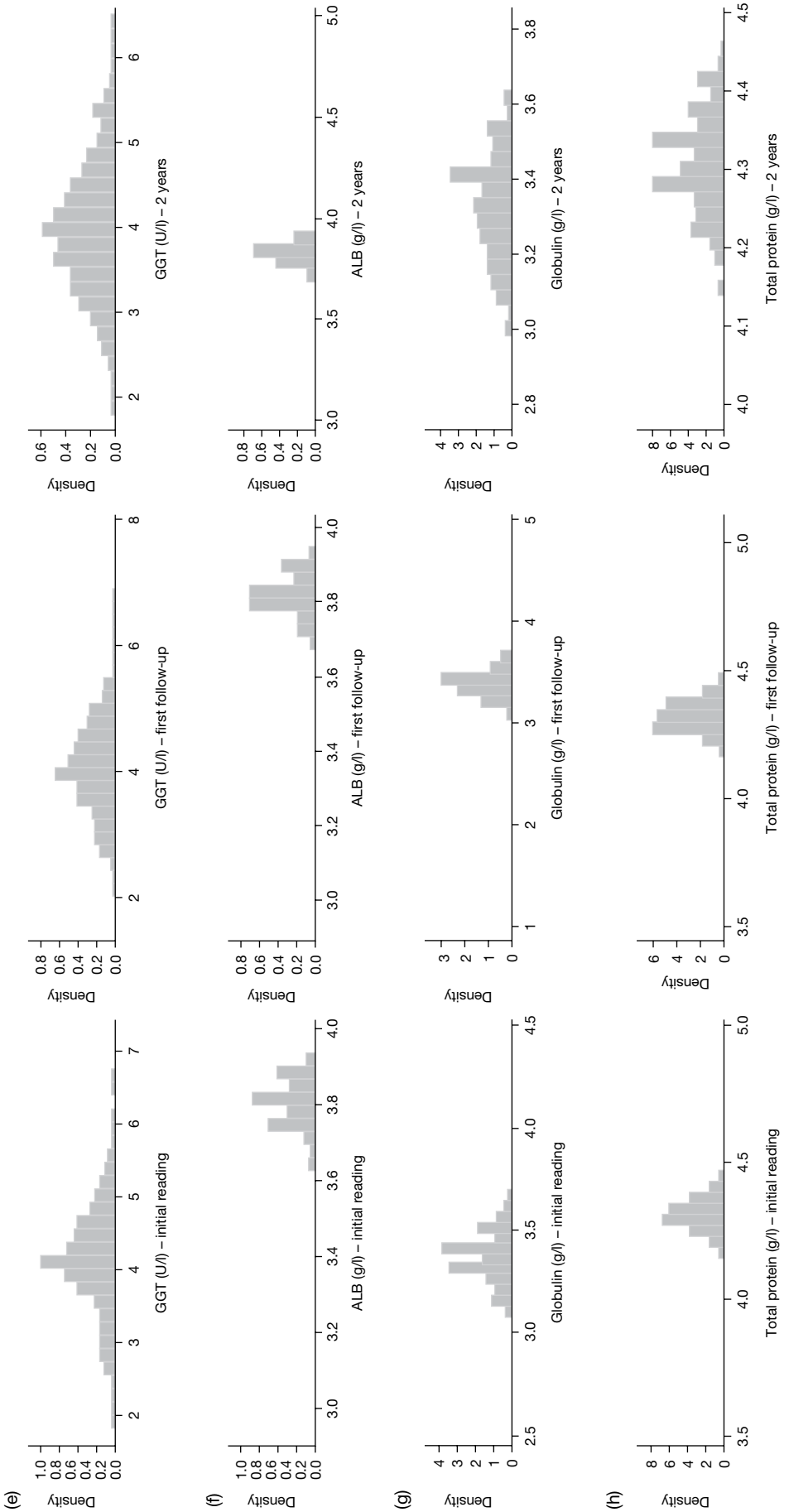


FIGURE 25 Histograms of LFT results (logged data) at baseline, FU1 and FU2.

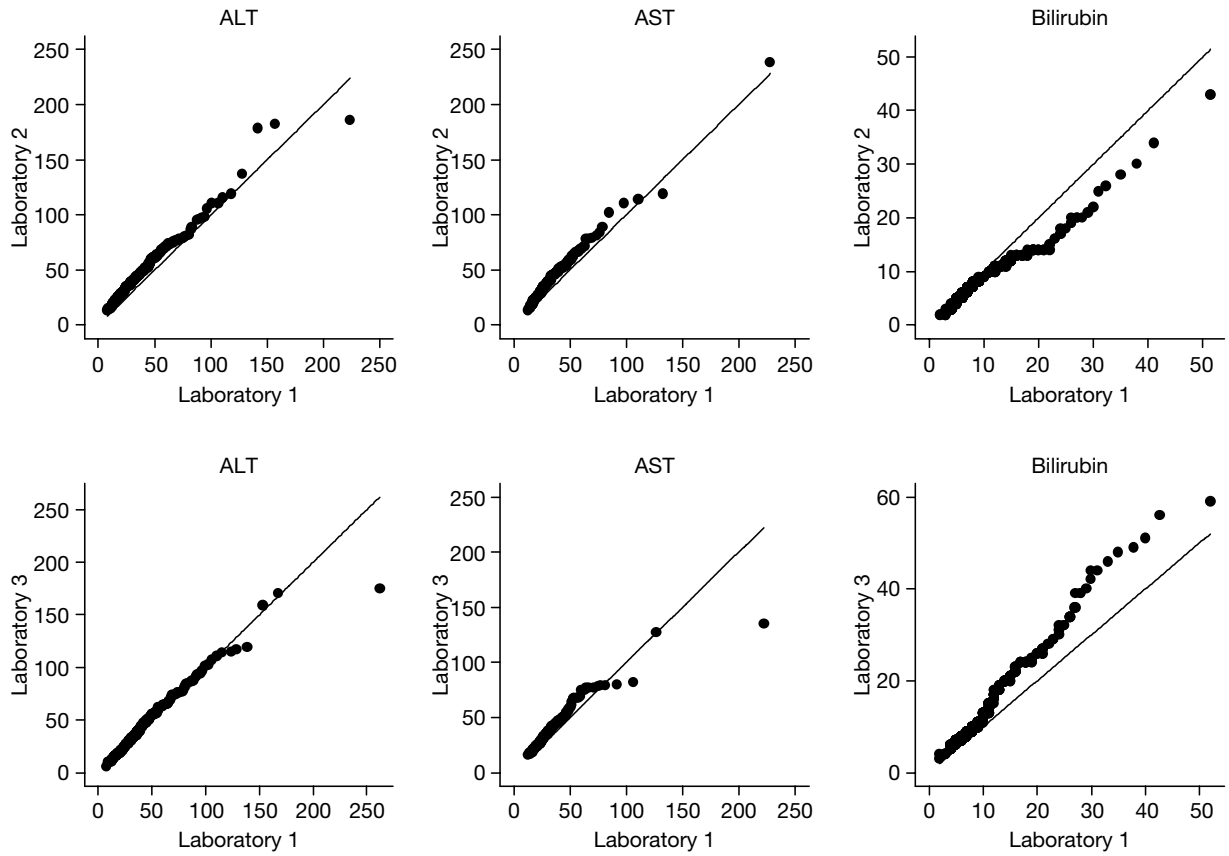


FIGURE 26 Quantile–quantile plots by laboratory for ALT, AST and bilirubin.

