National Institute for Health and Care Excellence

Final version

Cirrhosis in over 16s

Assessment and management

NICE guideline NG50

Methods, evidence and recommendations

July 2016

Developed by the National Guideline Centre, hosted by the Royal College of Physicians

ISBN
978-1-4731-1997-0

Contents

Guideline Development Group members

Name	Role
Iain Brew	General Practitioner, Leeds Community Healthcare NHS Trust
Andrew Fowell	Consultant Hepatologist, Portsmouth Hospitals NHS Trust
Lynda Greenslade	Clinical Nurse Specialist in Hepatology, Royal Free London Foundation Trust
Phillip Harrison	Consultant Hepatologist, King's College Hospital (Chair)
Mark Hudson	Consultant Hepatologist, Newcastle Hospitals NHS Trust
Andrew Langford	Patient/Carer Member
Susan McRae	Patient/Carer Member
Marsha Morgan	Principal Research Associate and Honorary Consultant Physician, UCL Institute for Liver and Digestive Health, University College London Medical School
Gerri Mortimore	Lead CNS Liver, Derby Hospitals NHS Foundation Trust
Valerie Ross	Lead Pharmacist Hepatology, Barts Health NHS Trust

NGC technical team members

Name	Role
Katie Broomfield	Document Editor/Process Assistant
Jill Cobb	Information Scientist
Emily Davies	Senior Research Fellow
Lefteris Floros	Health Economist
Martin Harker	Senior Health Economist
Brian Hogan	Clinical Trainee, Specialist Registrar in Hepatology and Intensive Medicine
Katie Jones	Senior Project Manager (until December 2014)
Stacey Lockyer	Research Fellow (from September 2014 to May 2015)
Jill Parnham	Operations Director and Guideline Lead (until December 2014)
Ben Pordes	Project Manager (from December 2014)
Gill Ritchie	Associate Director and Guideline Lead (from December 2014)
Grit Scheffler	Research Fellow (from May 2015)
Heather Stegenga	Research Fellow (until November 2014)

Co-optees

Name	Role
David Fitzmaurice	Professor of Primary Care, University of Birmingham
John O'Grady	Professor of Hepatology, King's College Hospital
Phillip Johnson	Professor of Oncology, University of Liverpool & Clatterbridge Cancer Centre
Rachel Pryke	General Practitioner Partner, Winyates Health Centre, Redditch
Roy Sherwood	Professor of Clinical Biochemistry, King's College London

Acknowledgements

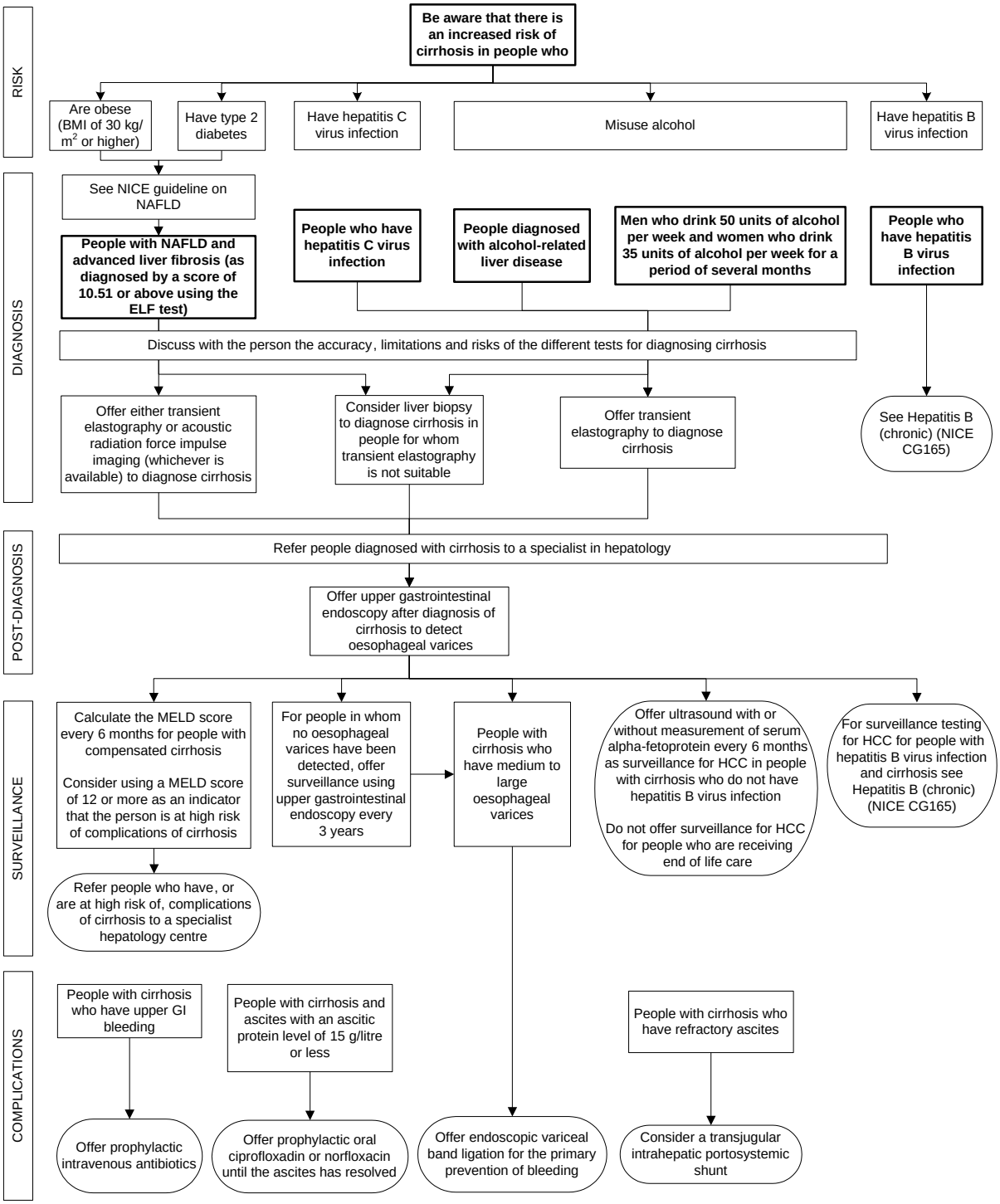
The development of this guideline was greatly assisted by the following people:

Joanna Ashe, Kate Ashmore, Serena Carville, Angela Cooper, Jessica Glen, Qudsia Malik, Josh South, Amelia Unsworth, Claire Wallnut, David Wonderling.

1 Guideline summary

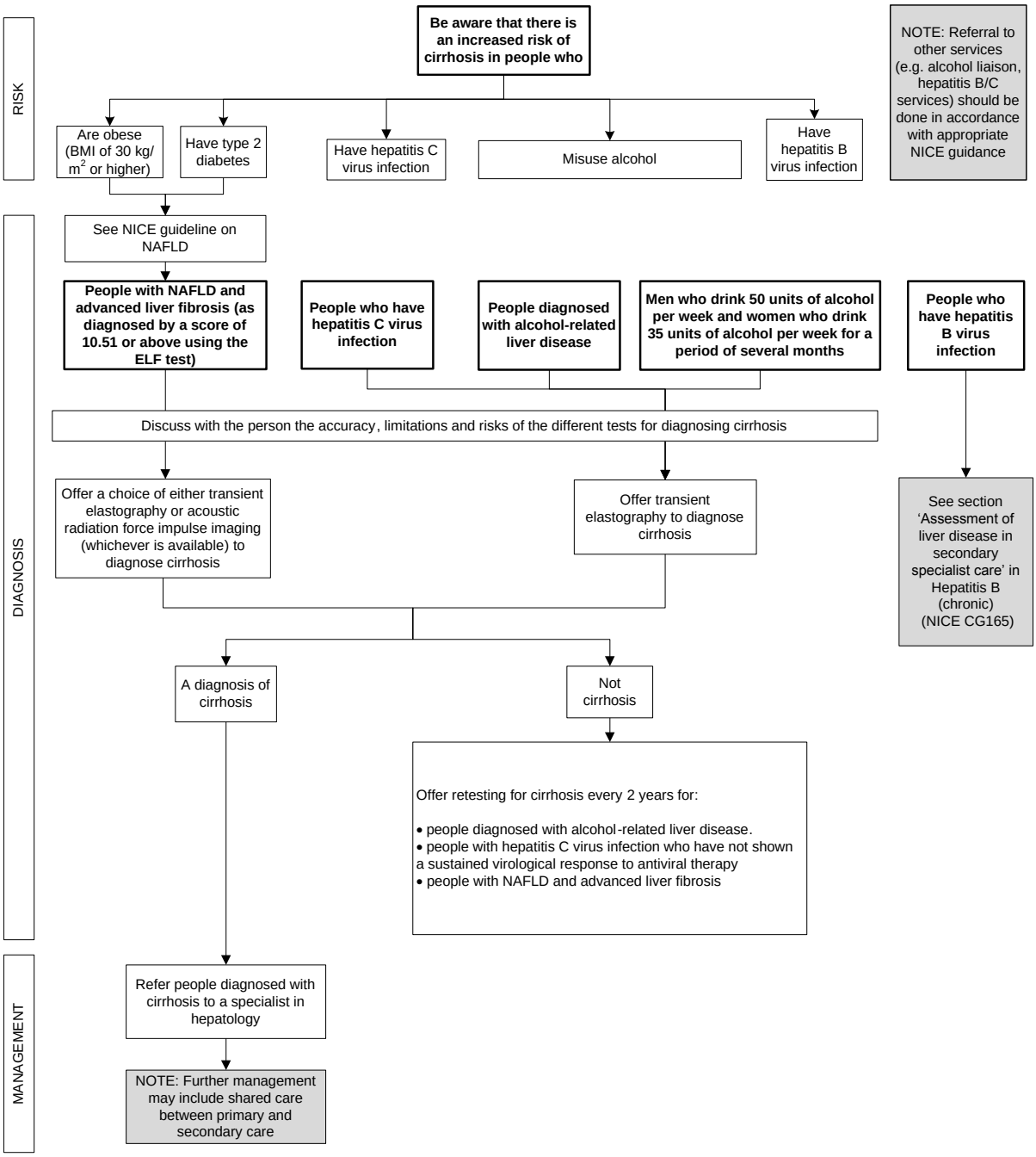
1.1 Algorithms

1.1.1 Cirrhosis in over 16s: assessment and management



1.1.2 Cirrhosis in over 16s: primary care pathway

Grey boxes indicate management outside of primary care.



1.2 Full list of recommendations

1. Be aware that there is an increased risk of cirrhosis in people who:
   - have hepatitis B virus infection
   - have hepatitis C virus infection
   - misuse alcohol
   - are obese (BMI of 30 kg/m^2 or higher)
   - have type 2 diabetes.

   Also see the NICE guidelines on: non-alcoholic fatty liver disease (NAFLD); alcohol-use disorders: diagnosis and management of physical complications; alcohol-use disorders: prevention; alcohol-use disorders: diagnosis, assessment and management of harmful drinking and alcohol dependence; type 2 diabetes in adults; obesity; and hepatitis B (chronic).
2. Discuss with the person the accuracy, limitations and risks of the different tests for diagnosing cirrhosis.
3. Offer transient elastography to diagnose cirrhosis for:
   - people with hepatitis C virus infection
   - men who drink over 50 units of alcohol per week and women who drink over 35 units of alcohol per week and have done so for several months
   - people diagnosed with alcohol-related liver disease.
4. Offer either transient elastography or acoustic radiation force impulse imaging (whichever is available) to diagnose cirrhosis for people with NAFLD and advanced liver fibrosis (as diagnosed by a score of 10.51 or above using the enhanced liver fibrosis [ELF] test). Also see the assessment for advanced liver fibrosis section in NICE's NAFLD guideline.
5. Consider liver biopsy to diagnose cirrhosis in people for whom transient elastography is not suitable.
6. For recommendations on diagnosing cirrhosis in people with hepatitis B virus infection, see the assessment of liver disease in secondary specialist care section in NICE's hepatitis B (chronic) guideline.
7. Do not offer tests to diagnose cirrhosis for people who are obese (BMI of 30 kg/m^2 or higher) or who have type 2 diabetes, unless they have NAFLD and advanced liver fibrosis (as diagnosed by a score of 10.51 or above using the ELF test). Also see the assessment for advanced liver fibrosis section in NICE's NAFLD guideline.
8. Ensure that healthcare professionals who perform or interpret non-invasive tests are trained to do so.
9. Do not use routine laboratory liver blood tests to rule out cirrhosis.
10. Refer people diagnosed with cirrhosis to a specialist in hepatology.
11. Offer retesting for cirrhosis every 2 years for:
    - people diagnosed with alcohol-related liver disease
    - people with hepatitis C virus infection who have not shown a sustained virological response to antiviral therapy

- people with NAFLD and advanced liver fibrosis.

12. For recommendations on reassessing liver disease in hepatitis B virus infection, see the assessment of liver disease in secondary specialist care section in NICE's hepatitis B (chronic) guideline.
13. Refer people who have, or are at high risk of, complications of cirrhosis to a specialist hepatology centre.
14. Calculate the Model for End-Stage Liver Disease (MELD) score every 6 months for people with compensated cirrhosis.
15. Consider using a MELD score of 12 or more as an indicator that the person is at high risk of complications of cirrhosis.
16. Offer ultrasound (with or without measurement of serum alpha-fetoprotein) every 6 months as surveillance for hepatocellular carcinoma (HCC) for people with cirrhosis who do not have hepatitis B virus infection.
17. For people with cirrhosis and hepatitis B virus infection, see the surveillance testing for hepatocellular carcinoma in adults with chronic hepatitis B section in NICE's hepatitis B (chronic) guideline.
18. Do not offer surveillance for HCC for people who are receiving end of life care.
19. After a diagnosis of cirrhosis, offer upper gastrointestinal endoscopy to detect oesophageal varices.
20. For people in whom no oesophageal varices have been detected, offer surveillance using upper gastrointestinal endoscopy every 3 years.
21. Offer endoscopic variceal band ligation for the primary prevention of bleeding for people with cirrhosis who have medium to large oesophageal varices.
22. Offer prophylactic intravenous antibiotics for people with cirrhosis who have upper gastrointestinal bleeding.
23. Review intravenous antibiotics prescriptions in line with the prescribing intravenous antimicrobials section in NICE's antimicrobial stewardship guideline.
24. Consider a transjugular intrahepatic portosystemic shunt for people with cirrhosis who have refractory ascites.
25. Offer prophylactic oral ciprofloxacin or norfloxacin[a] for people with cirrhosis and ascites with an ascitic protein of 15 g/litre or less, until the ascites has resolved.

[a] At the time of publication (July 2016), neither ciprofloxacin nor norfloxacin had a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

1.3 Key research recommendations

1. Development of a risk tool to identify people at risk of cirrhosis
2. Do non-selective beta-blockers improve survival and prevent first variceal bleeds in people with cirrhosis that is associated with small oesophageal varices?
3. What is the quality of life in people who have had a transjugular intrahepatic portosystemic shunt (TIPS)?
4. How frequently does antibiotic resistance occur, and how significant are antibiotic treatment-related complications when antibiotics are used for the primary prevention of spontaneous bacterial peritonitis in people at high risk of having, or developing, cirrhosis?
5. What is the most clinically and cost-effective volume replacer for patients with hepatorenal syndrome due to cirrhosis who are also receiving vasoactive drugs?
6. In people with cirrhosis and an acute episode of hepatic encephalopathy secondary to a clearly identified, potentially reversible precipitating factor, does management of the precipitating event alone improve the hepatic encephalopathy without specific treatment?

