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

Scientific summary

Birmingham and Lambeth Liver Evaluation Testing Strategies (BALLETS)

Health Technology Assessment 2013; Vol. 17: No. 28
DOI: 10.3310/hta17280

NIHR Journals Library www.journalslibrary.nihr.ac.uk

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.

Publication

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

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).