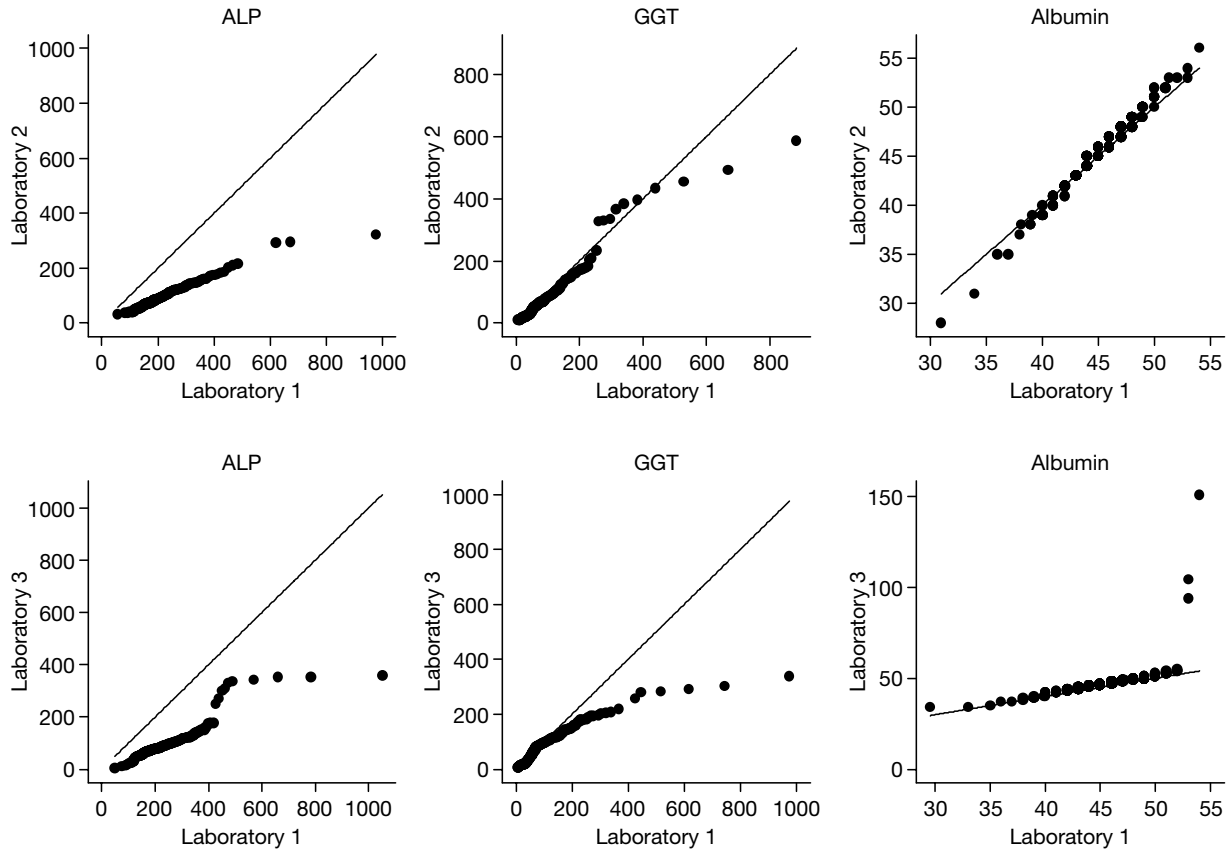


FIGURE 27 Quantile–quantile plots by laboratory for ALP, GGT and albumin.

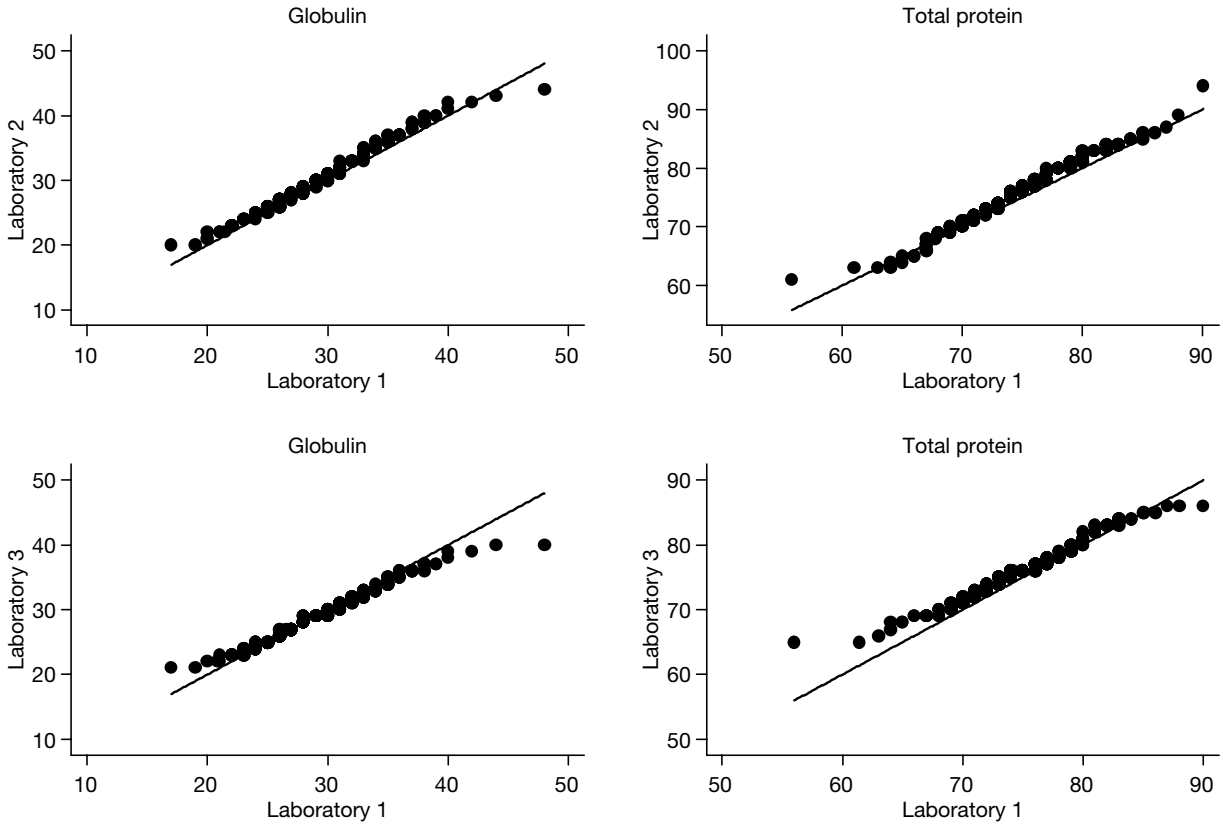


FIGURE 28 Quantile–quantile plots by laboratory for globulin and total protein.

available within additive (or generalised linear) models fitted to the log-transformed LFTs. This approach was applied to all eight analytes: it does no harm when the laboratory effect is absent, and has the capacity to adjust (at least partially) for differences between laboratories, whether or not these have been anticipated.

Figures 29–31 show Q–Q plots of log-transformed LFT results for laboratory 2 against laboratory 1 (top) and for laboratory 3 against laboratory 1 (bottom). LFT results have been pooled over baseline and both follow-ups (FU1 and FU2). The straight line on each plot represents distributional equivalence. Vertical (or horizontal) shifts to the line represent a multiplicative laboratory effect.

Summary of analyses of liver function test results

Univariate analyses

Method

Results are presented at each epoch (index, FU1 and FU2). All characteristics (except sex) are described by more than two categories. For these covariates – age, ethnic group, BMI, alcohol – one-way ANOVA was conducted on the log-LFT data, with adjustment for laboratory effects. An *F*-test for equality of LFT values across categories was conducted and the results displayed together with the marginal means for each category, computed as for laboratory 1 and back-transformed to the natural scale. Results for sex are displayed as multiplicative effects derived from *t*-tests applied to the logged data, after adjustment for laboratory effects.

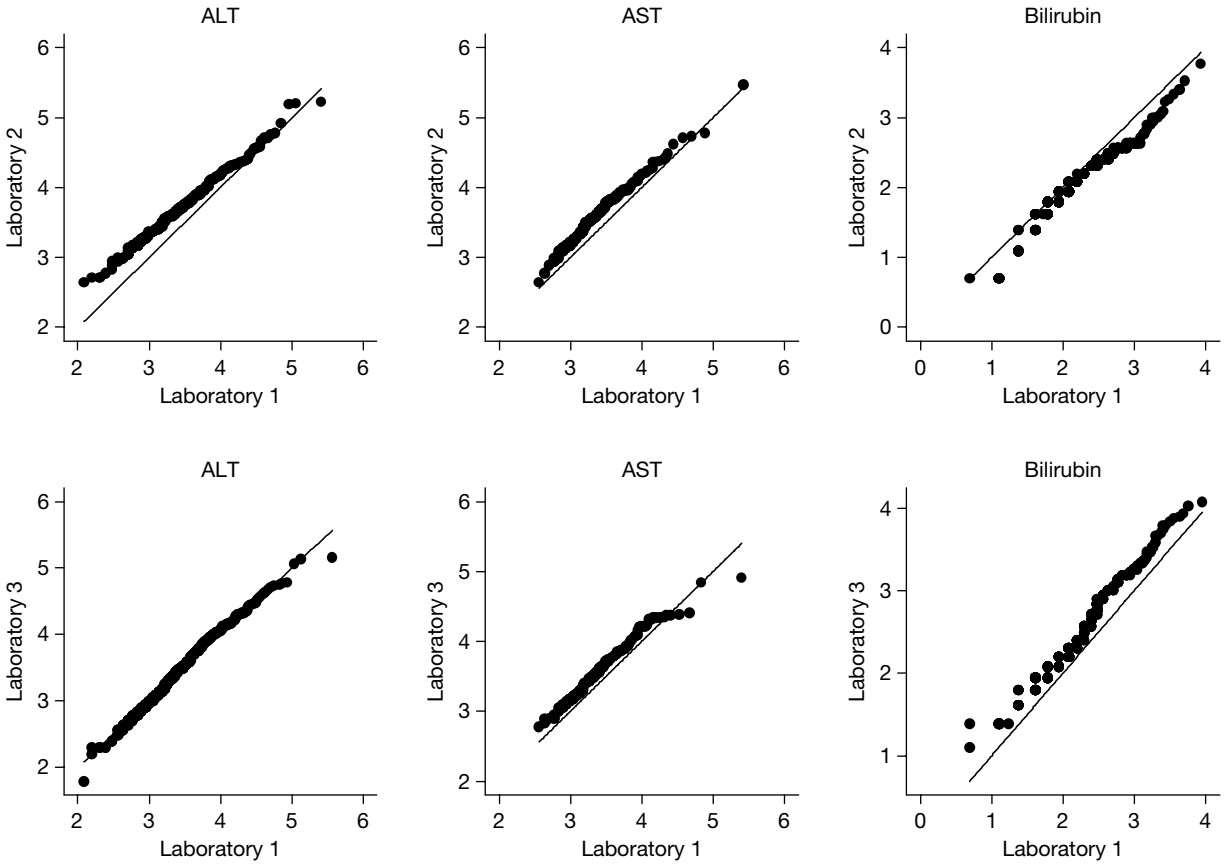


FIGURE 29 Quantile–quantile plots by laboratory for ALT, AST and bilirubin (logged data).