2 Introduction

People admitted to hospital with liver disease in England in 2012 were more likely to die compared to people admitted with other conditions (8.8% compared to 1.4%). Nearly half of the liver-related hospital admissions were for people with alcohol-related liver injury and 12.3% of these admissions resulted in death. Men accounted for over two-thirds of admissions for alcohol-related liver disease (ALD).[98] Finished admission episodes with a primary diagnosis of cirrhosis in English NHS hospitals rose from 3783 in 2005/06 to 5621 in 2014/15 (48.6% increase).[99] Consequently, the Chief Medical Officer has identified liver disease as one of the key issues for population health in England because it is the only major cause of mortality and morbidity that is increasing.[49]

Cirrhosis is a condition that occurs as a response to liver damage. It is characterised by disruption of the normal liver architecture and its replacement by fibrous bands of tissue and nodules of regenerating liver tissue. Cirrhosis usually develops over a period of years following exposure to one or more risk factors such as alcohol misuse, hepatitis B or C and non-alcoholic fatty liver disease (NAFLD), which cause inflammation and cell death within the liver. However, not everyone who is at risk will eventually develop cirrhosis; thus, the proportion of individuals who develop cirrhosis who abuse alcohol or have chronic viral hepatitis is around 10–20%, whereas with NAFLD it is around 5–10%.[235] Better recognition of individuals at risk of cirrhosis would allow for more timely intervention.

There is currently variation in practice across England and Wales for diagnostic tests for cirrhosis. Liver biopsy, performed in secondary care, has been the definitive diagnostic method for confirming cirrhosis but several non-invasive tests to predict cirrhosis have been developed, based on combining the results of routine laboratory liver blood tests, proprietary blood test panels that are surrogate markers of fibrogenesis, or imaging methods to measure the 'stiffness' of the liver. These tests can be performed in primary care or in an outpatient setting in secondary care; the results of the imaging tests are available immediately. Consequently, guidance is required to assess the clinical and cost-effectiveness, and patient acceptability of liver biopsy compared to the non-invasive tests of fibrosis in the confirmation of a diagnosis of cirrhosis.

People with cirrhosis may show no symptoms or signs of liver disease for many years and so do not come to attention until their disease progresses and they develop one or other of the major complications such as jaundice, fluid retention manifest as swelling of the abdomen and/or lower limbs, bleeding from their upper gastrointestinal tract or changes in their mental status. Thus, opportunities to intervene often come late. Presently there are no standard criteria for referral by primary care of a person with suspected cirrhosis for assessment in secondary care. Also there is a clear need for a cirrhosis risk assessment tool to assist in the identification of people who are at high risk of liver decompensation before they experience a defining event. Such a tool would inform the timing of referral of a person with compensated cirrhosis to specialist hepatology services for further assessment including suitability for liver transplantation.

This guideline provides advice on the assessment and management of people aged 16 years or older who are suspected or confirmed to have cirrhosis for clinicians in primary and secondary NHS-commissioned care. The aetiologies of cirrhosis in children and young people are generally different to those in adults (for example, biliary atresia), and the assessment and management of these conditions is different. However, the recommendations in this guideline may be useful to clinicians who are caring for young people in transition. The guideline sets out the critical pathways in the management of complications of cirrhosis and provides advice when to offer surveillance testing for the detection of hepatocellular carcinoma and oesophageal varices. It also gives recommendations for the prophylactic treatment of oesophageal varices to prevent variceal haemorrhage and the use of antibiotics for the primary prevention of spontaneous bacterial peritonitis. These recommendations establish the concepts that effort in the management of patients with confirmed cirrhosis should be directed either to the prevention of complications or early intervention to stabilise disease progression in order to avoid or delay clinical decompensation and the need for liver

transplantation. The recommendations in this guideline will facilitate the generation of a set of quality indicators for measurement of quality of care in people with cirrhosis.

3 Development of the guideline

3.1 What is a NICE clinical guideline?

NICE clinical guidelines are recommendations for the care of individuals in specific clinical conditions or circumstances within the NHS – from prevention and self-care through primary and secondary care to more specialised services. We base our clinical guidelines on the best available research evidence, with the aim of improving the quality of healthcare. We use predetermined and systematic methods to identify and evaluate the evidence relating to specific review questions.

NICE clinical guidelines can:

- provide recommendations for the treatment and care of people by health professionals
- be used to develop standards to assess the clinical practice of individual health professionals
- be used in the education and training of health professionals
- help patients to make informed decisions
- improve communication between patient and health professional.

While guidelines assist the practice of healthcare professionals, they do not replace their knowledge and skills.

We produce our guidelines using the following steps:

- Guideline topic is referred to NICE from the Department of Health.
- Stakeholders register an interest in the guideline and are consulted throughout the development process.
- The scope is prepared by the National Guideline Centre (NGC).
- The NGC establishes a Guideline Development Group.
- A draft guideline is produced after the group assesses the available evidence and makes recommendations.
- There is a consultation on the draft guideline.
- The final guideline is produced.

The NGC and NICE produce a number of versions of this guideline:

- the 'full guideline' contains all the recommendations, plus details of the methods used and the underpinning evidence
- the 'NICE guideline' lists the recommendations
- 'information for the public' is written using suitable language for people without specialist medical knowledge
- NICE Pathways brings together all connected NICE guidance.

This version is the full version. The other versions can be downloaded from NICE at www.nice.org.uk.

3.2 Remit

NICE received the remit for this guideline from the Department of Health. NICE commissioned the NGC to produce the guideline.

The remit for this guideline is:

To develop a clinical guideline on assessment and management of Cirrhosis

3.3 Who developed this guideline?

A multidisciplinary Guideline Development Group (GDG) comprising health professionals and researchers as well as lay members developed this guideline (see the list of Guideline Development Group members and the acknowledgements).

The National Institute for Health and Care Excellence (NICE) funds the National Guideline Centre (NGC) and thus supported the development of this guideline. The GDG was convened by the NGC and chaired by Dr Phil Harrison in accordance with guidance from NICE.

The group met approximately every 5–6 weeks during the development of the guideline. At the start of the guideline development process all GDG members declared interests including consultancies, fee-paid work, shareholdings, fellowships and support from the healthcare industry. At all subsequent GDG meetings, members declared arising conflicts of interest.

Members were either required to withdraw completely or for part of the discussion if their declared interest made it appropriate. The details of declared interests and the actions taken are shown in Appendix B.

Staff from the NGC provided methodological support and guidance for the development process. The team working on the guideline included a project manager, systematic reviewers (research fellows), health economists and information scientists. They undertook systematic searches of the literature, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate and drafted the guideline in collaboration with the GDG.

3.3.1 What this guideline covers

The guideline covers the identification and assessment of suspected cirrhosis, monitoring to detect complications and management of complications such as ascites and hepatorenal syndrome and referral for tertiary care. For further details please refer to the scope in Appendix A and the review questions in Section 4.1.

3.3.2 What this guideline does not cover

This guideline does not provide recommendations for people whose cirrhosis is diagnosed before the age of 16.

The guideline does not cover diagnosis, investigation and management of the underlying cause of cirrhosis, complications specific to the underlying cause of cirrhosis, liver transplantation, management of hepatocellular carcinoma or management of variceal haemorrhage.

3.3.3 Relationships between the guideline and other NICE guidance

Related NICE technology appraisals:

- Virtual Touch Quantification to diagnose and monitor liver fibrosis in chronic hepatitis B and C. NICE medical technology guidance 27 (2015).
- Sofosbuvir for treating chronic hepatitis C. NICE technology appraisal guidance 330 (2015).
- Simeprevir in combination with peginterferon alfa and ribavirin for treating genotypes 1 and 4 chronic hepatitis C. NICE technology appraisal guidance 331 (2015).
- Rifaximin for preventing episodes of overt hepatic encephalopathy. NICE medical technology guidance 337 (2015).

Related NICE diagnostics guidance:

- SonoVue (sulphur hexafluoride microbubbles) – contrast agent for contrast-enhanced ultrasound imaging of the liver. NICE diagnostics guidance 5 (2012).

Related NICE interventional procedures guidance:

- Subcutaneous implantation of a battery-operated catheter drainage system for managing refractory and recurrent ascites. NICE interventional procedure guidance 479 (2014).
- Stent insertion for bleeding oesophageal varices. NICE interventional procedure guidance 392 (2011).
- Extracorporeal albumin dialysis for acute liver failure. NICE interventional procedure guidance 316 (2009).

Related NICE public health guidance:

- Hepatitis B and C – ways to promote and offer testing. NICE public health guidance 43 (2013).
- Alcohol-use disorders: preventing harmful drinking. NICE public health guidance 24 (2010).

Related NICE guidelines:

- Non-alcoholic fatty liver disease (NAFLD): assessment and management. NICE guideline 49 (2016)
- Type 2 diabetes in adults: management. NICE guideline 28 (2015)
- Suspected cancer: recognition and referral. NICE guideline 12 (2015).
- Obesity: identification, assessment and management. NICE guideline 189 (2014)
- Hepatitis B (chronic). NICE clinical guideline 165 (2013).
- Acute upper gastrointestinal bleeding. NICE clinical guideline 141 (2012).
- Alcohol-use disorders: diagnosis, assessment and management of harmful drinking and alcohol dependence. NICE clinical guideline 115 (2011).
- Alcohol-use disorders: diagnosis and clinical management of alcohol-related physical complications. NICE clinical guideline 100 (2010).

Related NICE guidance currently in development:

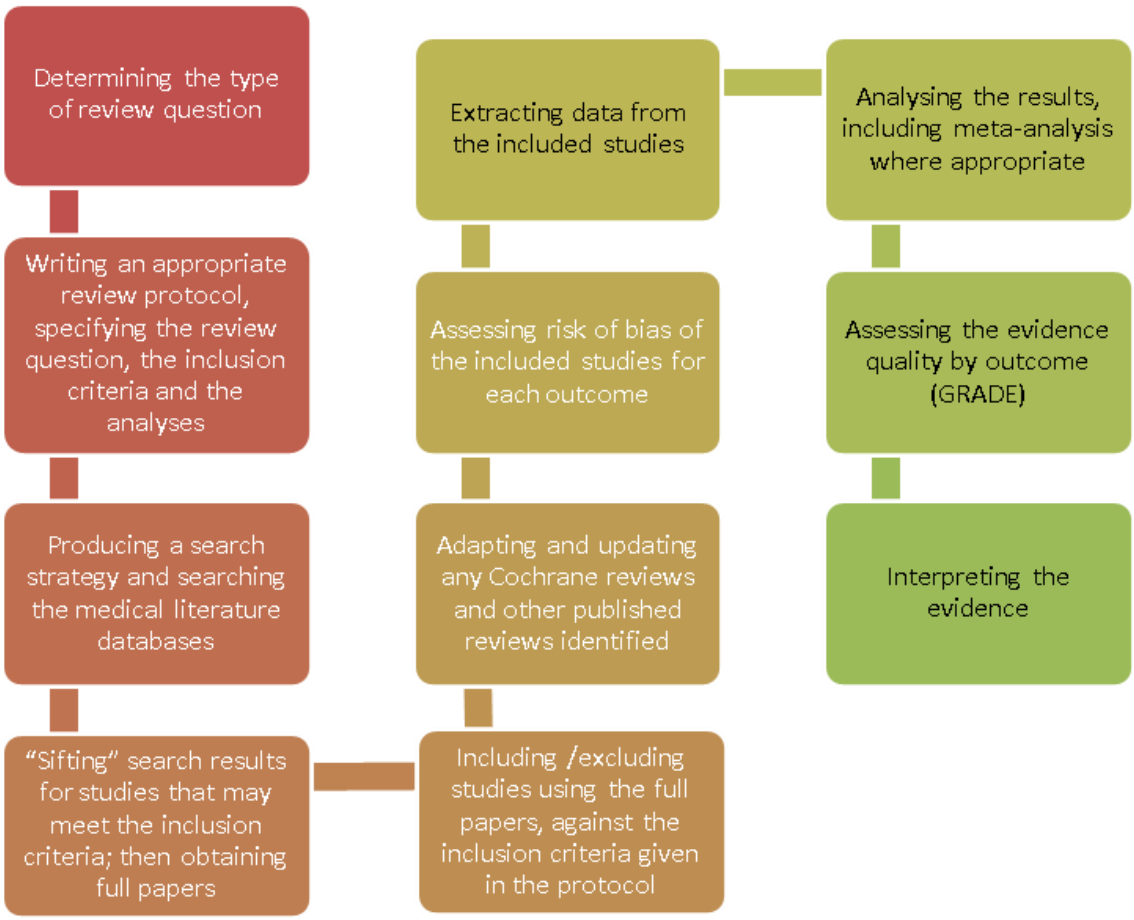
- Hepatitis C (chronic) – ombitasvir/paritaprevir/ritonavir (with or without dasabuvir). NICE technology appraisal 731. Publication expected November 2015.
- Hepatitis C (chronic) – ledipasvir-sofosbuvir. NICE technology appraisal 742. Publication expected November 2015.
- Hepatitis C (chronic) – daclatasvir. NICE technology appraisal 766. Publication expected November 2015.
- Hepatitis C (chronic) – simeprevir with sofosbuvir. NICE technology appraisal 887. Publication expected June 2016.
- Hepatitis C: diagnosis and management of hepatitis C. NICE guideline. Publication expected May 2018.

4 Methods

This chapter sets out in detail the methods used to review the evidence and to develop the recommendations that are presented in subsequent chapters of this guideline. This guidance was developed in accordance with the methods outlined in the NICE guidelines manual (the 2012 version was followed until consultation and 2014 version was followed from the start of consultation).[143,145]

Sections 4.1 to 4.3 describe the process used to identify and review clinical evidence (summarised in Figure 1), Sections 4.2 and 4.4 describe the process used to identify and review the health economic evidence, and Section 4.5 describes the process used to develop recommendations.

Figure 1: Step-by-step process of review of evidence in the guideline



4.1 Developing the review questions and outcomes

Review questions were developed using a PICO framework (patient, intervention, comparison and outcome) for intervention reviews; using a framework of population, index tests, reference standard and target condition for reviews of diagnostic test accuracy; and using population, presence or absence of factors under investigation (for example prognostic factors) and outcomes for prognostic reviews.

This use of a framework guided the literature searching process, critical appraisal and synthesis of evidence, and facilitated the development of recommendations by the GDG. The review questions were drafted by the NGC technical team and refined and validated by the GDG. The questions were based on the key clinical areas identified in the scope (Appendix A).

A total of 17 review questions were identified.

Full literature searches, critical appraisals and evidence reviews were completed for all the specified review questions.