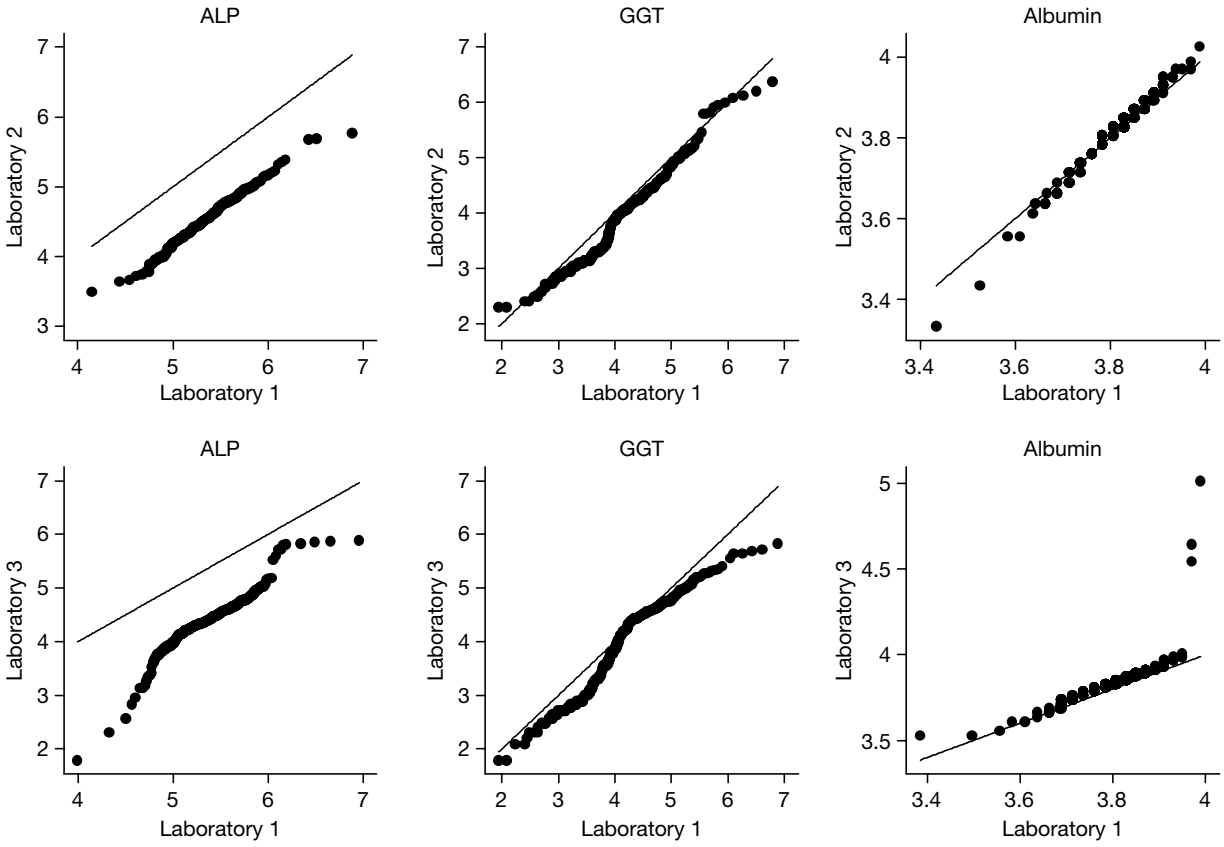


FIGURE 30 Quantile–quantile plots by laboratory for ALP, GGT and albumin (logged data).

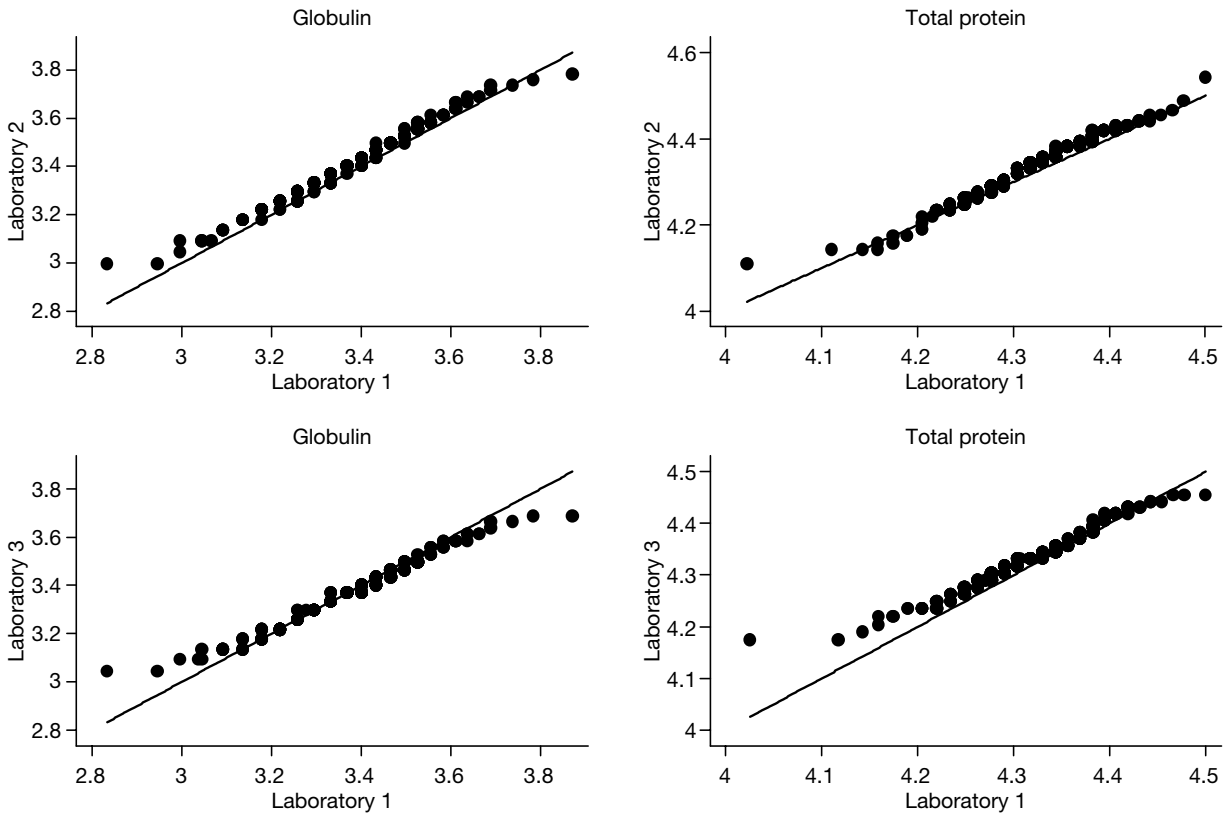


FIGURE 31 Quantile–quantile plots by laboratory for globulin and total protein (logged data).

Liver function test by sex

Tables 84–86 show average multiplicative sex effects, estimated from ANOVA models applied to log-transformed LFTs with adjustment for laboratory effects.

TABLE 84 Index LFTs

	Female/male			
Analyte	**Ratio**	**95% CI**	**T-statistic**	***p*-value**
ALT	0.78	0.73 to 0.84	–7.15	0.000
AST	0.92	0.88 to 0.97	–3.05	0.002
Bilirubin	0.73	0.69 to 0.77	–10.43	0.000
ALP	1.18	1.14 to 1.23	7.90	0.000
GGT	0.84	0.77 to 0.92	–3.87	0.000
Albumin	0.98	0.97 to 0.99	–5.13	0.000
Globulin	1.03	1.01 to 1.05	2.95	0.003
Total protein	1.00	0.99 to 1.01	–0.14	0.889

TABLE 85 First follow-up (FU1)

	Female/male			
Analyte	**Ratio**	**95% CI**	**T-statistic**	***p*-value**
ALT	0.77	0.72 to 0.82	–8.42	0.000
AST	0.91	0.87 to 0.96	–3.85	0.000
Bilirubin	0.75	0.71 to 0.80	–9.26	0.000
ALP	1.18	1.13 to 1.22	8.20	0.000
GGT	0.87	0.80 to 0.95	–3.13	0.002
Albumin	0.98	0.97 to 0.99	–5.39	0.000
Globulin	1.02	1.00 to 1.04	2.47	0.014
Total protein	0.99	0.99 to 1.00	–1.36	0.175

TABLE 86 Two-year follow-up (FU2)

	Female/male			
Analyte	**Ratio**	**95% CI**	**T-statistic**	***p*-value**
ALT	0.79	0.74 to 0.85	–6.18	0.000
AST	0.93	0.89 to 0.99	–2.52	0.012
Bilirubin	0.78	0.72 to 0.84	–6.38	0.000
ALP	1.13	1.08 to 1.20	4.70	0.000
GGT	0.84	0.75 to 0.95	–2.86	0.004
Albumin	0.99	0.98 to 1.01	–1.03	0.304
Globulin	1.01	0.99 to 1.03	0.77	0.441
Total protein	1.00	0.99 to 1.01	–0.58	0.562

Liver function test by age

Tables 87–89 show marginal estimates of average LFTs within age groups, as for laboratory 1, estimated from ANOVA models applied to log-transformed LFTs with adjustment for laboratory effects. The *F*-statistics are taken from the ANOVAs, and indicate the strength of the relationship with age.