Table 1: Review questions

Chapter	Type of review	Review questions	Outcomes
5	Prognostic	What are the risk factors that indicate the populations at specific risk for cirrhosis?	**Critical outcome:** • Diagnosis of cirrhosis
5	Prognostic risk tool	Are there any validated risk tools that indicate the populations at specific risk for cirrhosis?	**Critical outcomes:** • Area under curve • Sensitivity • Specificity • Calibration plot
6	Diagnostic	In people with suspected (or under investigation for) cirrhosis what is the most accurate blood fibrosis test to identify whether cirrhosis is present?	**Critical outcomes:** • Sensitivity • Specificity **Important outcome:** • Area under curve
6	Diagnostic	In people with suspected (or under investigation for) cirrhosis what is the most accurate non-invasive imaging test to identify whether cirrhosis is present?	**Critical outcomes:** • Sensitivity • Specificity **Important outcome:** • Area under curve
6	Diagnostic	In people with suspected (or under investigation for) cirrhosis is the most accurate blood fibrosis test more accurate compared to an individual blood test to identify whether cirrhosis is present?	**Critical outcomes:** • Sensitivity • Specificity **Important outcomes:** • Area under curve
6	Diagnostic	In people with suspected (or under investigation for) cirrhosis is a combination of 2 non-invasive tests more accurate compared to a blood fibrosis test alone or an imaging test alone to identify whether cirrhosis is present?	**Critical outcomes:** • Sensitivity • Specificity **Important outcome:** • Area under curve
7	Prognostic risk tool	Which risk assessment tool is the most accurate and cost-effective for predicting the risk of morbidity and mortality in people with compensated cirrhosis?	**Critical outcomes:** • Area under curve • Sensitivity • Specificity • Calibration plot
7	Prognostic risk tool	When (at what severity score on the risk assessment tool) should people with cirrhosis be referred to specialist care?	**Critical outcomes:** • Area under curve • Sensitivity • Specificity • Calibration plot
8	Intervention	When and how frequently should	**Critical outcomes:**

Chapter	Type of review	Review questions	Outcomes
		surveillance testing be offered for the early detection of hepatocellular carcinoma (HCC) in people with cirrhosis?	• Transplant-free survival • Quality of life **Important outcomes:** • HCC occurrence • Lesion of HCC less than or equal to 3 cm • Number of lesions • Liver cancer staging • Liver transplant
9	Intervention	How frequently should surveillance testing using endoscopy be offered for the detection of oesophageal varices and isolated gastric varices in people with cirrhosis?	**Critical outcomes:** • Survival • Free from variceal bleeding • Quality of life **Important outcomes:** • Free from varices • Occurrence of moderate or large varices • Size of varices • Number receiving prophylactic treatment
10	Intervention	What is the clinical and cost-effectiveness of non-selective beta-blockers for the primary prevention of bleeding in people with oesophageal varices due to cirrhosis?	**Critical outcomes:** • Quality of life • Survival • Free from primary variceal bleeding **Important outcomes:** • Hospital admission • Hospital length of stay • Primary upper gastrointestinal bleeding (irrespective of bleeding source) • Bleeding-related mortality • Adverse events
10	Intervention	What is the clinical and cost-effectiveness of endoscopic band ligation for the primary prevention of bleeding in people with oesophageal varices due to cirrhosis?	**Critical outcomes:** • Quality of life • Survival • Free from primary variceal bleeding **Important outcomes:** • Hospital admission • Hospital length of stay • Primary upper gastrointestinal bleeding (irrespective of bleeding source) • Bleeding-related mortality • Adverse events
10	Intervention	What is the clinical and cost-effectiveness of non-selective beta-blockers compared with endoscopic	**Critical outcomes:** • Quality of life

Chapter	Type of review	Review questions	Outcomes
		band ligation for the primary prevention of bleeding in people with oesophageal varices due to cirrhosis?	• Survival • Free from primary variceal bleeding **Important outcomes:** • Hospital admission • Hospital length of stay • Primary upper gastrointestinal bleeding (irrespective of bleeding source) • Bleeding-related mortality • Adverse events
11	Intervention	What is the most clinically and cost-effective prophylactic antibiotic for the primary prevention of bacterial infections in people with cirrhosis and upper gastrointestinal bleeding?	**Critical outcomes:** • Occurrence of bacterial infections • Quality of life • All-cause mortality **Important outcomes:** • Adverse effect: renal failure • Length of hospital stay • Readmission rate • Antibiotic complications
12	Intervention	What is the clinical and cost-effectiveness of transjugular intrahepatic portosystemic shunt (TIPS) compared with large-volume paracentesis (LVP) with albumin in the management of diuretic-resistant ascites due to cirrhosis?	**Critical outcomes:** • Re-accumulation of ascites • Quality of life • Transplant-free survival **Important outcomes:** • Spontaneous bacterial peritonitis • Renal failure • Hepatic encephalopathy • Length of hospital stay • Readmission rate
13	Intervention	What is the clinical and cost-effectiveness of antibiotics compared with placebo for the primary prevention of spontaneous bacterial peritonitis (SBP) in people with cirrhosis and ascites?	**Critical outcomes:** • Occurrence of SBP • Quality of life • All-cause mortality **Important outcomes:** • Adverse event: incidence of resistant organisms • Adverse effect: renal failure • Adverse effect: liver failure • Length of hospital stay • Readmission rate
14	Intervention	Which is the most clinically and cost-effective volume replacer for patients with hepatorenal syndrome due to cirrhosis who are also receiving vasoactive drugs?	**Critical outcomes:** • Survival • Quality of life (continuous) • Reversal of hepatorenal syndrome or improved renal function

Chapter	Type of review	Review questions	Outcomes
			Important outcomes: • Time to discharge from hospital • Readmission to hospital • Adverse events of volume replacement (infection) • Adverse events of volume replacement (heart failure)
15	Intervention	What is the most clinically and cost-effective intervention for the first-line treatment of an episode of acute hepatic encephalopathy in people with cirrhosis?	**Critical outcomes:** • Survival • No improvement in hepatic encephalopathy • Quality of life **Important outcomes:** • Time to discharge from hospital • Adverse events (diarrhoea, flatulence, abdominal pain, nausea, gastrointestinal bleeding, renal failure)

4.2 Searching for evidence

4.2.1 Clinical literature search

Systematic literature searches were undertaken to identify all published clinical evidence relevant to the review questions. Searches were undertaken according to the parameters stipulated within the NICE guidelines manual.[143,145] Databases were searched using relevant medical subject headings, free-text terms and study-type filters where appropriate. Where possible, searches were restricted to articles published in English. Studies published in languages other than English were not reviewed. All searches were conducted in Medline, Embase, and The Cochrane Library. All searches were updated on 24 August 2015. No papers published after this date were considered.

Search strategies were quality assured by cross-checking reference lists of highly relevant papers, analysing search strategies in other systematic reviews, and asking GDG members to highlight any additional studies. Searches were quality assured by a second information scientist before being run. The questions, the study types applied, the databases searched and the years covered can be found in Appendix G.

The titles and abstracts of records retrieved by the searches were sifted for relevance, with potentially significant publications obtained in full text. These were assessed against the inclusion criteria.

During the scoping stage, a search was conducted for guidelines and reports on the websites listed below from organisations relevant to the topic.

- Guidelines International Network database (www.g-i-n.net)
- National Guideline Clearing House (www.guideline.gov)
- National Institute for Health and Care Excellence (NICE) (www.nice.org.uk)
- National Institutes of Health Consensus Development Program (consensus.nih.gov)
- NHS Evidence Search (www.evidence.nhs.uk).

All references sent by stakeholders were considered. Searching for unpublished literature was not undertaken. The NGC and NICE do not have access to drug manufacturers' unpublished clinical trial results, so the clinical evidence considered by the GDG for pharmaceutical interventions may be different from that considered by the Medicines and Healthcare Products Regulatory Agency (MHRA) and European Medicines Agency for the purposes of licensing and safety regulation.

4.2.2 Health economic literature search

Systematic literature searches were also undertaken to identify health economic evidence within published literature relevant to the review questions. The evidence was identified by conducting a broad search relating to cirrhosis in the NHS Economic Evaluation Database (NHS EED), the Health Technology Assessment database (HTA) and the Health Economic Evaluations Database (HEED) with no date restrictions (NHS EED ceased to be updated after March 2015; HEED was used for searches up to 27 August 2014 but subsequently ceased to be available). Additionally, the search was run on Medline and Embase using a health economic filter, from 2013 to ensure recent publications that had not yet been indexed by the economic databases were identified. This was supplemented by an additional search that looked for economic papers specifically relating to the modelling of liver disease on Medline, Embase, HTA, NHS EED and HEED to ensure no modelling studies were missed. Where possible, searches were restricted to articles published in English. Studies published in languages other than English were not reviewed.

The health economic search strategies are included in Appendix G. All searches were updated on 27 August 2015. No papers published after this date were considered.

4.3 Identifying and analysing evidence of effectiveness

Research fellows conducted the tasks listed below, which are described in further detail in the rest of this section:

- Identified potentially relevant studies for each review question from the relevant search results by reviewing titles and abstracts. Full papers were then obtained.
- Reviewed full papers against prespecified inclusion and exclusion criteria to identify studies that addressed the review question in the appropriate population, and reported on outcomes of interest (review protocols are included in Appendix C).
- Critically appraised relevant studies using the appropriate study design checklist as specified in the NICE guidelines manual.[143,145] Prognostic or qualitative studies were critically appraised using NGC checklists.
- Extracted key information about interventional study methods and results using 'Evibase', NGC's purpose-built software. Evibase produces summary evidence tables, including critical appraisal ratings. Key information about non-interventional study methods and results was manually extracted onto standard evidence tables and critically appraised separately (evidence tables are included in Appendix H).
- Generated summaries of the evidence by outcome. Outcome data were combined, analysed and reported according to study design:
    - Randomised data were meta-analysed where appropriate and reported in GRADE profile tables.
    - Observational data were presented as a range of values in GRADE profile tables or meta-analysed if appropriate.
    - Prognostic data were meta-analysed where appropriate and reported in GRADE profile tables.
    - Diagnostic data studies were meta-analysed where appropriate or presented as a range of values in adapted GRADE profile tables

- A sample of a minimum of 10% of the abstract lists of the first 3 sifts by new reviewers were double-sifted by a senior research fellow. As no papers were missed by any reviewers, no further double-sifting was carried out. All of the evidence reviews were quality assured by a senior research fellow. This included checking:
  - papers were included or excluded appropriately,
  - a sample of the data extractions,
  - correct methods were used to synthesise data
  - a sample of the risk of bias assessments.

4.3.1 Inclusion and exclusion criteria

The inclusion and exclusion of studies was based on the criteria defined in the review protocols, which can be found in Appendix C. Excluded studies by review question (with the reasons for their exclusion) are listed in Appendix L. The GDG was consulted about any uncertainty regarding inclusion or exclusion.

The key population inclusion criterion was:

- Adults and young people (16 years and over) with cirrhosis

The key population exclusion criterion was:

- Children <16 years with cirrhosis

Conference abstracts were not automatically excluded from any review. The abstracts were initially assessed against the inclusion criteria for the review question and further processed when a full publication was not available for that review question. If the abstracts were included the authors were contacted for further information. No relevant conference abstracts were identified for this guideline. Literature reviews, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded.

4.3.2 Type of studies

Randomised trials, non-randomised trials, and observational studies (including diagnostic or prognostic studies) were included in the evidence reviews as appropriate.

For most intervention reviews in this guideline, parallel randomised controlled trials (RCTs) were included because they are considered the most robust type of study design that can produce an unbiased estimate of the intervention effects. If non-randomised studies were appropriate for inclusion (for example, non-drug trials with no randomised evidence) the GDG stated a priori in the protocol that either certain identified variables must be equivalent at baseline or else the analysis had to adjust for any baseline differences. If the study did not fulfil either criterion it was excluded. Please refer to the review protocols in Appendix C for full details on the study design of studies selected for each review question.

For diagnostic review questions, diagnostic RCTs, cross-sectional studies and retrospective studies were included. For prognostic review questions, prospective and retrospective cohort studies were included. Case-control studies were not included.

Where data from observational studies were included, the results for each outcome were presented separately for each study and meta-analysis was not conducted.

4.3.3 Methods of combining clinical studies

4.3.3.1 Data synthesis for intervention reviews

Where possible, meta-analyses were conducted using Cochrane Review Manager (RevMan5)[2] software to combine the data given in all studies for each of the outcomes of interest for the review question.

For some questions stratification was used, and this is documented in the individual review question protocols (see Appendix C).

4.3.3.1.1 Analysis of different types of data

Dichotomous outcomes

Fixed-effects (Mantel-Haenszel) techniques (using an inverse variance method for pooling) were used to calculate risk ratios (relative risk, RR) for the binary outcomes. The absolute risk difference was also calculated using GRADEpro[91] software, using the median event rate in the control arm of the pooled results.

For binary variables where there were zero events in either arm or a less than 1% event rate, Peto odds ratios, rather than risk ratios, were calculated. Peto odds ratios are more appropriate for data with a low number of events.

Where there was sufficient information provided, Hazard Ratios were calculated in preference for outcomes such as mortality where the time to the event occurring was important for decision-making.

Continuous outcomes

Continuous outcomes were analysed using an inverse variance method for pooling weighted mean differences. Where the studies within a single meta-analysis had different scales of measurement, standardised mean differences were used (providing all studies reported either change from baseline or final values rather than a mixture of the 2); each different measure in each study was 'normalised' to the standard deviation value pooled between the intervention and comparator groups in that same study.

The means and standard deviations of continuous outcomes are required for meta-analysis. However, in cases where standard deviations were not reported, the standard error was calculated if the p values or 95% confidence intervals (95% CI) were reported, and meta-analysis was undertaken with the mean and standard error using the generic inverse variance method in Cochrane Review Manager (RevMan5)[2] software. Where p values were reported as 'less than', a conservative approach was undertaken. For example, if a p value was reported as 'p≤0.001', the calculations for standard deviations were based on a p value of 0.001. If these statistical measures were not available then the methods described in Section 16.1.3 of the Cochrane Handbook (version 5.1.0, updated March 2011) were applied.

4.3.3.1.2 Generic inverse variance

If a study reported only the summary statistic and 95% CI the generic inverse variance method was used to enter data into RevMan5.[2] If the control event rate was reported this was used to generate the absolute risk difference in GRADEpro.[91] If multivariate analysis was used to derive the summary statistic but no adjusted control event rate was reported no absolute risk difference was calculated.

4.3.3.1.3 Heterogeneity

Statistical heterogeneity was assessed for each meta-analysis estimate by considering the chi-squared test for significance at $p<0.1$ or an I-squared (I^2) inconsistency statistic (with an I-squared value of more than 50% indicating significant heterogeneity) as well as the distribution of effects. Where significant heterogeneity was present, predefined subgrouping of studies was carried out.

If the subgroup analysis resolved heterogeneity within all of the derived subgroups, then each of the derived subgroups were adopted as separate outcomes (providing at least 1 study remained in each subgroup). Assessments of potential differences in effect between subgroups were based on the chi-squared tests for heterogeneity statistics between subgroups. Any subgroup differences were interpreted with caution as separating the groups breaks the study randomisation and as such is subject to uncontrolled confounding.

If all predefined strategies of subgrouping were unable to explain statistical heterogeneity within each derived subgroup, then a random effects (DerSimonian and Laird) model was employed to the entire group of studies in the meta-analysis. A random effects model assumes a distribution of populations, rather than a single population. This leads to a widening of the confidence interval around the overall estimate, thus providing a more realistic interpretation of the true distribution of effects across more than 1 population. If, however, the GDG considered the heterogeneity was so large that meta-analysis was inappropriate, then the results were described narratively.