TABLE 87 Index LFTs

	Age (years)						*F*-test for age effect		
Analyte	**≤ 34**	**35–44**	**45–54**	**55–64**	**65–74**	**75+**	***F***	**df**	***p*-value**
ALT	37	42	41	35	30	25	20.0	(5, 1106)	0.000
AST	33	31	33	31	30	28	4.5	(5, 1150)	0.000
Bilirubin	11	8	9	9	10	10	5.5	(5, 1257)	0.000
ALP	208	201	200	208	210	245	7.2	(5, 1264)	0.000
GGT	44	67	81	68	70	65	8.7	(5, 1144)	0.000
Albumin	46	45	45	45	44	43	21.4	(5, 1270)	0.000
Globulin	29	29	29	29	29	30	0.9	(5, 969)	0.457
Total protein	75	75	75	74	73	73	6.4	(5, 973)	0.000

TABLE 88 First follow-up (FU1)

	Age (years)						*F*-test for age effect		
Analyte	**≤ 34**	**35–44**	**45–54**	**55–64**	**65–74**	**75+**	***F***	**df**	***p*-value**
ALT	31	38	39	35	29	23	27.8	(5, 1226)	0.000
AST	27	30	32	30	28	27	5.8	(5, 1204)	0.000
Bilirubin	10	8	8	9	10	10	4.6	(5, 1225)	0.000
ALP	195	192	196	207	207	241	9.5	(5, 1228)	0.000
GGT	42	56	75	64	63	57	9.2	(5, 1235)	0.000
Albumin	47	46	46	45	45	44	17.6	(5, 1246)	0.000
Globulin	29	30	30	30	30	30	1.2	(5, 1206)	0.329
Total protein	77	76	77	75	75	74	7.2	(5, 1227)	0.000

TABLE 89 2-year follow-up (FU2)

	Age (years)						*F*-test for age effect		
Analyte	**≤ 34**	**35–44**	**45–54**	**55–64**	**65–74**	**75+**	***F***	**df**	***p*-value**
ALT	29	30	34	29	26	20	12.2	(5, 699)	0.000
AST	25	26	29	28	27	25	2.6	(5, 716)	0.023
Bilirubin	9	7	8	8	9	9	3.5	(5, 753)	0.004
ALP	185	189	191	196	201	228	3.3	(5, 750)	0.006
GGT	44	51	67	60	56	50	2.7	(5, 708)	0.020
Albumin	46	46	46	46	46	45	3.0	(5, 770)	0.012
Globulin	28	28	27	28	28	28	0.3	(5, 689)	0.903
Total protein	75	75	74	74	74	72	2.0	(5, 699)	0.080

Liver function test by ethnic group

Tables 90–92 show marginal estimates of average LFTs within ethnic groups, as for laboratory 1, estimated from ANOVA models applied to log-transformed LFTs with adjustment for laboratory effects. The *F*-statistics are taken from the ANOVAs, and indicate the strength of the relationship with ethnic group.

TABLE 90 Analysis of index LFTs by ethnic group

Analyte	White	Asian	Black	Other	*F*	df	*p*-value
ALT	34	36	33	38	0.7	(3, 1074)	0.538
AST	31	31	32	30	0.1	(3, 1118)	0.938
Bilirubin	9	9	9	9	0.3	(3, 1220)	0.844
ALP	211	220	198	198	1.1	(3, 1227)	0.354
GGT	70	51	65	65	4.4	(3, 1111)	0.004
Albumin	45	44	44	45	1.9	(3, 1233)	0.135
Globulin	29	32	32	30	16.6	(3, 947)	0.000
Total protein	74	77	77	76	10.9	(3, 951)	0.000

TABLE 91 Analysis of LFT by ethnic group at first follow-up (FU1)

Analyte	White	Asian	Black	Other	*F*	df	*p*-value
ALT	32	33	28	31	1.0	(3, 1192)	0.387
AST	29	30	28	28	0.3	(3, 1169)	0.829
Bilirubin	9	9	9	8	0.3	(3, 1190)	0.834
ALP	207	215	199	198	0.8	(3, 1193)	0.500
GGT	63	48	61	56	2.8	(3, 1200)	0.041
Albumin	45	45	45	46	2.8	(3, 1210)	0.041
Globulin	30	33	34	31	22.6	(3, 1172)	0.000
Total protein	75	78	79	79	16.7	(3, 1193)	0.000

TABLE 92 Analysis of LFT by ethnic group at 2-year follow-up (FU2)

Analyte	White	Asian	Black	Other	*F*	df	*p*-value
ALT	28	31	26	28	0.9	(3, 683)	0.444
AST	27	28	24	25	1.2	(3, 699)	0.296
Bilirubin	8	7	7	7	2.0	(3, 735)	0.119
ALP	198	212	227	194	1.7	(3, 732)	0.158
GGT	58	49	44	58	1.3	(3, 691)	0.264
Albumin	46	46	45	45	0.4	(3, 752)	0.757
Globulin	28	30	31	27	7.5	(3, 675)	0.000
Total protein	74	77	76	73	5.6	(3, 685)	0.001

Liver function test by body mass index

Tables 93–95 show marginal estimates of average LFTs within BMI categories, as for laboratory 1, estimated from ANOVA models applied to log-transformed LFTs with adjustment for laboratory effects. The *F*-statistics are taken from the ANOVAs, and indicate the strength of the relationship with BMI.

TABLE 93 Analysis of index LFTs by BMI at first follow-up (FU1)

	BMI (kg/m^2)						
Analyte	**< 20**	**20–24.99**	**25–29.99**	**30+**	***F***	**df**	***p*-value**
ALT	24	29	35	37	15.3	(3, 1074)	0.000
AST	33	29	31	31	1.7	(3, 1120)	0.164
Bilirubin	12	10	10	9	9.1	(3, 1222)	0.000
ALP	220	219	214	205	2.2	(3, 1228)	0.085
GGT	40	62	70	72	9.2	(3, 1112)	0.000
Albumin	44	44	45	44	1.6	(3, 1233)	0.193
Globulin	30	29	29	30	5.2	(3, 943)	0.001
Total protein	74	74	74	75	2.4	(3, 947)	0.068

TABLE 94 Analysis of LFT by BMI at first follow-up (FU1)

	BMI (kg/m^2)						
Analyte	**< 20**	**20–24.99**	**25–29.99**	**30+**	***F***	**df**	***p*-value**
ALT	22	26	33	35	21.4	(3, 1190)	0.000
AST	28	28	30	30	2.1	(3, 1169)	0.100
Bilirubin	11	10	10	8	6.6	(3, 1188)	0.000
ALP	215	215	209	202	1.8	(3, 1191)	0.148
GGT	42	54	63	66	7.2	(3, 1199)	0.000
Albumin	45	45	46	45	3.0	(3, 1209)	0.030
Globulin	30	29	29	31	7.1	(3, 1169)	0.000
Total protein	76	75	76	76	3.0	(3, 1190)	0.028

TABLE 95 Analysis of LFT by BMI at 2-year follow-up (FU2)

	BMI (kg/m^2)						
Analyte	**< 20**	**20–24.99**	**25–29.99**	**30+**	***F***	**df**	***p*-value**
ALT	18	23	28	31	15.3	(3, 654)	0.000
AST	24	26	27	28	1.4	(3, 648)	0.247
Bilirubin	14	9	9	8	8.5	(3, 669)	0.000
ALP	216	211	193	199	1.9	(3, 668)	0.128
GGT	33	54	56	61	2.8	(3, 662)	0.041
Albumin	46	47	46	46	2.5	(3, 688)	0.062
Globulin	28	27	27	28	2.1	(3, 658)	0.101
Total protein	74	74	74	74	0.4	(3, 666)	0.730

Liver function by self-reported alcohol intake

Tables 96–98 show marginal estimates of average LFTs within categories of alcohol intake, as for laboratory 1, estimated from ANOVA models applied to log-transformed LFTs with adjustment for laboratory effects. The *F*-statistics are taken from the ANOVAs, and indicate the strength of the relationship with alcohol.