4.3.3.1.4 Complex analysis

Network meta-analysis was considered for the comparison of interventional treatments for acute hepatic encephalopathy, but was not pursued because of insufficient data available for the relevant outcomes.

4.3.3.2 Data synthesis for prognostic factor reviews

Odds ratios (ORs), risk ratios (RRs), or hazard ratios (HRs), with their 95% CIs, for the effect of the prespecified prognostic factors were extracted from the studies. Studies were only included if the confounders prespecified by the GDG were either matched at baseline or were adjusted for in multivariate analysis.

Studies of lower risk of bias were preferred, taking into account the analysis and the study design. In particular, prospective cohort studies were preferred if they reported multivariable analyses that adjusted for key confounders identified by the GDG at the protocol stage for that outcome. Data were not combined in meta-analyses for prognostic studies.

4.3.3.3 Data synthesis for diagnostic test accuracy reviews

Two review protocols were produced to reflect the 2 different diagnostic study designs.

4.3.3.3.1 Diagnostic RCTs

Diagnostic RCTs (sometimes referred to as test and treat trials) are a randomised comparison of 2 diagnostic tests, with study outcomes being clinically important consequences of the diagnosis (patient-related outcome measures similar to those in intervention trials, such as mortality). Patients are randomised to receive test A or test B, followed by identical therapeutic interventions based on the results of the test (so someone with a positive result would receive the same treatment regardless of whether they were diagnosed by test A or test B). Downstream patient outcomes are then compared between the 2 groups. As treatment is the same in both arms of the trial, any differences in patient outcomes will reflect the accuracy of the tests in correctly establishing who does and does not have the condition. Data were synthesised using the same methods for intervention reviews (see Section 4.3.3.1.1 above).

4.3.3.3.2 Diagnostic accuracy studies

For diagnostic test accuracy studies, a positive result on the index test was found if the patient had values of the measured quantity above or below a threshold value, and different thresholds could be used. The thresholds were prespecified by the GDG including whether or not data could be pooled across a range of thresholds. Diagnostic test accuracy measures used in the analysis were: area under the receiver operating characteristics (ROC) curve (AUC), and, for different thresholds (if appropriate), sensitivity and specificity. The threshold of a diagnostic test is defined as the value at which the test can best differentiate between those with and without the target condition. In practice this varies amongst studies. If a test has a high sensitivity then very few people with the condition will be missed (few false negatives). For example, a test with a sensitivity of 97% will only miss 3% of people with the condition. Conversely, if a test has a high specificity then few people without the condition would be incorrectly diagnosed (few false positives). For example, a test with a specificity of 97% will only incorrectly diagnose 3% of people who do not have the condition as positive. For this guideline, sensitivity was considered more important than specificity due to the consequences of a missed diagnosis of cirrhosis (false negative result).

Coupled forest plots of sensitivity and specificity with their 95% CIs across studies (at various thresholds) were produced for each test, using RevMan5.[2] In order to do this, 2×2 tables (the number of true positives, false positives, true negatives and false negatives) were directly taken from the study if given, or else were derived from raw data or calculated from the set of test accuracy statistics.

Diagnostic meta-analysis was conducted where appropriate, that is, when 3 or more studies were available per threshold. Test accuracy for the studies was pooled using the bivariate method for the direct estimation of summary sensitivity and specificity using a random effects approach in WinBUGS software.[3] The advantage of this approach is that it produces summary estimates of sensitivity and specificity that account for the correlation between the 2 statistics. Other advantages of this method have been described elsewhere.[172,238,239] The bivariate method uses logistic regression on the true positives, true negatives, false positives and false negatives reported in the studies. Overall sensitivity and specificity and confidence regions were plotted (using methods outlined by Novielli 2010.[150]) Pooled sensitivity and specificity and their 95% CIs were reported in the clinical evidence summary tables. For thresholds with fewer than 3 studies, median sensitivity and the paired specificity were reported where possible.

Heterogeneity or inconsistency amongst studies was visually inspected in the forest plots.

Area under the ROC curve (AUC) data for each study was also plotted on a graph, for each diagnostic test. The AUC describes the overall diagnostic accuracy across the full range of thresholds. The following criteria were used for evaluating AUCs:

- ≤0.50: worse than chance
- 0.50–0.60: very poor
- 0.61–0.70: poor
- 0.71–0.80: moderate
- 0.81–0.92: good
- 0.91–1.00: excellent or perfect test.

Heterogeneity or inconsistency amongst studies was visually inspected.

4.3.3.4 Data synthesis for risk prediction rules

Evidence reviews on risk prediction rules or risk prediction tool results were presented separately for discrimination and calibration. The discrimination data were analysed according to the principles of

data synthesis for diagnostic accuracy studies as outlined in Section 4.3.3.3.2. Calibration data such as r-squared (R^2), if reported, were presented separately to the discrimination data. The results were presented for each study separately along with the quality rating for the study.

4.3.4 Appraising the quality of evidence by outcomes

4.3.4.1 Intervention reviews

The evidence for outcomes from the included RCTs and, where appropriate, observational studies were evaluated and presented using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group (http://www.gradeworkinggroup.org/). The software (GRADEpro[91]) developed by the GRADE working group was used to assess the quality of each outcome, taking into account individual study quality and the meta-analysis results.

Each outcome was first examined for each of the quality elements listed and defined in Table 2.

Table 2: Description of quality elements in GRADE for intervention studies

Quality element	Description
Risk of bias	Limitations in the study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect. Examples of such limitations are selection bias (often due to poor allocation concealment), performance and detection bias (often due to a lack of blinding of the patient, healthcare professional or assessor) and attrition bias (due to missing data causing systematic bias in the analysis).
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of effect estimates between studies in the same meta-analysis.
Imprecision	Results are imprecise when studies include relatively few patients and few events (or highly variable measures) and thus have wide confidence intervals around the estimate of the effect relative to clinically important thresholds. 95% confidence intervals denote the possible range of locations of the true population effect at a 95% probability, and so wide confidence intervals may denote a result that is consistent with conflicting interpretations (for example a result may be consistent with both clinical benefit AND clinical harm) and thus be imprecise.
Publication bias	Publication bias is a systematic underestimate or overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. A closely related phenomenon is where some papers fail to report an outcome that is inconclusive, thus leading to an overestimate of the effectiveness of that outcome.
Other issues	Sometimes randomisation may not adequately lead to group equivalence of confounders, and if so this may lead to bias, which should be taken into account. Potential conflicts of interest, often caused by excessive pharmaceutical company involvement in the publication of a study, should also be noted.

Details of how the 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) were appraised for each outcome are given below. Publication or other bias was only taken into consideration in the quality assessment if it was apparent.

4.3.4.1.1 Risk of bias

The main domains of bias for RCTs are listed in Table 3. Each outcome had its risk of bias assessed within each study first. For each study, if there were no risks of bias in any domain, the risk of bias was given a rating of 0. If there was risk of bias in just 1 domain, the risk of bias was given a 'serious'

rating of −1, but if there was risk of bias in 2 or more domains the risk of bias was given a 'very serious' rating of −2. A weighted average score was then calculated across all studies contributing to the outcome, by taking into account the weighting of studies according to study precision. For example if the most precise studies tended to each have a score of −1 for that outcome, the overall score for that outcome would tend towards −1.

Table 3: Principle domains of bias in randomised controlled trials

Limitation	Explanation
Selection bias (sequence generation and allocation concealment)	If those enrolling patients are aware of the group to which the next enrolled patient will be allocated, either because of a non-random sequence that is predictable, or because a truly random sequence was not concealed from the researcher, this may translate into systematic selection bias. This may occur if the researcher chooses not to recruit a participant into that specific group because of: • knowledge of that participant's likely prognostic characteristics, and • a desire for one group to do better than the other.
Performance and detection bias (lack of blinding of patients and healthcare professionals)	Patients, caregivers, those adjudicating or recording outcomes, and data analysts should not be aware of the arm to which patients are allocated. Knowledge of the group can influence: • the experience of the placebo effect • performance in outcome measures • the level of care and attention received, and • the methods of measurement or analysis all of which can contribute to systematic bias.
Attrition bias	Attrition bias results from an unaccounted for loss of data beyond a certain level (a differential of 10% between groups). Loss of data can occur when participants are compulsorily withdrawn from a group by the researchers (for example, when a per-protocol approach is used) or when participants do not attend assessment sessions. If the missing data are likely to be different from the data of those remaining in the groups, and there is a differential rate of such missing data from groups, systematic attrition bias may result.
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results can also lead to bias, as this may distort the overall impression of efficacy.
Other limitations	For example: • Stopping early for benefit observed in randomised trials, in particular in the absence of adequate stopping rules. • Use of unvalidated patient-reported outcome measures. • Lack of washout periods to avoid carry-over effects in crossover trials. • Recruitment bias in cluster-randomised trials.

4.3.4.1.2 Indirectness

Indirectness refers to the extent to which the populations, interventions, comparisons and outcome measures are dissimilar to those defined in the inclusion criteria for the reviews. Indirectness is important when these differences are expected to contribute to a difference in effect size, or may affect the balance of harms and benefits considered for an intervention. As for the risk of bias, each outcome had its indirectness assessed within each study first. For each study, if there were no sources of indirectness, indirectness was given a rating of 0. If there was indirectness in just 1 source (for example in terms of population), indirectness was given a 'serious' rating of −1, but if there was indirectness in 2 or more sources (for example, in terms of population and treatment) the indirectness was given a 'very serious' rating of −2. A weighted average score was then calculated across all studies contributing to the outcome by taking into account study precision. For example, if

the most precise studies tended to have an indirectness score of −1 each for that outcome, the overall score for that outcome would tend towards −1.

4.3.4.1.3 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. When estimates of the treatment effect across studies differ widely, this suggests true differences in the underlying treatment effect, which may be due to differences in populations, settings or doses. When heterogeneity existed within an outcome (chi-squared $p<0.1$, or $I^2>50\%$), but no plausible explanation could be found, the quality of evidence for that outcome was downgraded. Inconsistency for that outcome was given a 'serious' score of −1 if the I^2 was 50–74%, and a 'very serious' score of −2 if the I^2 was 75% or more.

If inconsistency could be explained based on prespecified subgroup analysis (that is, each subgroup had an $I^2<50\%$), the GDG took this into account and considered whether to make separate recommendations on new outcomes based on the subgroups defined by the assumed explanatory factors. In such a situation the quality of evidence was not downgraded for those emergent outcomes.

Since the inconsistency score was based on the meta-analysis results, the score represented the whole outcome and so weighted averaging across studies was not necessary.

4.3.4.1.4 Imprecision

The criteria applied for imprecision were based on the 95% CIs for the pooled estimate of effect, and the minimal important differences (MID) for the outcome. The MIDs are the threshold for appreciable benefits and harms, separated by a zone either side of the line of no effect where there is assumed to be no clinically important effect. If either end of the 95% CI of the overall estimate of effect crossed 1 of the MID lines, imprecision was regarded as serious and a 'serious' score of −1 was given. This was because the overall result, as represented by the span of the confidence interval, was consistent with 2 interpretations as defined by the MID (for example, both no clinically important effect and clinical benefit were possible interpretations). If both MID lines were crossed by either or both ends of the 95% CI then imprecision was regarded as very serious and a 'very serious' score of −2 was given. This was because the overall result was consistent with all 3 interpretations defined by the MID (no clinically important effect, clinical benefit and clinical harm). This is illustrated in Figure 2. As for inconsistency, since the imprecision score was based on the meta-analysis results, the score represented the whole outcome and so weighted averaging across studies was not necessary.

The position of the MID lines is ideally determined by values reported in the literature. 'Anchor-based' methods aim to establish clinically meaningful changes in a continuous outcome variable by relating or 'anchoring' them to patient-centred measures of clinical effectiveness that could be regarded as gold standards with a high level of face validity. For example, a MID for an outcome could be defined by the minimum amount of change in that outcome necessary to make patients feel their quality of life had 'significantly improved'. MIDs in the literature may also be based on expert clinician or consensus opinion concerning the minimum amount of change in a variable deemed to affect quality of life or health. For binary variables, any MIDs reported in the literature will inevitably be based on expert consensus, as such MIDs relate to all-or-nothing population effects rather than measurable effects on an individual, and so are not amenable to patient-centred 'anchor' methods.

In the absence of values identified in the literature, the alternative approach to deciding on MID levels is the 'default' method, as follows:

- For categorical outcomes the MIDs were taken to be RRs of 0.75 and 1.25. For 'positive' outcomes such as 'patient satisfaction', the RR of 0.75 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant

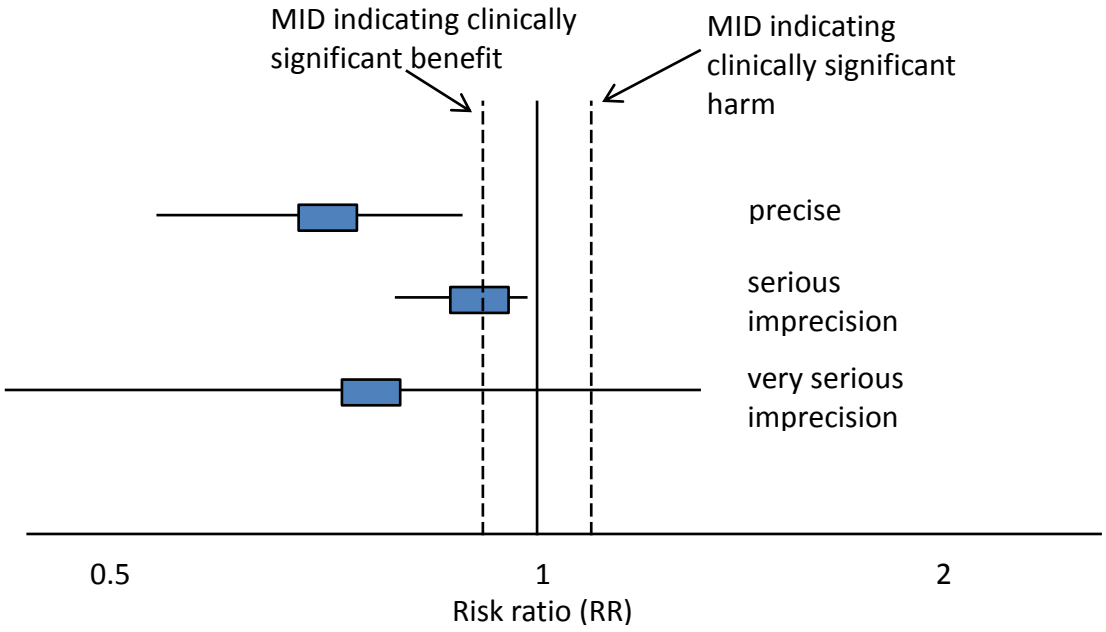
benefit. For 'negative' outcomes such as 'bleeding', the opposite occurs, so the RR of 0.75 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm.