TABLE 96 Analysis of index LFT by alcohol at first follow-up (FU1)

	Units per week								
Analyte	**0**	**1–14**	**15–29**	**30–49**	**50–99**	**100+**	***F***	**df**	***p*-value**
ALT	32	33	35	39	41	49	5.7	(5, 1099)	0.000
AST	29	30	31	36	37	42	10.3	(5, 1144)	0.000
Bilirubin	8	10	10	10	10	10	6.2	(5, 1249)	0.000
ALP	230	208	195	191	198	179	9.7	(5, 1256)	0.000
GGT	60	64	73	87	109	102	14.0	(5, 1137)	0.000
Albumin	44	44	45	45	45	46	2.9	(5, 1262)	0.013
Globulin	30	29	29	29	29	30	1.7	(5, 962)	0.135
Total protein	74	74	74	74	74	76	0.9	(5, 966)	0.456

TABLE 97 Analysis of LFT by alcohol at first follow-up (FU1)

	Units per week								
Analyte	**0**	**1–14**	**15–29**	**30–49**	**50–99**	**100+**	***F***	**df**	***p*-value**
ALT	29	31	34	39	40	45	11.6	(5, 1219)	0.000
AST	28	28	29	34	36	38	10.7	(5, 1198)	0.000
Bilirubin	8	10	10	10	9	9	4.7	(5, 1218)	0.000
ALP	222	205	198	188	192	176	7.9	(5, 1221)	0.000
GGT	54	57	67	80	113	88	18.9	(5, 1227)	0.000
Albumin	45	46	46	46	46	46	3.5	(5, 1238)	0.004
Globulin	31	30	29	30	30	30	3.4	(5, 1199)	0.005
Total protein	76	75	75	75	77	76	0.9	(5, 1220)	0.492

TABLE 98 Analysis of LFT by alcohol at 2-year follow-up (FU2)

	Units per week								
Analyte	**0**	**1–14**	**15–29**	**30–49**	**50–99**	**100+**	***F***	**df**	***p*-value**
ALT	26	27	32	33	34	23	4.8	(5, 666)	0.000
AST	26	26	29	31	35	25	7.6	(5, 661)	0.000
Bilirubin	8	9	9	10	9	9	2.4	(5, 682)	0.033
ALP	211	192	193	179	195	177	3.0	(5, 681)	0.011
GGT	52	51	67	95	86	42	9.3	(5, 675)	0.000
Albumin	46	46	46	46	45	47	0.3	(5, 701)	0.893
Globulin	28	27	28	28	28	30	2.7	(5, 669)	0.020
Total protein	74	73	75	74	74	78	2.8	(5, 677)	0.018

Multivariate analyses

Method

The approach is described in the main report (see *Chapter 4*, *Multivariate analysis*).

Results

The tables in this section give an account of the parameter estimates obtained when fitting the final covariate models from *Table 34* in the main report. These are expressed as multiplicative factors that represent the effect of the patient characteristic on the 'average' LFT level. The average level referred to here is a technically a geometric mean since the factors have been derived by back-transformation from an analysis of log-transformed data. Reference categories for the categorical variables have been chosen as follows:

- age: ≤ 34 years
- sex: male
- ethnic group: white
- BMI: from 20 to 24.999
- alcohol: 0 units per week.

The estimates and CIs in the tables are computed relative to these categories. Estimates for reference categories are necessarily equal to 1, and the confidence limits are left blank. Where interaction terms are present, some choices have to be made in representing the effects. These choices have been made by reference to a hierarchy of covariates in which age and sex have precedence. For instance, the joint effect of age and BMI is presented in terms of the effect of BMI at each fixed age, rather than the effect of age at each fixed level of BMI.

For completeness, laboratory effects have been included in the tables using laboratory 1 as reference.

The 'random component' is represented as a 95% prediction interval for the (multiplicative) residual error term in the model. It gives some perspective on the practical importance of the error limits expressed by the CIs. In this regard, the relative tightness of the errors for the protein measures compared with the other analytes is a consequence of differences in natural scaling of these analytes.

Temporal modelling of liver function tests

Introduction

Between-patient differences in measured LFTs arise from a combination of two sources of variation:

1. measurement error associated with laboratory processes and (which is indistinguishable from) short-term fluctuations in the concentration of the analyte in the subject's blood (this component may be attributed to 'random error', as such effects are uncorrelated over relatively short time horizons, i.e. hours or days)
2. variation that reflects genuine differences between patients.

The second source of variation (see '2', above) may be further subdivided into:

- long-term 'persistent' differences in average concentrations between individual patients
- ephemeral or 'transient' variation attributable to medium-term (perhaps seasonal) fluctuations in the patient's environment or behaviour.

Persistent effects will reflect demographic, genetic and environmental factors and the patient's state of health and might be associated with measured (and unmeasured) patient-level covariates. Transient or ephemeral effects can arise as a result of subtle variation in patterns of behaviour from week to week or month to month (e.g. eating, drinking and exercise) or short-term fluctuations in a subject's state of health. In such cases the effect might persist for some time but is not indicative of a long-term shift in average level.

For clinical diagnosis and monitoring the 'persistent' part of the patient component, specifically that portion of the persistent component that cannot be attributed to background genetic and demographic factors, has special significance as it is here that any 'signal' associated with a serious disease must reside. With this in mind, the aim of the present analysis is to quantify the relative magnitude of the different components for each analyte, in particular to identify the proportion of the variation attributable to long-term differences between patients that cannot be explained by background factors.

Statistical model

[For reasons discussed elsewhere (see *Chapter 4*, *Summary of liver function test data*, and *Distribution of liver function test data*, above) a log transformation was applied to each analyte before analysis. So, in what follows, the LFT value, X, is to be understood as the logarithm of a measured concentration.]

The variance decomposition described above is captured in the following model for the LFT value X, as measured for patient p at time t:

$$X_t^{(p)} = \mu + Y^{(p)} + Z_t^{(p)} + \varepsilon_t^{(p)},$$

where:

- μ is the population average for this LFT
- $Y^{(p)}$ is the long-run average deviation of patient p from the population average μ
- $Z_t^{(p)}$ is an auto-correlated time-series, with mean 0, which represents transient deviations of patient p from his or her long-run average
- $\varepsilon_t^{(p)}$ is an uncorrelated process with mean 0 and standard deviation σ_ε representing measurement error.

The components of the decomposition for different patients are statistically independent of one another, and all quantities are assumed normally distributed.

The transient process Z is modelled as an autoregressive process with parameter λ (≥ 0) and standard deviation σ_Z. Thus, the correlation function between times s and t is given by

$$\mathrm{corr}\left(Z_s^{(p)}, Z_t^{(p)}\right) = \lambda^{|t-s|}$$

The persistent component Y is modelled either as:

- A: a simple random effect with mean 0 and standard deviation σ_Y, or as
- B: the aggregate of a random effect and a fixed effect modelled as a linear function of patient-level covariates.

The more general model – model B – can be written:

$$Y^{(p)} = w^{(p)}\beta + \eta^{(p)}$$

where $w^{(p)}$ is a column vector of covariate values specific to patient p, β is a vector of parameter values and $\eta^{(p)}$ is the residual term with mean *0* and SD σ_η, and represents variation in the patient level that cannot be explained in terms of the available covariates. Where models A and B are both fitted to the same set of data, it is to be anticipated that the estimate of σ_Y (from model A) will exceed the estimate of σ_η (model B), as part of the variation will have been 'explained' by the covariates. The proportion of explained variation will then be a focus of attention.

The theoretical development is given in terms of the more general model B. Under model B, for each patient $\left\{X_t^{(p)};t\right\}$ is a stationary Gaussian process with mean

$$\mathrm{E}\left(X_t^{(p)}\right)=\mu+w^{(p)}\beta \tag{1}$$

variance

$$\mathrm{var}\left(X_t^{(p)}\right)=\sigma_\eta^2+\sigma_Z^2+\sigma_\varepsilon^2 \tag{2}$$

and correlation function

$$\mathrm{corr}\left(X_s^{(p)},X_t^{(p)}\right)=a+b\lambda^{|t-s|},s\neq t \tag{3}$$

where:

$$a=\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_Z^2+\sigma_\varepsilon^2} \tag{4}$$

and

$$b=\frac{\sigma_Z^2}{\sigma_\eta^2+\sigma_Z^2+\sigma_\varepsilon^2} \tag{5}$$

are the proportions of the variance attributable, respectively, to the persistent and ephemeral components of interpatient variation.

[Random measurement error accounts for the remaining proportion of the variance $(=1-a-b)$.]

Model-fitting

Subjects were included in the current study only if the index panel of LFTs exhibited at least one abnormal result. Thus, the LFTs observed in the current study are subject to selection bias. To address this problem, conditional versions of the models were fitted taking the initial LFT value as a given.

Suppose that the initial measurement for patient p occurs at time 0. In model B it follows that the conditional behaviour of the subsequent LFTs, given the initial LFT measurement, is governed by a non-stationary Gaussian process characterised by:

a conditional mean function:

$$\mathrm{E}\left(X_t\,|\,X_0=x\right)=\left(1-a-b\lambda^{|t|}\right)\left(\mu+w'\beta\right)+\left(a+b\lambda^{|t|}\right)x \tag{6}$$

a conditional variance function:

$$\mathrm{var}\left(X_t\,|\,X_0=x\right)=\left(1-\rho_t^2\right)\tau^2 \tag{7}$$

and a conditional covariance function:

$$\operatorname{cov}\left(X_s, X_t \mid X_0 = x\right) = \left(\rho_{t-s} - \rho_s \rho_t\right)\tau^2, s \neq t \tag{8}$$

Here $\rho_t = a + b\lambda^{|t|}$ is the correlation function from the stationary model and $\tau^2 = \sigma_\eta^2 + \sigma_Z^2 + \sigma_\varepsilon^2$ is the unconditional variance of X_t. (For simplicity, '*(p)*' has been suppressed in these expressions.)

Using these expressions, a (conditional) normal likelihood can be written down for a series of LFTs taken in the same patient at times $t_1, t_2, ..., t_{n_p}$, where $n_p + 1$ is the number of measurements taken on patient p. The full conditional likelihood is obtained by multiplication over all patients.

An algorithm

In the present study, each patient contributed zero, one, two or three measurements on each analyte. Given the conditional nature of the model, it is fitted using data only from patients in whom the LFT was measured two or three times. (Cases in which only a single measurement was available did not contribute to the temporal analysis for the analyte under discussion.) Thus, $n_p = 1$ or 2 for all patients contributing to the likelihood.

The models were fitted using an iteratively re-weighted least squares algorithm, which may be described as follows:

1. Using current estimates for *a* and *b*, set $\rho_t = a + b\lambda^{|t|}$, and use this to compute the conditional variances and covariances in *Equations* 7 and *8* above.
2. Maximise the likelihood over the parameters in *Equation 6*, treating the variances and covariances as fixed. (This second step is equivalent to a non-linear weighted least squares fit and does not involve τ^2, which can be set to any non-zero value at this point.) Steps 1 and 2 are repeated until convergence is obtained.
3. The variance τ^2 is estimated from the final implementation of the non-linear least squares fit.

For the current study this algorithm was implemented using the Stata non-linear regression command (nl) to carry out step 2. For patients with three separate LFT measurements ($np = 2$), the data vector $\left(X_{t_1}^{(p)}, X_{t_2}^{(p)}\right)$ was replaced by the transformed vector $\left(U_1^{(p)}, U_2^{(p)}\right)$ defined by:

$$U_1^{(p)} = \left(1 - \rho_{t_1}^2\right)^{-\frac{1}{2}} X_{t_1}^{(p)} \tag{9}$$

$$U_1^{(p)} = \left(1 - \rho_{12\cdot0}^2\right)^{-\frac{1}{2}} \left[\rho_{12\cdot0}\left(1 - \rho_{t_1}^2\right)^{-\frac{1}{2}} X_{t_1}^{(p)} - \left(1 - \rho_{12\cdot0}^2\right)^{-\frac{1}{2}} \left(1 - \rho_{t_2}^2\right)^{-\frac{1}{2}} X_{t_2}^{(p)}\right] \tag{10}$$

Here

$$\rho_{12\cdot0} = \frac{\rho_{t_2 - t_1} - \rho_{t_1}\rho_{t_2}}{\sqrt{\left(1 - \rho_{t_1}^2\right)\left(1 - \rho_{t_1}^2\right)}} \tag{11}$$

is the conditional correlation between $X_{t_1}^{(p)}$ and $X_{t_2}^{(p)}$ given the value of $X^{(p)}$. For patients with just two measurements ($n_p = 1$), $X_{t_1}^{(p)}$ was replaced by:

$$U_1^{(p)} = \left(1 - \rho_{t_1}^2\right)^{-\frac{1}{2}} X_{t_1}^{(p)}, \tag{12}$$

and the second component $U_2^{(p)}$ does not arise.

When t is the true correlation function, the components $\left(U_1^{(p)}, U_2^{(p)}\right)$ are uncorrelated with one another. Also $\text{var}\left(U_i^{(p)}\right) = \tau^2$ for $i = 1, 2$.

Step 2 of the iterative algorithm is executed by means of a non-linear regression applied to a derived data set consisting of the values of the transformed variables $U_i^{(p)}$, including contributions from all patients. The mean function for the regression is obtained by inserting the conditional mean expression from *Equation 6* into *Equations 9* and *10* in place of $X_t^{(p)}$.

For technical reasons the quantities a and b were parameterised as

$$a = \frac{e^{a'}}{1 + e^{a'} + e^{b'}} \tag{13}$$

and

$$b = \frac{e^{b'}}{1 + e^{a'} + e^{b'}}, \tag{14}$$

where a' and b' occupy an unrestricted domain $\left(-\infty < a', b' < \infty\right)$.

The autoregressive parameter λ was parameterised in terms of the logarithm of the 'half-life' h, where h is defined as time interval over which the autoregressive correlation λ^h reduces to half.

Model-fitting strategy

In principle, the algorithm described above is capable of fitting the full model B to each of the LFTs once a suitable set of explanatory covariates has been identified. In practice, the following strategy was adopted for each analyte:

1. Model B was fitted using the iterative algorithm using laboratory effects as the only explanatory covariates.
2. The contribution of patient-level covariates was explored by fitting model B again using the full set of explanatory variables identified elsewhere.

Results

Variance decomposition

The model partitions the variance of the LFT values into three components, as described above. In the full model, all three components are present. There are three natural submodels in which one or two of these components are missing (*Table 99*).

The deviances associated with each model are tabulated below (*Table 100*) for each analyte in turn. At this stage, adjustment was made for laboratory effects, but not for any patient-level covariates.

Where possible, a simpler model was preferred unless a more complex model was associated with a significant reduction in deviance. For albumin, this strategy was problematic. In this case, the 'best' model appears to be the 'no-patient' model. However, the estimated half-life of the decay in this model is very large (~65 months). In practice, this is hardly distinguishable from a model with no temporal decay. However, the 'full' model fits much better than the 'no-decay' model, so this has been preferred. The preferred models have been highlighted in the table. The 'full' model was preferred for all but three analytes: for ALP and GGT there was no evidence of any 'measurement error,' whereas the analysis of total protein suggested a simple decomposition into a patient component and an error term, with no evidence of decay in the temporal correlation.

The parameter estimates in the preferred models are presented in *Table 101*, below. These comprise the components of variance (in percentage terms) and the 'half-life' of the ephemeral component, i.e. the time (in months) at which correlation function of the ephemeral component falls to half.

The same information is presented graphically in the panels of *Figure 32*, which show the correlation function for each analyte over a 30-month time horizon. The variance components are displayed as percentages on the right-hand axis scales.

From these results, it appears that most of the variation in LFTs is associated with patient effects rather than random error. Indeed, for the models for ALP and GGT suggest that measurement error is entirely absent. This may seem implausible, but it is possible that the error component is particularly low for these analytes (at least as a proportion of the total variance), leading to a

TABLE 99 Candidate models

Model	No. of parameters	Correlation function	Components of variance		
			Between patients (persistent)	Between patients (ephemeral)	Measurement error
Full	3	$\rho_t = a + b\lambda^{\lvert t \rvert}$	a	b	$1-a-b$
'No error'	2	$\rho_t = a + (1-a)\lambda^{\lvert t \rvert}$	a	$1-a$	0
'No patient'	2	$\rho_t = b\lambda^{\lvert t \rvert}$	0	b	$1-b$
'No decay'	1	$\rho_t = a$	a	0	$1-a$

TABLE 100 Deviances from fitting candidate models to the LFT data

Model	df	Analyte							
		ALT	AST	Bilirubin	ALP	GGT	Albumin	Globulin	Total protein
Full	3	2778.152	1385.267	3598.937	967.5404	4887.153	–6290.488	–1562.914	–5560.725
'No error'	2	2789.109	1424.504	3610.029	967.5404	4887.170	–6252.988	–1553.160	–5560.555
'No patient'	2	2799.723	1399.177	3602.974	†	4928.983	–6289.263	–1557.625	†
'No decay'	1	2914.509	1523.109	3611.458	984.5354	5000.326	–6251.368	–1541.471	–5560.555

† Convergence not achieved.

TABLE 101 Parameter estimates

Analyte	Random error		Patient (persistent)		Patient (ephemeral)		Decay 'half-life'	
	%	95% CI	%	95% CI	%	95% CI	Months	95% CI
ALT	11.72	5.83 to 17.61	52.50	47.65 to 57.36	35.78	28.71 to 42.85	2.22	1.28 to 3.87
AST	23.39	17.96 to 28.82	38.33	32.89 to 43.77	38.28	31.57 to 45.00	3.48	1.73 to 7.02
Bilirubin	19.62	13.37 to 25.86	69.57	64.90 to 74.25	10.81	3.95 to 17.67	2.85	0.35 to 23.08
ALP			66.73	62.23 to 71.23	33.27	28.77 to 37.77	0.30	0.18 to 0.48
GGT			76.54	72.73 to 80.35	23.46	19.65 to 27.27	1.12	0.85 to 1.49
Albumin	32.63	28.24 to 37.02	46.72	25.60 to 67.83	20.65	1.84 to 39.46	10.42	0.74 to 146.95
Globulin	14.82	6.41 to 23.23	51.77	42.71 to 60.83	33.41	23.73 to 43.10	4.33	1.05 to 17.86
Total protein	37.08	33.03 to 41.13	62.92	58.87 to 66.97				