- For mortality any change was considered to be clinically important and the imprecision was assessed on the basis of whether the confidence intervals crossed the line of no effect: that is, whether the result was consistent with both benefit and harm.
- For continuous outcome variables the MID was taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically significant benefit was positive for a 'positive' outcome (for example, a quality of life measure where a higher score denotes better health), and negative for a 'negative' outcome (for example, a visual analogue scale [VAS] pain score). Clinically significant harms will be the converse of these. If baseline values are unavailable, then half the median comparator group standard deviation of that variable will be taken as the MID.
- If standardised mean differences have been used, then the MID will be set at the absolute value of +0.5. This follows because standardised mean differences are mean differences normalised to the pooled standard deviation of the 2 groups, and are thus effectively expressed in units of 'numbers of standard deviations'. The 0.5 MID value in this context therefore indicates half a standard deviation, the same definition of MID as used for non-standardised mean differences.

The default MID value was subject to amendment after discussion with the GDG. If the GDG decided that the MID level should be altered, after consideration of absolute as well as relative effects, this was allowed, provided that any such decision was not influenced by any bias towards making stronger or weaker recommendations for specific outcomes.

For this guideline, no appropriate MIDs for continuous or dichotomous outcomes were found in the literature, and so the default method was adopted.

Figure 2: Illustration of precise and imprecise outcomes based on the 95% CI of dichotomous outcomes in a forest plot (Note that all 3 results would be pooled estimates, and would not, in practice, be placed on the same forest plot)



4.3.4.1.5 Overall grading of the quality of clinical evidence

Once an outcome had been appraised for the main quality elements, as above, an overall quality grade was calculated for that outcome. The scores (0, −1 or −2) from each of the main quality elements were summed to give a score that could be anything from 0 (the best possible) to −8 (the worst possible). However scores were capped at −3. This final score was then applied to the starting grade that had originally been applied to the outcome by default, based on study design. All RCTs started as High and the overall quality became Moderate, Low or Very Low if the overall score was −1, −2 or −3 points respectively. The significance of these overall ratings is explained in Table 4. The reasons for downgrading in each case were specified in the footnotes of the GRADE tables.

Observational interventional studies started at Low, and so a score of −1 would be enough to take the grade to the lowest level of Very Low. Observational studies could, however, be upgraded if there were all of: a large magnitude of effect, a dose-response gradient, and if all plausible confounding would reduce the demonstrated effect.

Table 4: Overall quality of outcome evidence in GRADE

Level	Description
High	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low	Any estimate of effect is very uncertain

4.3.4.2 Prognostic reviews

The quality of evidence for prognostic studies was evaluated according to the criteria given in Table 5. If data were meta-analysed, the quality for pooled studies was presented. If the data were not pooled, then a quality rating was presented for each study.

Table 5: Description of quality elements for prospective studies

Quality element	Description of cases where the quality measure would be downgraded
Study design	Case–control studies rather than prospective cohort studies
Patient recruitment	If potential for selection bias
Validity of risk factor measure(s)	If non-validated and no reasonable face validity
Validity of outcome measure	If non-validated and no reasonable face validity
Blinding	If assessors of outcome not blinded to risk factor measurement (or vice versa)
Adequate duration of follow-up (or retrospective duration)	If follow-up (or retrospective) period inadequate to allow events to occur, or retrospective period so short that causality is in doubt because the outcome may have preceded the risk factor
Confounder consideration	If there is a lack of consideration of all reasonable confounders in a multivariable analysis
Attrition	If attrition is too high and there is no attempt to adjust for this.
Directness	If the population, risk factors or outcome differ from that in the review question

4.3.4.2.1 Inconsistency

Inconsistency was assessed as for intervention studies.

4.3.4.2.2 Imprecision

In meta-analysed outcomes, or for non-pooled outcomes, the position of the 95% CIs in relation to the null line determined the existence of imprecision. If the 95% CI did not cross the null line then no serious imprecision was recorded. If the 95% CI crossed the null line then serious imprecision was recorded.

4.3.4.2.3 Overall grading

Quality rating started at High for prospective studies, and each major limitation brought the rating down by 1 increment to a minimum grade of Very Low, as explained for interventional reviews. For prognostic reviews prospective cohort studies with a multivariate analysis are regarded as the gold standard because RCTs are usually inappropriate for these types of review for ethical or pragmatic reasons. Furthermore, if the study is looking at more than 1 risk factor of interest then randomisation would be inappropriate as it can only be applied to 1 of the risk factors.

4.3.4.3 Diagnostic studies

Risk of bias and indirectness of evidence for diagnostic data were evaluated by study using: the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) checklists (see Appendix H in the NICE guidelines manual 2014[143]). Risk of bias and applicability in primary diagnostic accuracy studies in QUADAS-2 consists of 4 domains (see Table 6):

- patient selection
- index test
- reference standard

- flow and timing.

Table 6: Summary of QUADAS-2 with list of signalling, risk of bias and applicability questions.

Domain	Patient selection	Index test	Reference standard	Flow and timing
Description	Describe methods of patient selection. Describe included patients (prior testing, presentation, intended use of index test and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2×2 table (refer to flow diagram). Describe the time interval and any interventions between index test(s) and reference standard
Signalling questions (yes/no/ unclear)	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case–control design avoided?	If a threshold was used, was it pre-specified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all patients receive a reference standard?
	Did the study avoid inappropriate exclusions?			Did all patients receive the same reference standard?
				Were all patients included in the analysis?
Risk of bias; (high/low/ unclear)	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability (high/low/ unclear)	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

4.3.4.3.1 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. Inconsistency was assessed by inspection of the sensitivity (based on the primary measure) using the point estimates and 95% CIs of the individual studies on the forest plots. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and a 95% threshold set by the GDG (the threshold above which it would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example, 0–50%, 50–95% and 95–100%).

4.3.4.3.2 Imprecision

The judgement of precision was based on visual inspection of the confidence region around the summary sensitivity and specificity point from the diagnostic meta-analysis, if conducted. Where a diagnostic meta-analysis was not performed, imprecision was assessed according to the sensitivity confidence region of the largest study. Imprecision was assessed on the sensitivity confidence region as the primary measure for decision-making. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%).

4.3.4.3.3 Overall grading

Quality rating started at High for prospective and retrospective cross-sectional studies, and each major limitation (risk of bias, indirectness, inconsistency and imprecision) brought the rating down by 1 increment to a minimum grade of Very Low, as explained for intervention reviews.

4.3.5 Assessing clinical importance

The GDG assessed the evidence by outcome in order to determine if there was, or potentially was, a clinically important benefit, a clinically important harm or no clinically important difference between interventions. To facilitate this, binary outcomes were converted into absolute risk differences (ARDs) using GRADEpro[91] software: the median control group risk across studies was used to calculate the ARD and its 95% CI from the pooled risk ratio.

The assessment of clinical benefit, harm, or no benefit or harm was based on the point estimate of absolute effect for intervention studies, which was standardised across the reviews. The GDG considered for most of the outcomes in the intervention reviews that if at least 100 participants per 1000 (10%) achieved the outcome of interest (for a positive outcome) in the intervention group compared to the comparison group then this intervention would be considered beneficial. The same point estimate but in the opposite direction applied if the outcome was negative. For adverse events 50 events or more per 1000 (5%) represented clinical harm. For continuous outcomes if the mean difference was greater than the minimally important difference (MID) then this resented a clinical benefit or harm. For critical outcomes such as mortality any reduction or increase was considered to be clinically important.

This assessment was carried out by the GDG for each critical outcome, and an evidence summary table was produced to compile the GDG's assessments of clinical importance per outcome, alongside the evidence quality and the uncertainty in the effect estimate (imprecision).

4.3.6 Clinical evidence statements

Clinical evidence statements are summary statements that are included in each review chapter, and which summarise the key features of the clinical effectiveness evidence presented. The wording of the evidence statements reflects the certainty or uncertainty in the estimate of effect. The evidence statements are presented by outcome and encompass the following key features of the evidence:

- The number of studies and the number of participants for a particular outcome.
- An indication of the direction of clinical importance (if one treatment is beneficial or harmful compared to the other, or whether there is no difference between the 2 tested treatments).
- A description of the overall quality of the evidence (GRADE overall quality).

4.4 Identifying and analysing evidence of cost-effectiveness

The GDG is required to make decisions based on the best available evidence of both clinical effectiveness and cost-effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is, their 'cost-effectiveness') rather than the total implementation cost.[143] Thus, if the evidence suggests that a strategy provides significant health benefits at an acceptable cost per patient treated, it should be recommended even if it would be expensive to implement across the whole population.

Health economic evidence was sought relating to the key clinical issues being addressed in the guideline. Health economists:

- Undertook a systematic review of the published economic literature.
- Undertook new cost-effectiveness analysis in priority areas.

4.4.1 Literature review

The health economists:

- Identified potentially relevant studies for each review question from the health economic search results by reviewing titles and abstracts. Full papers were then obtained.
- Reviewed full papers against prespecified inclusion and exclusion criteria to identify relevant studies (see below for details).
- Critically appraised relevant studies using economic evaluations checklist as specified in the NICE guidelines manual.[143,145]
- Extracted key information about the studies' methods and results into economic evidence tables (included in Appendix I).
- Generated summaries of the evidence in NICE economic evidence profile tables (included in the relevant chapter for each review question) – see below for details.

4.4.1.1 Inclusion and exclusion criteria

Full economic evaluations (studies comparing costs and health consequences of alternative courses of action: cost-utility, cost-effectiveness, cost-benefit and cost-consequences analyses) and comparative costing studies that addressed the review question in the relevant population were considered potentially includable as economic evidence.

Studies that only reported cost per hospital (not per patient), or only reported average cost-effectiveness without disaggregated costs and effects, were excluded. Literature reviews, abstracts, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded. Studies published before 1999 and studies from non-OECD countries or the USA were also excluded, on the basis that the applicability of such studies to the present UK NHS context is likely to be too low for them to be helpful for decision-making.

Remaining health economic studies were prioritised for inclusion based on their relative applicability to the development of this guideline and the study limitations. For example, if a High quality, directly applicable UK analysis was available, then other less relevant studies may not have been included. However, in this guideline, no economic studies were excluded on the basis that more applicable evidence was available.

For more details about the assessment of applicability and methodological quality see Table 7 below and the economic evaluation checklist (Appendix G of the 2012 NICE guidelines manual[145]) and the health economics review protocol in Appendix D.

When no relevant health economic studies were found from the economic literature review, relevant UK NHS unit costs related to the compared interventions were presented to the GDG to inform the possible economic implications of the recommendations.

4.4.1.2 NICE economic evidence profiles

NICE economic evidence profile tables were used to summarise cost and cost-effectiveness estimates for the included health economic studies in each review chapter. The economic evidence profile shows an assessment of applicability and methodological quality for each economic study, with footnotes indicating the reasons for the assessment. These assessments were made by the health economist using the economic evaluation checklist from the NICE guidelines manual.[145] It also shows the incremental costs, incremental effects (for example, quality-adjusted life years [QALYs]) and incremental cost-effectiveness ratio (ICER) for the base case analysis in the study, as well as information about the assessment of uncertainty in the analysis. See Table 7 for more details.

When a non-UK study was included in the profile, the results were converted into pounds sterling using the appropriate purchasing power parity.[151]

Table 7: Content of NICE economic evidence profile

Item	Description
Study	Surname of first author, date of study publication and country perspective with a reference to full information on the study.
Applicability	An assessment of applicability of the study to this guideline, the current NHS situation and NICE decision-making:[(a)] • Directly applicable – the study meets all applicability criteria, or fails to meet 1 or more applicability criteria but this is unlikely to change the conclusions about cost-effectiveness. • Partially applicable – the study fails to meet 1 or more applicability criteria, and this could change the conclusions about cost-effectiveness. • Not applicable – the study fails to meet 1 or more of the applicability criteria, and this is likely to change the conclusions about cost-effectiveness. Such studies would usually be excluded from the review.
Limitations	An assessment of methodological quality of the study:[(a)] • Minor limitations – the study meets all quality criteria, or fails to meet 1 or more quality criteria, but this is unlikely to change the conclusions about cost-effectiveness. • Potentially serious limitations – the study fails to meet 1 or more quality criteria, and this could change the conclusions about cost-effectiveness. • Very serious limitations – the study fails to meet 1 or more quality criteria, and this is highly likely to change the conclusions about cost-effectiveness. Such studies would usually be excluded from the review.
Other comments	Information about the design of the study and particular issues that should be considered when interpreting it.
Incremental cost	The mean cost associated with one strategy minus the mean cost of a comparator strategy.
Incremental effects	The mean QALYs (or other selected measure of health outcome) associated with one strategy minus the mean QALYs of a comparator strategy.
Cost-effectiveness	Incremental cost-effectiveness ratio (ICER): the incremental cost divided by the incremental effects (usually in £ per QALY gained).
Uncertainty	A summary of the extent of uncertainty about the ICER reflecting the results of deterministic or probabilistic sensitivity analyses, or stochastic analyses of trial data, as appropriate.

(a) Applicability and limitations were assessed using the economic evaluation checklist in Appendix G of the 2012 NICE guidelines manual[145]

4.4.2 Undertaking new health economic analysis

As well as reviewing the published health economic literature for each review question, as described above, new health economic analysis was undertaken by the health economist in selected areas. Priority areas for new analysis were agreed by the GDG after formation of the review questions and consideration of the existing health economic evidence.

The GDG identified the highest priority areas for original health economic modelling as:

- risk factors for cirrhosis
- the appropriate tests (blood tests, non-invasive tests or a combination) for diagnosing cirrhosis
- frequency of surveillance testing for the early detection of hepatocellular carcinoma
- frequency of surveillance testing for the detection of oesophageal varices.

This was due to the number of people affected by these questions and the current uncertainty as to what the most cost-effective solutions would be, due to the lack of published economic models encompassing the whole pathway of cirrhosis from diagnosis to end-stage liver disease. New work was therefore conducted, which entailed the development of the NGC Liver Disease Pathway Model to address all of the questions prioritised for this guideline.

The following general principles were adhered to in developing the cost-effectiveness analysis:

- Methods were consistent with the NICE reference case for interventions with health outcomes in NHS settings.[143,146]
- The GDG was involved in the design of the model, selection of inputs and interpretation of the results.
- Model inputs were based on the systematic review of the clinical literature supplemented with other published data sources where possible.
- When published data were not available GDG expert opinion was used to populate the model.
- Model inputs and assumptions were reported fully and transparently.
- The results were subject to sensitivity analysis and limitations were discussed.
- The model was peer-reviewed by another health economist at the NGC.

Full methods for the cost-effectiveness analysis are described in Appendix N.

4.4.3 Cost-effectiveness criteria

NICE's report 'Social value judgements: principles for the development of NICE guidance' sets out the principles that GDGs should consider when judging whether an intervention offers good value for money.[144] In general, an intervention was considered to be cost-effective (given that the estimate was considered plausible) if either of the following criteria applied:

- the intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
- the intervention cost less than £20,000 per QALY gained compared with the next best strategy.

If the GDG recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000 per QALY gained, the reasons for this decision are discussed explicitly in the 'Recommendations and link to evidence' section of the relevant chapter, with reference to issues regarding the plausibility of the estimate or

to the factors set out in 'Social value judgements: principles for the development of NICE guidance'.[144]

When QALYs or life years gained are not used in the analysis, results are difficult to interpret unless one strategy dominates the others with respect to every relevant health outcome and cost.

4.4.4 In the absence of economic evidence

When no relevant published health economic studies were found, and a new analysis was not prioritised, the GDG made a qualitative judgement about cost-effectiveness by considering expected differences in resource use between options and relevant UK NHS unit costs, alongside the results of the review of clinical effectiveness evidence.

The UK NHS costs reported in the guideline are those that were presented to the GDG and were correct at the time recommendations were drafted. They may have changed subsequently before the time of publication. However, we have no reason to believe they have changed substantially.

4.5 Developing recommendations

Over the course of the guideline development process, the GDG was presented with:

- Evidence tables of the clinical and economic evidence reviewed from the literature. All evidence tables are in Appendices H and I.
- Summaries of clinical and economic evidence and quality (as presented in Chapters 5–15).
- Forest plots (Appendix K).
- A description of the methods and results of the cost-effectiveness analysis undertaken for the guideline (Appendix N).

Recommendations were drafted on the basis of the GDG's interpretation of the available evidence, taking into account the balance of benefits, harms and costs between different courses of action. This was either done formally in an economic model, or informally. Firstly, the net clinical benefit over harm (clinical effectiveness) was considered, focusing on the critical outcomes. When this was done informally, the GDG took into account the clinical benefits and harms when one intervention was compared with another. The assessment of net clinical benefit was moderated by the importance placed on the outcomes (the GDG's values and preferences), and the confidence the GDG had in the evidence (evidence quality). Secondly, the GDG assessed whether the net clinical benefit justified any differences in costs between the alternative interventions.

When clinical and economic evidence was of poor quality, conflicting or absent, the GDG drafted recommendations based on its expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs compared to the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The consensus recommendations were agreed through discussions in the GDG meeting. The GDG also considered whether the uncertainty was sufficient to justify delaying making a recommendation to await further research, taking into account the potential harm of failing to make a clear recommendation (see Section 4.5.1 below).

The GDG considered the appropriate 'strength' of each recommendation. This takes into account the quality of the evidence but is conceptually different. Some recommendations are 'strong' in that the GDG believes that the vast majority of healthcare and other professionals and patients would choose a particular intervention if they considered the evidence in the same way that the GDG has. This is generally the case if the benefits clearly outweigh the harms for most people and the intervention is likely to be cost-effective. However, there is often a closer balance between benefits and harms, and some patients would not choose an intervention whereas others would. This may happen, for

example, if some patients are particularly averse to some side effect and others are not. In these circumstances the recommendation is generally weaker, although it may be possible to make stronger recommendations about specific groups of patients.

The GDG focused on the following factors in agreeing the wording of the recommendations:

- The actions health professionals need to take.
- The information readers need to know.
- The strength of the recommendation (for example the word 'offer' was used for strong recommendations and 'consider' for weak recommendations).
- The involvement of patients (and their carers if needed) in decisions on treatment and care.
- Consistency with NICE's standard advice on recommendations about drugs, waiting times and ineffective interventions (see Section 9.2 in the 2014 NICE guidelines manual[143]).

The main considerations specific to each recommendation are outlined in the 'Recommendations and link to evidence' sections within each chapter.

4.5.1 Research recommendations

When areas were identified for which good evidence was lacking, the GDG considered making recommendations for future research. Decisions about inclusion were based on factors such as:

- the importance to patients or the population
- national priorities
- potential impact on the NHS and future NICE guidance
- ethical and technical feasibility.

4.5.2 Validation process

This guidance is subject to a 6-week public consultation and feedback as part of the quality assurance and peer review of the document. All comments received from registered stakeholders are responded to in turn and posted on the NICE website

4.5.3 Updating the guideline

Following publication, and in accordance with the NICE guidelines manual, NICE will undertake a review of whether the evidence base has progressed significantly to alter the guideline recommendations and warrant an update.

4.5.4 Disclaimer

Healthcare providers need to use clinical judgement, knowledge and expertise when deciding whether it is appropriate to apply guidelines. The recommendations cited here are a guide and may not be appropriate for use in all situations. The decision to adopt any of the recommendations cited here must be made by practitioners in light of individual patient circumstances, the wishes of the patient, clinical expertise and resources.

The National Guideline Centre disclaims any responsibility for damages arising out of the use or non-use of this guideline and the literature used in support of this guideline.

4.5.5 Funding

The National Guideline Centre was commissioned by the National Institute for Health and Care Excellence to undertake the work on this guideline.

5 Risk factors and risk assessment tools

5.1 Introduction

In the UK, the most common causes of cirrhosis are drinking alcohol at harmful levels, non-alcohol-related steatohepatitis due to diabetes mellitus or metabolic syndrome, and chronic infection with either hepatitis B virus or hepatitis C virus.[148] Less common causes include autoimmune liver diseases (autoimmune hepatitis, primary biliary cholangitis or primary sclerosing cholangitis), genetic conditions (haemochromatosis, Wilson's disease, alpha-1 antitrypsin deficiency), prolonged exposure to certain chemicals or medications (amiodarone or methotrexate), Budd-Chiari syndrome or veno-occlusive disease, sarcoidosis and glycogen storage disease.[148]

People with cirrhosis may present to their primary care physician with non-specific signs and symptoms of the liver disease such as fatigue, loss of appetite or itchy skin, or with the features suggestive of liver failure such as, jaundice and fluid retention manifesting as ankle swelling or abdominal distension. However, many people with cirrhosis remain asymptomatic and hence go unrecognised until their liver begins to fail. The GDG thought that it might be helpful to identify people at risk of having cirrhosis before they developed evidence of liver decompensation and to determine whether there are any validated risk tools that identify these populations.

5.2 Review question 1: What are the risk factors that indicate the populations at specific risk for cirrhosis?

Table 8: Characteristics of review question 1

Population	People who are 16 years or older who do not have cirrhosis at baseline
Prognostic variables under consideration	• Obesity (BMI ≥30, or a lower BMI for people of Asian family origin) • Alcohol misuse • Viral hepatitis B • Viral hepatitis C • Type 2 diabetes
Confounding factors	• Obesity (BMI ≥30, BMI >25 for people of an Asian family origin): age, ethnicity, treatments for obesity (weight loss or surgery), all of the other risk factors • Alcohol misuse: gender, age, ethnicity, level and pattern of alcohol misuse, all of the other risk factors • Viral hepatitis B: gender, age, ethnicity, treatment for hepatitis B, all of the other risk factors • Viral hepatitis C: gender, age, ethnicity, treatment for hepatitis C, all of the other risk factors • Type 2 diabetes: gender, age, ethnicity, treatment for type 2 diabetes, all of the other risk factors.
Outcomes	**Critical outcomes:** Diagnosis of cirrhosis: time-to-event. If time-to-event data are not available, categorical data will be used (that is, the relative risk of developing cirrhosis at different time points).
Study design	• Prospective and retrospective cohort • Systematic reviews of the above.

5.3 Review question 2: Are there any validated risk tools that indicate the populations at specific risk for cirrhosis?

Table 9: Characteristics of review question 2

Population	People who are 16 years or older who do not have cirrhosis at baseline
Risks stratification tools	Any validated risk factor tools
Target condition	Development of cirrhosis (confirmed on liver biopsy)
Outcomes (in terms of discrimination and calibration)	**Critical outcomes:** • ROC area under the curve (of each risk tool for each outcome)/concordance c-statistic • Sensitivity, specificity, predictive values • Predicted risk, observed risk/calibration plot (reproduced with author permissions) (that is, predicted x-year mean risk % verses Kaplan-Meier x-year event rate). Narrative of agreement between observed and predicted risk and whether underestimation/overestimation of predicted risk) • Other outcomes: D statistics, R^2 statistic and Brier score.
Study design	Cohort (preferably prospective)

For full details see review protocol in Appendix C.

5.4 Clinical evidence

For question 1, we searched for prospective and retrospective cohort studies investigating the association of the following factors*:* alcohol intake, BMI, diabetes, hepatitis C and hepatitis B with future development of cirrhosis. Eight studies were included in the review which reported the relative risk of cirrhosis in people with the risk factor compared to people without the risk factor[13,15,21,79,103,113,121,197]. Evidence from these are summarised in the clinical GRADE evidence profile below. See also the study selection flow chart in Appendix E, forest plots in Appendix K, study evidence tables in Appendix H and exclusion list in Appendix L.

Six studies reported the association of alcohol consumption with the risk of cirrhosis[13,15,21,79,113,197]. Four studies reported the association of BMI with the risk of cirrhosis[15,103,121,197]. One study reported the association of diabetes status with the risk of cirrhosis in people with a BMI of 22.5 to <25[121]. All studies conducted a multivariable analysis and reported an adjusted risk which accounted for the influence of confounding factors, but different variables were included in the analysis between the different studies (see Table 10). Some studies did not account for confounding factors which are specific to cirrhosis (such as hepatitis status) as they were not initially designed to investigate cirrhosis as an outcome. These limitations were taken into account when assessing the risk of bias. Studies reporting relative risks that were not adjusted for any key confounders were not included in the analysis. Studies were not pooled in the analysis because different thresholds were used for analysis of the same outcome in different primary studies.

No studies were identified that reported the association of the following risk factors with the risk of cirrhosis in comparison to people without these risk factors: hepatitis C and hepatitis B.

The limitations of the included studies were that some studies reported the risk of cirrhosis taken from death certificates, with death from cirrhosis as the underlying cause. This might be confounded by the fact that people may reduce their alcohol consumption following a diagnosis, and the risk of developing cirrhosis may differ from the risk of death from cirrhosis. For alcohol consumption as a risk factor, some studies looked at alcohol consumption at baseline and then assessed the risk over

many years in which no information was collected. Collecting alcohol intake data only once ignores the fact that people change their drinking habits over time, so longitudinal studies that assess alcohol consumption only at baseline may result in misclassification. In the baseline questionnaire, some studies did not question subjects on their past drinking habits, meaning that ex-drinkers who were currently abstaining would have been included in the non-drinking group, therefore biasing the results. These limitations were taken into account when assessing the risk of bias. For question 2, we searched for any validated risk tools that incorporate the risk populations specified in question 1 to indicate the populations at specific risk for cirrhosis (see Appendix **C**). No validated risk tools encompassing any of the 5 risk factors (obesity, alcohol misuse, viral hepatitis B, viral hepatitis C, type 2 diabetes) were identified from the literature.

Table 10: Summary of studies included in the review

Study	Population	Analysis	Prognostic variable(s)	Confounders	Outcomes	Limitations
Askgaard 2015[13]	Data taken from a Danish prospective cohort study originally designed to investigate associations between diet and other lifestyle exposures and cancer in middle-aged individuals. From December 1993 to May 1997, women and men aged 50 to 64 years, born in Denmark and not previously diagnosed with cancer, were invited to participate in the Diet, Cancer and Health study. Total n=27,178, number of events (diagnosis of alcohol-related cirrhosis)=342	Multivariate	1. Alcohol consumption frequency (drinking days per week: lifetime abstainers, current abstainers, <1, 1, 2–4, 5–6, 7) **At start of study**	Sex, age, smoking, waist circumference, length of education	Diagnosis of alcohol-related cirrhosis **Follow-up at about 15 years**	Alcohol intake was only assessed at baseline. Participants were asked to retrospectively recall how many units they drank since age 20 but this data had not been incorporated into the hazard ratio analysis: that is, it is unknown how many units participants drank on their 'drinking days'. Confounding factors of hepatitis B and C status and diabetes not taken into account. No previous alcohol consumption data was collected for current abstainers who may have been heavy drinkers previously and therefore of uncertain risk if they returned to drinking during the follow-up period.
Becker 2002[15]	Subjects from several cohort studies: Copenhagen County Centre of Preventative Medicine birth cohorts, World Health Organisation Monitoring of Trends and Determinants in Cardiovascular Diseases (MONICA) I,	Multivariate	1. Alcohol intake (drinks/week: <1, 1–7, 8–21, 22–35, >35). 2. BMI (kg/m^2: <20,	Age, smoking habits, number of years in school education, percentage wine of total alcohol	Death or hospital discharge with alcohol-related cirrhosis **Average follow-**	Some of the included cohorts only assessed alcohol intake on 1 occasion. Confounding factors of hepatitis B and C status and diabetes not

Study	Population	Analysis	Prognostic variable(s)	Confounders	Outcomes	Limitations
	MONICA II and MONICA III the Copenhagen City Heart Study and the Copenhagen Male Study. Total n=30,630, number of events =292.		20–24, >24–30, >30). **At start of study**	intake	**up not reported**	taken into account. Non-drinking group could have included individuals who had previously suffered from illness due to alcohol or current abstainers who were previously heavy drinkers and therefore of uncertain risk if they returned to drinking during the follow-up period.
Blackwelder 1980[21]	Honolulu Heart Study, a prospective study among men of Japanese descent in Hawaii, born between 1900 and 1919 and residing on the island of Oahu in 1965	Multivariate	1. Alcohol consumption (ml per day of ethanol) **At start of study**	Age, cigarettes smoked per day, systolic blood pressure, serum cholesterol, relative weight	Death due to cirrhosis **Follow-up 8 years**	A proportion of subjects were diabetic at baseline and this was not adjusted for. The hepatitis status of subjects was unknown.
Fuchs 1995[79]	Prospective cohort. Nurses' Health Study Women only	Multivariate	1. Alcohol consumption (g/day) **At start of study –also performed sensitivity analyses using updated measurements of alcohol consumption obtained at 4 and 6 years (only reported for association between alcohol intake and all-cause mortality, not**	Age, smoking status, BMI, regular aspirin use, regular vigorous exercise, high plasma cholesterol level	Death due to cirrhosis **Follow-up 12 years**	Confounding factors of hepatitis status and ethnicity not accounted for

Study	Population	Analysis	Prognostic variable(s)	Confounders	Outcomes	Limitations
			mortality due to cirrhosis)			
Ioannou 2003[103]	Prospective cohort. Participants aged 25–74 years in the United States. National Health and Nutrition Examination Survey (NHANES I)	Multivariate	1. BMI (normal <25, overweight 25- <30, obese ≥30) **At start of study**	Age, alcohol consumption, sex, race, education, household income, geographic location in the United States. Adjusted HR reported in this review also adjusted for presence of diabetes.	Hospitalisation or death due to cirrhosis **Mean follow-up 12.6 years**	Did not account for confounding factors of hepatitis B or C status
Klatsky 1992[113]	Prospective cohort of people who underwent health examinations at the Oakland and San Francisco facilities of the Kaiser Permanente Medical Care Program	Multivariate	1. Alcohol use (categorical: never-drinkers, ex-drinkers, and five categories of drinkers up to 6 drinks per day or more) **At start of study**	Age, gender, race, education, BMI, marital status, upper gastrointestinal history, smoking, tea and coffee consumption	Hospitalisation or death due to cirrhosis **Recruitment and data collection period from 1878 to 1985. Last follow-up 1988.**	Did not account for confounding factors of hepatitis B or C status, diabetes or gender. Alcohol intake only assessed on 1 occasion.

Schult 2011[197]	Longitudinal cohort study conducted in Gothenburg during a 40 year study period. Total n=792. Number of events =14.	Multivariate	1. Alcohol abuse using 2 definitions **Data obtained in 1967** 2. BMI (elevated BMI versus normal). Note: elevated BMI presumed >30 but unclear from paper. **At start of study**	BMI, triglycerides, 2 definitions of alcohol abuse	Patients who were hospitalised and/or died with a diagnosis of cirrhosis **Data collection period 40 years 1963–2003**	Hepatitis status of subjects unknown. Ethnicity not adjusted for. One of the methods used to assess alcohol intake was based on records detailing individuals who had sought help for alcohol addiction, been convicted for drunkenness or provided with institutional care by the authorities. The second measure of alcohol intake was based on a one-off questionnaire.