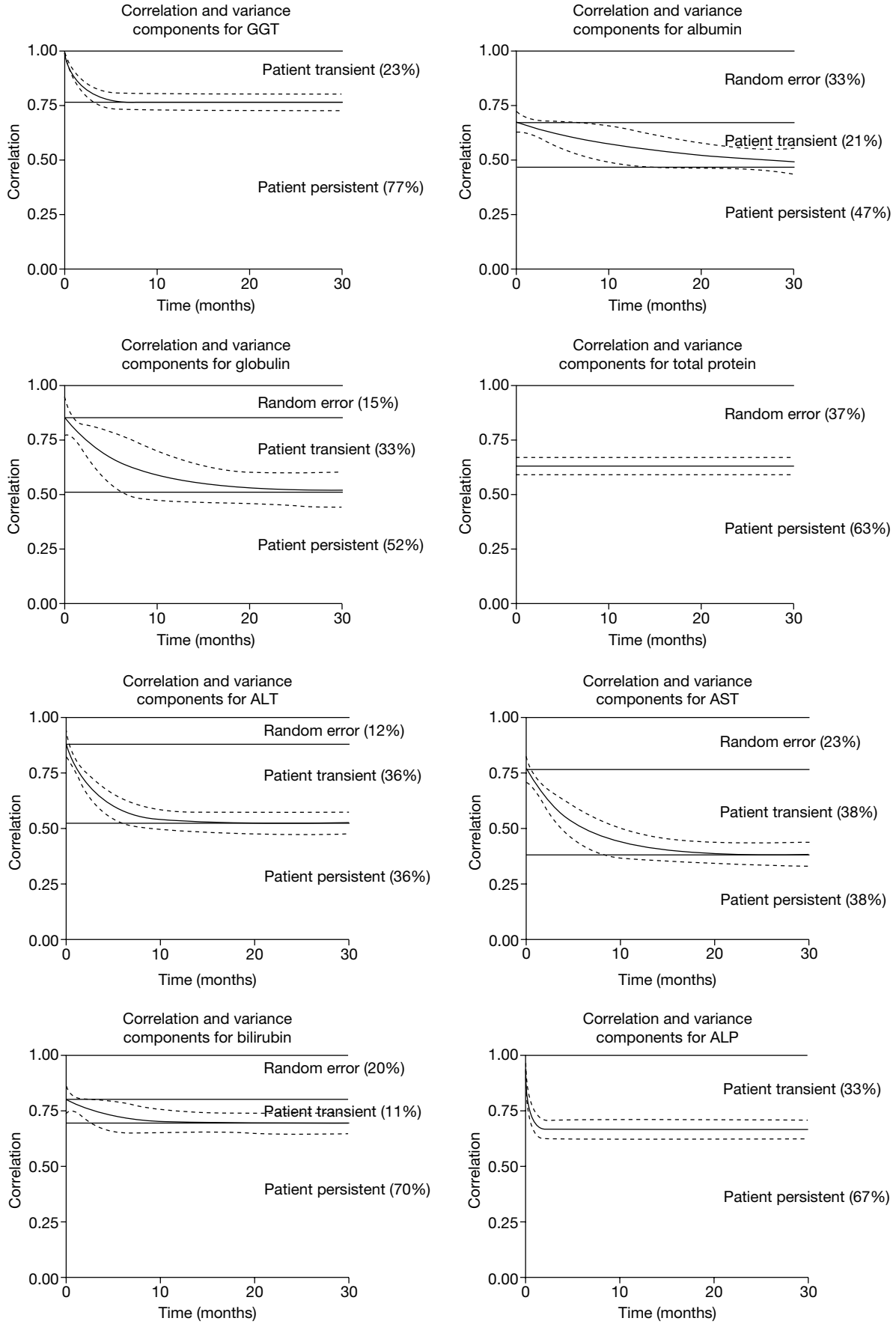


FIGURE 32 Estimated temporal correlation and components of variance for eight analytes from preferred models (without adjustment for patient-level covariates).

statistical inability to detect it in this sample. In any event, both these analytes exhibit significant transient patient effects, which may be difficult to distinguish from measurement error if the 'half-life' is small. Thus, for ALP, the transient component has an estimated half-life of only around 9 days and the correlation plot is close in character to that for an analyte (such as total protein) where the transient is replaced by random error in the best-fitting model.

The potential usefulness of an analyte for long-term patient monitoring is limited by the size of the persistent patient effect. In principle, the proportion of the variance attributable to this component places an upper limit on the signal–noise ratio for detecting serious disease. From this point of view, the most promising analytes are GGT (76%), and bilirubin (70%), with ALP not far behind (67%). Perhaps the most striking finding is the relatively unpromising result for AST (38%). Although the aggregate between-patient component is 76%, around half of this is accounted for by short-term fluctuations within patients, with a half-life of about 4.5 months. Of course, this analysis cannot distinguish which analytes are in fact most useful for any particular purpose, as all exhibit substantial persistent variance components.

A simpler approach to the analysis of between-patient variation is to compute the ICC for patients using the index LFT and both follow-up LFTs for each patient. This approach disregards the study selection effects and does not allow for the possibility of correlated temporal fluctuations (i.e. the transient component). It is, of course, more straightforward to apply. For comparative purposes, the results from this method are displayed in *Table 102*.

The first set of estimates are ICCs, the second ignores transient components and the third represents the best estimate available. The performance of the ICCs relative to the best estimates is subject to two conflicting tendencies: by ignoring transient effects the ICC will overestimate long-run persistence, whereas ignoring selection effects will overestimate the error component and tend to underestimate the persistent effect, as the index LFT is more likely to have been subject to an unusually large measurement error. In most cases, it appears that crude ICC is an unreliable guide to the true long-run correlation based on the current model.

The effect of patient-level covariates

Part of the persistent component of interpatient variation can be attributed to differences in measured patient-level covariates. To estimate this component, the covariates identified in *Chapter 4* (see *Multivariate analysis*) were incorporated into the fitting process described above (model B). For this purpose, weakly significant interaction terms (alcohol × sex for albumin, ethnicity × age for total protein) were omitted. Convergence problems were encountered in

TABLE 102 Comparison of estimates for between-patient variance components

	Intrapatient correlation	Crude conditional analysis (from 'no-decay' model)		'Persistent' component (from preferred model)	
Analyte	%	%	95% CI	%	95% CI
ALT	68.8	72.7	70.0 to 75.4	52.5	47.7 to 57.4
AST	61.2	62.3	59.4 to 65.2	38.3	32.9 to 43.8
Bilirubin	72.5	75.2	72.2 to 78.1	69.6	64.9 to 74.3
ALP	76.0	70.3	66.4 to 74.3	66.7	62.2 to 71.2
GGT	80.9	90.2	88.1 to 92.3	76.5	72.8 to 80.4
Albumin	55.8	61.8	58.5 to 65.0	46.7	25.6 to 67.8
Globulin	55.2	71.1	67.1 to 75.2	51.8	42.7 to 60.8
Total protein	56.6	62.9	58.9 to 67.0	62.9	58.9 to 67.0

two cases (globulin, total protein). The available results are displayed in *Table 103* alongside adjusted R^2 results from simple ANOVAs applied at each epoch. The analyses were repeated with restriction to the non-specific diagnostic group.

For ALT, it appears that around one-third of the persistent component of inter-patient variation can be explained by identifiable differences between individuals. For other analytes, the explanatory power of the covariates is considerably less important.

It is striking that the proportion of the variance explained in the simple ANOVAs tends to diminish as time goes by, especially at the 2-year follow-up. This is hard to explain but might be related in some way to the attenuation of selection bias with time.

A note on the impact of selection effects

The natural population for a study of the current type consists of patients undergoing LFTs in primary care. Yet the available data refer to a selected sample from this population. The selection mechanism has several aspects.

1. First, only patients with an 'abnormal' panel of index LFTs are invited to join the study. This feature is a clear source of bias if LFTs are analysed as if they derive from a random sample of the 'natural' population.
2. Second, not all invited patients express effective consent for the study by undergoing tests at FU1. Those who do may not constitute a random sample of those who are eligible.
3. Finally, not all study patients attend for tests at FU2. Those with identified liver disease were not invited to do so and may be systematically excluded from analyses of 2-year data. Of greater concern is the possibility of differential rates of attendance within different patient subgroups. For example, it is clear that 2-year follow-up is more complete among middle-aged patients than in younger and older age groups.

Some aspects of the dependence of LFT results on patient-level covariates can be analysed using a conditional approach without undue concern for selection effects. By conditioning on the index LFTs, the analysis is able to consider subsequent LFT values, as they occur within a selected sample. However, the full range of patient variation is not necessarily reflected proportionately within the sample and, for this reason, one must be cautious when assessing the magnitude of inter-patient variation as presented here. Similarly, the component of variance attributable to patient-level covariates may not be accurately assessed even although, in principle, the regression

TABLE 103 Proportion of total variance attributable to patient-level covariates

		All subjects				Non-specific diagnoses			
Analyte	Fitted covariates (all models include c)	Temporal model	Index ANOVA	FU1 ANOVA	FU2 ANOVA	Temporal model	Index ANOVA	FU1 ANOVA	FU2 ANOVA
ALT	BMI × age, alcohol	17.8	14.9	19.7	14.5	19.0	16.0	21.0	16.3
AST	BMI × age, alcohol	5.1	6.1	6.7	5.7	5.0	5.4	6.0	7.3
Bilirubin	BMI, alcohol	6.9	12.4	9.8	8.1	7.3	13.6	10.5	8.6
ALP	Alcohol	4.6	8.0	9.3	4.2	4.3	7.6	9.1	4.3
GGT	BMI × age, alcohol	7.8	13.3	13.8	6.7	6.5	13.9	14.1	6.4
Albumin	BMI × age, ethnicity	6.5	12.8	11.3	2.5	4.5	13.8	11.7	2.8
Globulin	BMI × age, ethnicity	N/A	7.6	8.7	1.4	NA	6.7	8.9	1.1
Total protein	BMI × age, ethnicity	N/A	4.5	7.6	1.5	NA	4.4	8.1	1.1

N/A, not applicable.
Results for globulin and total protein could not be obtained from the temporal model because of computational problems.

coefficients associated with such covariates can be estimated without bias under appropriate modelling assumptions. Clearly, statements about proportions of attributable variance must be interpreted as restricted to the sample of subjects entering the study.