5.4.1 Prognostic factor: alcohol consumption

Six studies reported the association of alcohol consumption with the risk of cirrhosis[13,15,21,79,113,197]. The level of alcohol consumption was categorised differently between the studies. In addition, the studies adjusted for different confounding factors. Therefore, the results were not pooled in the analysis.

Table 11: Prognostic factor: alcohol consumption

Risk factors/outcomes	Number of studies	Number of events/people (%) with and without risk factor (if available)	Effect with 95% CIs	Imprecision	GRADE
Men <1 drink/week versus 1–7 drinks/week in men (reference) for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])	1	Not reported for men and women separately	HR=7.76 (3.35–18.0)	None[b]	Low
Women <1 drink/week versus 1–7 drinks/week in men (reference) for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])	1	Not reported for men and women separately	HR=1.32 (0.51–3.42)	Serious[b]	Very low
Women 1–7 drinks/week versus 1–7 drinks/week in men (reference) for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])	1	Not reported for men and women separately	HR=1.19 (0.54–2.62)	Serious[b]	Very low
Men 8–21 drinks/week versus 1–7 drinks/week in men (reference) for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])	1	Not reported for men and women separately	HR=2.34 (1.18–4.64)	None[b]	Low
Women 8–21 drinks/week versus 1–7 drinks/week	1	Not reported for men and	HR=5.33 (2.63–	None[b]	Low

Risk factors/outcomes	Number of studies	Number of events/people (%) with and without risk factor (if available)	Effect with 95% CIs	Imprecision	GRADE
in men (reference) for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])		women separately	10.8)		
Men 22–35 drinks/week versus 1–7 drinks/week in men (reference) for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])	1	Not reported for men and women separately	HR=10.4 (5.4–20.03)	None[b]	Low
Women 22–35 drinks/week versus 1–7 drinks/week in men (reference) for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])	1	Not reported for men and women separately	HR=10.8 (4.28–27.25)	None[b]	Low
Men >35 drinks/week versus 1–7 drinks/week in men (reference) for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])	1	Not reported for men and women separately	HR=20.4 (10.8–38.53)	None[b]	Low
Women >35 drinks/week versus 1–7 drinks/week in men (reference) for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])	1	Not reported for men and women separately	HR=14.1 (4.45–44.6)	None[b]	Low
'Model 1' (alcohol abuse definition 1) versus non-abusers for predicting death or hospitalisation with cirrhosis (adjusted ORs[c])	1	Not reported	OR=0.71 (0.17–2.97)	Serious[b]	Very low
'Model 2' (alcohol abuse definition 2) versus non-abusers for predicting death or hospitalisation with cirrhosis (adjusted ORs[c])	1	Not reported	OR=1.55 (0.36–6.78)	Serious[b]	Very low
0.1–1.4 g/day versus 0 g/day for predicting death due to cirrhosis (adjusted HRs[d])	1	With: 1/11,304 (0.009%) Without: 12/25,535 (0.05%)	HR=0.21 (0.03–1.59)	Serious[b]	Low
1.5–4.9 g/day versus 0 g/day for predicting death due to cirrhosis (adjusted HRs[d])	1	With: 5/18,406 (0.03%) Without: 12/25,535 (0.05%)	HR=0.69 (0.24–1.98)	Serious[b]	Low
5.0–14.9 g/day versus 0 g/day for predicting death due to cirrhosis (adjusted HRs[d])	1	With: 10/17,783 (0.06%) Without: 12/25,535 (0.05%)	HR=1.27 (0.54–3.01)	Serious[b]	Low
15.0–29.9 g/day versus 0 g/day for predicting death	1	With: 9/8,106 (0.11%)	HR=1.86 (0.76–	Serious[b]	Low

Association of alcohol intake with death from cirrhosis (standardised coefficient from multivariate analysis)[e]	1	Not reported	Standardised coefficient from multivariate analysis =0.341 (t=3.11, estimated coefficient divided by its standard error, p<0.01)	None[b]	Low
Risk factors/outcomes	**Number of studies**	**Number of events/people (%) with and without risk factor (if available)**	**Effect with 95% CIs**	**Imprecision**	**GRADE**
due to cirrhosis (adjusted HRs[d])		Without: 12/25,535 (0.05%)	4.59)		
≥30 g/day versus 0 g/day for predicting death due to cirrhosis (adjusted HRs[d])	1	With: 15/4,521 (0.33%) Without: 12/25,535 (0.05%)	HR=2.55 (1.06–6.11)	None[b]	Moderate
Men current abstainers versus 2–4 drinking days/week in men (reference) for predicting diagnosis of alcohol-related cirrhosis (adjusted HRs[f])	1	With: 7/350 (2.00%) Without: 27/9165 (0.29%)	HR=10.0 (4.32–23.0)	None[b]	Low
Women current abstainers versus 2–4 drinking days/week in women (reference) for predicting diagnosis of alcohol-related cirrhosis (adjusted HRs[f])	1	With: 2/370 (0.54%) Without: 15/9481 (0.16%)	HR=4.03 (0.91–17.8)	Serious[b]	Very Low
Men <1 drinking days/week versus 2–4 drinking days/week in men (reference) for predicting diagnosis of alcohol-related cirrhosis (adjusted HRs[f])	1	With: 14/2946 0.48%) Without: 27/9165 (0.29%)	HR=1.34 (0.67–2.67)	Serious[b]	Very Low
Women <1 drinking days/week versus 2–4 drinking days/week in women (reference) for predicting diagnosis of alcohol-related cirrhosis (adjusted HRs[f])	1	With: 16/7682 (0.21%) Without: 15/9481 (0.16%)	HR=1.45 (0.71–2.96)	Serious[b]	Very Low
Men 1 drinking day/week versus 2–4 drinking days/week in men (reference) for predicting diagnosis of alcohol-related cirrhosis (adjusted HRs[f])	1	With: 8/2401 (0.33%) Without: 27/9165 (0.29%)	HR=1.30 (0.59–2.87)	Serious[b]	Very Low
Women 1 drinking day/week versus 2–4 drinking	1	With: 5/4345 (0.12%)	HR=0.81 (0.29–	Serious[b]	Very Low

days/week in women (reference) for predicting diagnosis of alcohol-related cirrhosis (adjusted HRs[f])		Without: 15/9481 (0.16%)	2.24)		
Men 5–6 drinking days/week versus 2–4 drinking days/week in men (reference) for predicting diagnosis of alcohol-related cirrhosis (adjusted HRs[f])	1	With: 30/4495 (0.67%) Without: 27/9165 (0.29%)	HR=1.43 (0.84–2.43)	Serious[b]	Very Low
Women 5–6 drinking days/week versus 2–4 drinking days/week in women (reference) for predicting diagnosis of alcohol-related cirrhosis (adjusted HRs[f])	1	With: 17/3147 (0.54%) Without: 15/9481 (0.16%)	HR=2.30 (1.14–4.67)	None[b]	Low
Men 7 drinking days/week versus 2–4 drinking days/week in men (reference) for predicting diagnosis of alcohol-related cirrhosis (adjusted HRs[f])	1	With: 171/7276 (2.35%) Without: 27/9165 (0.29%)	HR=3.65 (2.39–5.55)	None[b]	Low
Women 7 drinking days/week versus 2–4 drinking days/week in women (reference) for predicting diagnosis of alcohol-related cirrhosis (adjusted HRs[f])	1	With: 30/3931 (0.76%) Without: 15/9481 (0.16%)	HR=1.73 (0.85–3.52)	Serious[b]	Very Low

[a] *Methods multivariable analysis, key covariates included: age, smoking habits, number of years in school education, percentage wine of total alcohol intake*
[b] *If the confidence intervals did not cross the null line then no serious imprecision was recorded. If the confidence intervals crossed the null line then serious imprecision was recorded.*
[c] *Methods multivariable analysis, key covariates included: BMI, triglycerides, 2 definitions of alcohol abuse*
[d] *Methods multivariable analysis, key covariates included: Age, smoking status, BMI, regular aspirin use, regular vigorous exercise, high plasma cholesterol level*
[e] *Methods multivariable analysis, key covariates included: Age, cigarettes smoked per day, systolic blood pressure, serum cholesterol, relative weight*
[f] *Methods multivariable analysis, key covariates included: age, smoking, education, and waist circumference.*

Narrative information:

Klatsky 1992[113] reported the following information, however 95% CIs were not reported therefore the quality of the evidence could not be assessed. Adjusted HRs (adjusted for age, gender, race, education, BMI, marital status, upper gastrointestinal history, smoking, tea and coffee consumption):

Hospitalisation for alcohol-related cirrhosis

Drinks/day

<1 drink/day (reference) HR 1.0

Ex-drinkers HR 5.4

1–2	HR 7.7
3–5	HR 18.2
≥6	HR 33.1

Hospitalisation for non-alcohol-related cirrhosis

Drinks/day	
Lifelong non-drinkers (reference)	HR 1.0
Ex-drinkers	HR 1.2
1–2	HR 0.8
3–5	HR (analysis not performed because of the small number of cases)
≥6	HR 0.8

Death from alcohol-related cirrhosis

Drinks/day	
<1 drink/day (reference)	HR 1.0
Ex-drinkers	HR 17.1
1–2	HR 7.8
3–5	HR 21.6
≥6	HR 83.4

Death from non-alcohol-related cirrhosis

Drinks/day	
Lifelong non-drinkers (reference)	HR 1.0
Ex-drinkers	HR 16.3
1–2	HR 7.0
3–5	HR 6.4
≥6	HR 23.6

5.4.2 Prognostic factor: BMI

Four studies reported the association of BMI with the risk of cirrhosis[15,103,122,197]. The BMI was categorised differently between the studies. In addition, the studies adjusted for different confounding factors. Therefore, the results were not pooled in the analysis.

Table 12: Prognostic factor: BMI

Risk factors/outcomes	Number of studies	Number of events/people (%) with and without risk factor (if available)	Effect with 95% CIs	Imprecision	GRADE
BMI <20 versus 20–24 for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])	1	Not reported	HR=2.2(1.3–3.9)	None[b]	Low
BMI >30 versus 20–24 for predicting death or discharge with alcohol-related cirrhosis (adjusted HRs[a])	1	Not reported	HR=2.2(1.5–3.4)	None[b]	Low
'Model 1' (alcohol abuse definition 1) elevated BMI[f] versus non-obese for predicting death or hospitalisation with cirrhosis (adjusted ORs[c])	1	Not reported	OR=1.27 (1.09–1.48)	None[b]	Low
'Model 2' (alcohol abuse definition 2) elevated BMI[f] versus non-obese for predicting death or hospitalisation with cirrhosis (adjusted ORs[c])	1	Not reported	OR=1.26 (1.08–1.47)	None[b]	Low
BMI overweight 25- <30 versus normal <25 (adjusted HRs[d])	1	35/3774 (0.93%) versus 34/5752 (0.59%)	HR=1.08 (0.6–1.9)	Serious[b]	Low
BMI obese ≥30 versus normal <25 (adjusted HRs[d])	1	20/1939 (1.03%) versus 34/5752 (0.59%)	HR=1.65 (0.9–3.1)	Serious[b]	Low
BMI <22.5 versus 22.5 to <25 for predicting death or hospitalisation with cirrhosis (adjusted HRs[e])	1	414/237,619 (0.17%) versus 402/331,480 (0.12%)	HR=1.36 (1.23–1.50)	None[b]	Low
BMI 25 to <27.5 versus 22.5 to <25 for predicting	1	343/266,795 (0.13%)	HR=1.05 (0.94–1.17)	Serious[b]	Very Low

Risk factors/outcomes	Number of studies	Number of events/people (%) with and without risk factor (if available)	Effect with 95% CIs	Imprecision	GRADE
death or hospitalisation with cirrhosis (adjusted HRs[e])		versus 402/331,480 (0.12%)			
BMI 27.5 to <30 versus 22.5 to <25 for predicting death or hospitalisation with cirrhosis (adjusted HRs[e])	1	236/173,498 (0.14%) versus 402/331,480 (0.12%)	HR=1.11 (0.97–1.26)	Serious[b]	Very Low
BMI 30 to <35 versus 22.5 to <25 for predicting death or hospitalisation with cirrhosis (adjusted HRs[e])	1	283/156,733 (0.18%) versus 402/331,480 (0.12%)	HR=1.49 (1.33–1.68)	None[b]	Low
BMI ≥35 versus 22.5 to <25 for predicting death or hospitalisation with cirrhosis (adjusted HRs[e])	1	133/64,537 (0.21%) versus 402/331,480 (0.12%)	HR=1.77 (1.49–2.10)	None[b]	Low

[a] *Methods multivariable analysis, key covariates included: age, smoking habits, number of years in school education, percentage wine of total alcohol intake*
[b] *If the confidence intervals did not cross the null line then no serious imprecision was recorded. If the confidence intervals crossed the null line then serious imprecision was recorded.*
[c] *Methods multivariable analysis, key covariates included: BMI, triglycerides, 2 definitions of alcohol abuse*
[d] *Methods multivariable analysis, key covariates included: age, alcohol consumption, sex, race, education, household income, geographic location in the United States. Adjusted HR reported in this review also adjusted for presence of diabetes.*
[e] *Methods multivariable analysis, key covariates included: age, region, socioeconomic status, alcohol consumption, smoking, physical activity*
[f] *Elevated BMI presumed to be >30 but unclear as reported in paper*

Narrative information:

In addition to the adjusted HRs reported in the table above (BMI overweight 25- <30 and BMI obese ≥30 versus, both versus the reference of normal BMI and adjusted for alcohol as one of the confounding factors) the Ioannou 2003 study[103] also reported adjusted HRs for the association of obesity and being overweight with cirrhosis-related hospitalisation or death stratified by alcohol consumption. Among people who did not consume alcohol, they reported a strong association between obesity (adjusted HR 4.10, 95% CI 1.4–11.4) or being overweight (adjusted HR 1.93, 95% CI 0.7–5.3) and cirrhosis-related hospitalisation or death. This association was weaker among persons who consumed up to 0.3 alcoholic drinks/day (adjusted HR 2.48, 95% CI 0.7–8.4 for obesity; adjusted HR 1.31, 95% CI 0.4–4.2 for being overweight), and no association was identified among those who consumed more than 0.3 alcoholic drinks/day (adjusted HR 0.80, 95% CI 0.3–2.1 for obesity; adjusted HR 0.97, 95% CI 0.5–1.8 for being overweight). The cut-off level of 0.3 alcoholic drinks/day was chosen because it was the median level of alcohol consumption among alcohol drinkers. Ioannou 2003[103] also reported that when cirrhosis

was considered as an any-listed diagnosis instead of as the principal diagnosis, being overweight and obesity were associated with an increased risk for cirrhosis among non-drinkers (adjusted HR 3.0, 95% CI 1.5–6.4 for being overweight; adjusted HR 3.2, 95% CI 1.4–7.3 for obesity), but not among drinkers.