The 2-year follow-up data are subject to patient dropout, some of which is non-random. In principle, non-random dropout that is related to measured covariates can be handled using inverse probability weighting when analysing FU2 data. However, the possibility of informative dropout – perhaps associated with unmeasured covariates – cannot be excluded.

Appendix 3

BALLETS study: summary of ethics and substantial amendment approval

The main research ethics committee, St Thomas' Hospital Research Ethics Committee, gave favourable ethical opinion to the BALLETS study on 19 April 2005. The following research sites were approved: Hall Green Health (Principal Investigator Dr Masood Nazir) on 2 June 2005, Bellevue Medical Centre (Principal Investigator Dr Sukhdev Singh) on 16 May 2005, and Guy's, King's and St Thomas' School of Medicine, Department of General Practice and Primary Care, and general practices in Lambeth (Principal Investigator Dr David Armstrong) on 18 July 2005.

All amendments were approved by South Birmingham and Lambeth local research ethics and Research and Development committees. Approved patient documentation for the study is contained in the appendices of the study protocol.

During the recruitment and follow-up phases of the study, the Modifications Subcommittee of St Thomas' Hospital Research Ethics Committee approved the following substantial amendments to the BALLETS study protocol and documentation.

Substantial amendment no. 1 (approval date 10 May 2006)

Approved protocol changes included:

- *Closure of site* The Principal Investigator (PI) and partners at Bellevue Medical Practice withdrew from the study in December 2005, prior to recruitment.
- *Section 3.3.1 Recruitment of new practices and PIs* Lordswood House Medical Practice (PI Dr Ewan Hamnett), Greenridge Surgery (PI Dr Richard McManus) and Yardley Wood Health Centre (PI Dr Peter Clarke).
- *Change of PI* Owing to changes in the routine workload of the PI at Hall Green Health practice, the role was transferred to GP colleague, Dr Bill Strange.
- *Section 3.3.2 Clinical process alterations* were made to reflect differences in routine clinical practice at each of the surgeries involved in both Birmingham and Lambeth.
- *Section 3.7.1 Psychological pilot study* to inform the development of psychological questionnaires for use in the main study. Although interviews with participants were approved in the original ethical approval, additional pilot interviews were conducted prior to administration of the first psychology questionnaire (T_1).

Patient documentation approval included:

- *Psychological pilot study* consent and information forms.

Substantial amendment no. 2 (approval date 10 July 2006)

Approved protocol changes included:

- Section 3.3.2 *Formal enrolment in subsequent testing protocol: defining of the patient population and seeking consent* Altered to document changes to the study process at Lambeth sites.
- *Section 3.5.1 Broad aim* Altered to address the possibility of selection bias, which could occur when suitable patients decline to take part or when suitable patients are not selected by their GP to take part.

Patient documentation amendments included:

- New *Patient Information Sheets*, produced to reflect the study process at each site.
- Lambeth team produced a new *Consent Form* and obtained approval for a *Pre-Consent* form.

Substantial amendment no. 3 (approval date 16 August 2006)

Patient documentation approved:

- *T1 psychology questionnaire* First draft submitted for approval.

Substantial amendment no. 4 (approval date 8 September 2006)

Approved protocol changes included:

- *Psychology sections* 2.5.6 (altered to describe the contents of psychology questionnaires); 3.3.4 (indicating that psychology questionnaires would not be translated into languages other than English); 3.3.6 (documenting that alcohol consumption would be measured at 2 years); and 3.7.1 (study process updated regarding the measures and time points used for data collection).

Patient documentation amendments included:

- *T1 psychology questionnaire* Minor amendments to Section 1 and Section 2 of the questionnaire.

Substantial amendment no. 5 (approval date 2 February 2007)

Patient documentation amendments included:

- Woodland Road Surgery patient documentation (Patient Information Sheet, Patient Letter and Receptionist Script).

Substantial amendment no. 6 (not approved)

An application was submitted to the main ethics committee for approval of a substudy for the collection, cryogenic storage and later testing of blood samples.

As the committee required a copy of a Human Tissue Authority Licence and clarification on the issue of anonymising samples, it was decided that the application should be withdrawn and discussed at the next steering committee.

Substantial amendment no. 7 (approval date 1 September 2007)

Approved protocol changes included:

- *Section 3.3.2* An outline of the procedures to be undertaken by patients at three new Birmingham practices, including Wand Medical Centre, Cofton Medical Centre and Shenley Green Health Centre.

Patient documentation amendments included:

- *Patient Information Sheets* for three new practices, modification of the date and version number of Patient Information Sheets for Birmingham practices that were already recruiting, and a new version of the *Consent Form*.

Substantial amendment no. 8 (not approved)

A second application for approval of the 'cryogenic storage of blood sample' add-on study [see *Substantial amendment no. 6 (not approved)*] was submitted on 26 October 2007. This application was withdrawn as further advice was required from the study steering committee regarding the proposed length of time for cryogenic storage, future plans for further study and destruction of samples, and the implications for patients of the proposed genetic testing.

Substantial amendment no. 9 (approval date 3 April 2008)

Approved protocol changes:

- *Section 5.1 Cryogenic storage and later testing* This was the third and final application requesting approval to collect and store an anonymised blood sample from consenting Birmingham patients, attending for their 2-year follow-up appointment.

Regarding future genetic research, St Thomas' Hospital Research Ethics Committee advised that in the event of discovering that a polymorphism is associated with a poor prognosis and requiring a treatment adjustment, then de-anonymising of stored cells will need to be considered.

Patient documentation amendments included:

- An *Extra Consent Form* to obtain consent from patients, who agreed to take part in the add-on study, and for one anonymised blood sample to be collected, stored and tested.
- An *Extra Patient Information Sheet* containing a brief description of the add-on study process at 2-year follow-up appointments at Birmingham practices.

Substantial amendment no. 10 (approval date 28 July 2008)

Approved protocol changes included:

- *Section 3.3.6 Long term follow-up* For changes to the 2-year follow-up phase, including repeat USS and interview.

Patient documentation amendments included:

- *Version 2.0 Extra Consent Form* to obtain consent from patients, who agreed to take part in the add-on study, and for one anonymised blood sample to be collected, stored and tested.
- *Version 2.0 Extra Patient Information Sheet* containing a brief description of the add-on study process at 2-year follow-up appointments at Birmingham practices.
- *Second Psychology Questionnaire T4* (referred to in the main report as T2). Modified version of the questionnaire.

Substantial amendment no. 11 (approval date 28 October 2008)

Approved protocol changes included:

- *Section 5.2* A qualitative investigation into LFT ordering behaviour of GPs involved in the BALLETS study. A substudy designed to examine the non-clinical motives behind a GPs decision to order an LFT.

Patient documentation amendments included:

- *General Practitioner Interview Consent Form* to obtain consent from GPs taking part in the substudy.
- *Information Sheet for Birmingham GPs* Sent to GPs at Birmingham practices interested in taking part in the substudy.

Substantial amendment no. 12 (approval date 5 February 2009)

Approved protocol changes included:

- *Section 5.3 Follow-up of abnormal test results* In the course of the study some patients tested positive for some specific liver diseases, but many were not followed up according to the agreed algorithm for referral or further testing. Letters were prepared by the study hepatologist and chief investigator suggesting appropriate follow-up of individual study patients testing positive for particular diseases.

Documentation included for approval:

- primary biliary cirrhosis letter to GPs
- Wilson's disease letter to GPs
- haemochromatosis letter to GPs
- autoimmune hepatitis letter to GPs.

Substantial amendment no. 13 (approval date 23 March 2010)

Approved protocol changes included:

- *Section 5.4* Qualitative investigation exploring anecdotal and preliminary evidence that events associated with participation in the BALLETS study were motivational to patients with and without fatty liver. Patients and sonographers who undertook first and follow-up USSs for the study were interviewed, as there were several anecdotal accounts from patients at follow-up clinics reporting implementation of lifestyle changes following attendance at first clinics.

Documentation included for approval:

- *Participant Consent Form* patient to obtain consent from patients taking part in the substudy.
- *Participant Consent Form* sonographer to obtain consent from sonographers taking part in the substudy.
- *Participant Information Sheet* patient.
- *Participant Information Sheet* sonographer.
- *Semistructured interview script* patient.
- *Semistructured interview script* sonographer.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library