Table 13: Association of obesity and being overweight with cirrhosis-related hospitalisation or death stratified by alcohol consumption (Ioannou 2003)

	Reported alcohol consumption (adjusted HRs[a])		
BMI category	**None**	**Up to 0.3 drinks/day**	**>0.3 drinks/day**
Normal (reference)	1.0	1.0	1.0
Overweight	1.93 (0.7–5.3)	1.31 (0.4–4.2)	0.97 (0.5–1.8)
Obese	4.10 (1.4–11.4)	2.48 (0.7–8.4)	0.80 (0.3–2.1)

[a] *Adjusted for age, alcohol consumption, geographic region, sex, race, household income, and educational attainment.*

For the results summarised in Table 12 above, Liu 2010[121] reported the association of BMI with risk of cirrhosis-related hospitalisation or death, adjusted for alcohol consumption. This study also reported the association of BMI on the relative risk of cirrhosis in categories of alcohol consumption and diabetes reported at recruitment. They reported that the trend in the relative risk with increasing BMI did not differ significantly between drinkers with increasingly larger consumptions of alcohol (<70, 70 to <150, or ≥150 g/week) or between women who had diabetes or not.

Table 14: Association of BMI with cirrhosis-related hospitalisation or death stratified by alcohol consumption and presence of diabetes (Liu 2010)

	Reported alcohol consumption (adjusted HRs[a])			Diabetes status (adjusted HRs[a])	
BMI category at recruitment	**<70 g/week**	**70 to <150 g/week**	**≥150 g/week**	**No diabetes**	**Diabetes**
22.5 to <25	1.00 (0.85 to 1.17) (reference)	1.59 (1.31 to 1.92)	3.44 (2.70 to 4.37)	1.00 (0.90 to 1.11) (reference)	4.29 (2.74 to 6.73)
25 to <30	0.96 (0.84 to 1.10)	1.83 (1.56 to 2.16)	3.82 (3.09 to 4.72)	1.05 (0.96 to 1.15)	4.37 (3.30 to 5.78)
≥30	1.35 (1.15 to 1.59)	2.31 (1.81 to 2.94)	6.53 (4.98 to 8.55)	1.38 (1.24 to 1.54)	5.94 (4.83 to 7.31)

[a] *Adjusted for age, region, socioeconomic status, physical activity, and alcohol consumption and smoking as appropriate*

5.4.3 Prognostic factor: diabetes status

No studies specifically reported the association of diabetes with the risk of cirrhosis. However, Liu 2010 reported the association of BMI with the risk of cirrhosis in people stratified by diabetes status. As the reference group (without the risk factor) comprised people with a BMI of 22.5 to <25 without diabetes, the risk of diabetes could be assessed in this group (BMI 22.5 to <25).

Table 15: Prognostic factor: diabetes

Risk factors/outcomes	Number of studies	Number of events/people (%) with and without risk factor (if available)	Effect with 95% CIs	Imprecision	GRADE
Diabetes versus no diabetes (in people with BMI 22.5 to <25) for predicting death or hospitalisation with cirrhosis (adjusted HRs[a])	1	Not reported	4.29 (2.74 to 6.73)	None[b]	Low

[a] *Adjusted for age, region, socioeconomic status, physical activity, and alcohol consumption and smoking as appropriate*
[b] *If the confidence intervals did not cross the null line then no serious imprecision was recorded. If the confidence intervals crossed the null line then serious imprecision was recorded.*

Narrative information:

Ioannou 2003[103] reported little difference in the rates of death or hospitalisation caused by cirrhosis by diabetes mellitus status.

5.5 Economic evidence

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

5.6 Evidence statements

5.6.1 Clinical

Alcohol consumption

- One prospective cohort study of 30,630 people showed that men who drink <1, 8–21, 22–35 or >35 drinks/week are at higher risk of death or a discharge diagnosis of alcohol-related cirrhosis than men who drink 1–7 drinks/week (Low quality evidence). The same study showed that women who drink <1, 8–21, 22–35 or >35 drinks/week are at higher risk of death or discharge with alcohol-related cirrhosis than men who drink 1–7 drinks/week (Very low to Low quality evidence). Only women who drank 1–7 drinks/week had a similar risk of death or discharge with alcohol-related cirrhosis to men who drank 1–7 drinks/week in men (reference). This evidence was of Very Low quality.
- Very Low quality evidence from 1 prospective cohort study of 792 people reported that individuals who had sought help for alcohol addiction, been arrested for drunkenness or had been provided with institutional care by social authorities had lower odds of death or hospitalisation with cirrhosis than people who did not abuse alcohol. Very Low quality evidence from the same study showed that people who self-reported as having alcohol problems and/or daily alcohol consumption had higher odds of death or hospitalisation with cirrhosis than people who did not abuse alcohol.
- Low quality evidence from 1 prospective cohort study of 85,709 people reported lower risk of death due to cirrhosis for people who had an average alcohol intake of 0.1–1.4 or 1.5–4.9 g/day compared to people with an alcohol intake of 0 g/day. The same study showed a higher risk of death from cirrhosis for people who had an average alcohol intake of 5.0–14.9, 15.0–29.9 or ≥30 g/day compared to people with an intake of 0 g/day (Low to Moderate quality).
- Low quality evidence from 1 prospective cohort study of 8,008 people reported an increase in the risk of death from cirrhosis with increasing alcohol consumption.
- One prospective cohort study of 55,917 people showed that men, who currently abstain from drinking alcohol, or have <1, 1, 5–6 or 7 drinking days per week are at higher risk of receiving a diagnosis of alcohol-related cirrhosis compared to men who have 2–4 drinking days per week (Low to Very Low quality evidence). The same study showed that women who currently abstain from drinking alcohol, or have <1, 5–6 or 7 drinking days per week are at higher risk of receiving a diagnosis of alcohol-related cirrhosis compared to women who have 2–4 drinking days per week (Low to Very Low quality evidence). Only women with 1 drinking day per week had a lower risk of a diagnosis of alcohol-related cirrhosis compared to women with 2–4 drinking days per week (Very Low quality evidence).

BMI

- Low quality evidence from 1 prospective cohort study of 30,630 people reported increased risk of death or discharge with alcohol-related cirrhosis for people with a BMI of <20 or >30 compared to people with a BMI of 20–24.
- Low quality evidence from 1 prospective cohort study of 792 people showed increased odds of death or hospitalisation with cirrhosis for individuals who had sought help for alcohol addiction,

been arrested for drunkenness or had been provided with institutional care by social authorities and had an elevated BMI (presumed >30), as well as individuals self-reported as having alcohol problems and/or daily alcohol consumption and had an elevated BMI compared to non-obese individuals.

- Low quality evidence from 1 prospective cohort study of 11,465 people reported no changes in risk of death or hospitalisation with cirrhosis for people with a BMI in the overweight category (25- <30) but increased risk for people with a BMI in the obese category (≥30) compared to people with a BMI in the normal category (<25).
- Very Low quality evidence from 1 prospective cohort study of 1,230,662 women reported no difference in risk of death or hospitalisation with cirrhosis for those with a BMI of 25 to <27.5 or 27.5 to <30 compared to women with a BMI of 22.5 to <25. Low quality evidence from the same study showed an increased risk of death or hospitalisation with cirrhosis for women with a BMI of <22.5, 30 to <35, or ≥35 compared to those with a BMI of 22.5 to <25.

Diabetes

- Low quality evidence from 1 prospective cohort study of 1,230,662 women reported increased risk of death or hospitalisation with cirrhosis for women with diabetes (with BMI 22.5 to <25) compared to those without diabetes.

5.6.2 Economic

- No relevant economic evaluations were identified.

5.7 Recommendations and link to evidence

Recommendation	**1. Be aware that there is an increased risk of cirrhosis in people who:** • **have hepatitis B virus infection** • **have hepatitis C virus infection** • **misuse alcohol** • **are obese (BMI of 30 kg/m^2 or higher)** • **have type 2 diabetes.** **Also see the NICE guidelines on: non-alcoholic fatty liver disease (NAFLD); alcohol-use disorders: diagnosis and management of physical complications; alcohol-use disorders: prevention; alcohol-use disorders: diagnosis, assessment and management of harmful drinking and alcohol dependence; type 2 diabetes in adults; obesity; and hepatitis B (chronic).**
Research recommendation	**1. Development of a risk tool to identify people at risk of cirrhosis**
Relative values of different outcomes	The GDG was interested in the association of the following factors: alcohol intake, BMI, diabetes, hepatitis C and hepatitis B with future development of cirrhosis. The ideal way of predicting an individual's risk of cirrhosis, based on all the individual's risk factors, is to use a validated risk tool. In the absence of such a tool, the GDG looked for evidence on the association of each individual factor with the development of future cirrhosis from prognostic studies. The GDG agreed that this would help identify associations between each risk factor and cirrhosis, and therefore populations who may be at higher risk. However, this evidence cannot predict an individual's risk of cirrhosis in order to determine who should be tested for cirrhosis.

Trade-off between clinical benefits and harms	**Risk Tools** No validated risk tools to predict an individual's future risk of cirrhosis were identified. The GDG were aware of the 'love your liver' tool from the British Liver Trust website. However, it was thought that this tool has not yet been validated and no validation studies were identified from the search. It was also thought that this tool predicts the risk of liver disease in general, and not cirrhosis specifically. The GDG discussed that a validated risk tool is the only way to predict an individual's future risk of cirrhosis based on all their risk factors. The GDG made a research recommendation for the development and validation of a risk tool to identify people at highest risk of developing cirrhosis who should be investigated. **Risk factors** Although a risk tool is the best way to identify an individual's future risk of cirrhosis and identify individuals who should undergo diagnostic testing, knowledge of the risk factors associated with cirrhosis is useful in predicting the likely effect of treatment or avoidance of risk factors, and in planning diagnostic investigations. If the probability of cirrhosis is very low ('good prognosis'), any adverse effects related to invasive diagnostic tests, even if rare, will play a big part in any decision to perform such tests. If instead the probability of cirrhosis is high ('bad prognosis'), the impact of new diagnostic information may be large and patients may be ready to accept higher risks of diagnostic investigations. **Alcohol consumption** Six studies reported association of alcohol consumption and either risk of diagnosis of alcohol-related cirrhosis, or risk of hospitalisation or death from cirrhosis. Studies were not pooled in the analysis because they reported different categorisations for the level of alcohol consumption, and also due to heterogeneity in the confounding factors adjusted for in the analyses. However, the GDG noted the general increase in risk (hazard or odds ratio) associated with an increasing amount of alcohol consumed (above 7 drinks per week, or 5 grams per day) or the number of drinking days (4 or more days per week). They noted that there was a trend towards a higher risk of cirrhosis-related hospitalisation and death in women than men, at the same consumption levels. The GDG discussed the limitations of the prognostic studies, for example only assessing the alcohol consumption at baseline and not taking into account changes in the alcohol consumption over time. This may have resulted in people who were currently abstaining from alcohol being included in the non-drinking group, thereby biasing the results. Although the increased risk of developing cirrhosis associated with an increased alcohol consumption was apparent, due to the differences between studies in how the level of alcohol consumption was measured (that is, drinks per week or number of drinking days per week) it was difficult for the GDG to define a cut-off level of alcohol consumption above which to recommend a diagnostic assessment for cirrhosis. The GDG agreed that it was not necessary for everyone drinking above recommended safe limits to be assessed for cirrhosis. This would result in a large proportion of people who will not develop cirrhosis being tested. The GDG discussed that the people most at risk of developing cirrhosis, and therefore the people who should be tested, are people who fall within the current NHS definition of 'higher-risk' drinking: that is, men who drink over 50 units of alcohol per week and women who drink over 35 units of alcohol per week. The GDG agreed that consumption at these levels is associated with an increased risk of developing cirrhosis over time. However, it is important to note that there may be a case for testing in people who have otherwise been identified as potentially having ALD even if their current alcohol consumption is lower than the selected threshold levels, for example those shown incidentally to have abnormal blood tests. The GDG discussed that special consideration should be made in people who may be

	at higher risk. For example, lower levels of alcohol consumption may lead to increased risk in people of certain ethnic origins (for example those of South Asian origin). However, there was no evidence available on the risk of cirrhosis specifically in this group. The GDG also felt that people with a combination of risk factors should be considered to be at higher risk, for example, people with increased alcohol consumption and hepatitis C. These people should be offered diagnostic assessment for cirrhosis even if their alcohol consumption did not exceed the selected threshold level. **BMI** Four studies reported the association of BMI with the risk of cirrhosis. Studies were not pooled in the analysis due to the studies reporting different categorisations of BMI, and also due to heterogeneity in the confounding factors adjusted for in the analysis. However, the GDG noted the general increase in risk (hazard or odds ratio) associated with an increase in BMI. The GDG felt that the association of BMI and cirrhosis was not as strong as expected. Being overweight (BMI 25-<30) versus a normal BMI resulted in little increase in risk, with a hazard ratio of 1.08 from one study, and hazard ratios of 1.05 and 1.11 for a BMI of 25-<27.5 and 27.5-<30 respectively from another study. Obesity versus a normal BMI resulted in a larger increase in risk of cirrhosis, with hazard ratios of 1.65 and 2.2 for a BMI >30, and a hazard ratio of 1.49 for a BMI of 30-<35. The GDG noted Table 14 which suggested the added effect of alcohol and BMI and the stratification by diabetes. The reference group for this was low BMI and low alcohol intake. **Type 2 diabetes** One study reported the association of type 2 diabetes status with the risk of cirrhosis in people with a BMI of 22.5 to <25. There was a higher risk of cirrhosis in people with type 2 diabetes, with an adjusted hazard ratio of 4.29. The GDG agreed that health professionals should be aware of the increased risk of cirrhosis in people with obesity and type 2 diabetes. The GDG noted that 'Non-alcoholic fatty liver disease (NAFLD): assessment and management'[142] has made recommendations on which people to test for hepatic fibrosis and cirrhosis and that cross-reference should be made to this guideline. **Hepatitis B and C** No evidence was identified that reported the association of hepatitis B or C in comparison to people without these risk factors. The GDG agreed that all people with hepatitis B and C are at a higher risk of cirrhosis and should be tested.
Trade-off between net clinical effects and costs	No relevant published economic evidence was identified. 'Non-alcoholic fatty liver disease (NAFLD): assessment and management'[142] has made recommendations to identify people with NAFLD and advanced fibrosis (stage F3). Since an individual must have advanced fibrosis before developing cirrhosis, this group will include all those at risk of progressing to cirrhosis, whilst minimising the number of people with NAFLD needing to be tested for cirrhosis. Since the risk of cirrhosis in people who have obesity or type 2 diabetes comes through their increased likelihood of having NAFLD, and there is no known risk of cirrhosis in people with obesity or type 2 diabetes if they do not also have NAFLD (or an alternative risk factor), then the GDG agreed it was not appropriate to test people with obesity or type 2 diabetes more broadly for cirrhosis, but only the subpopulation identified by 'Non-alcoholic fatty liver disease (NAFLD): assessment and management'. Having agreed that people with hepatitis B or hepatitis C, and those considered likely to have alcohol-related liver disease have a much higher incidence of cirrhosis than

	the general population, the GDG agreed that testing these groups would be the most efficient use of resources. These 4 population groups (people with NAFLD, ALD, hepatitis B, hepatitis C) were therefore investigated in the original economic analysis conducted for this guideline, to compare the cost-effectiveness of a variety of diagnostic tests in each of these groups. For each group the diagnostic tests were also compared against an alternative of no testing, and the results showed that testing was cost-effective compared to no testing, thus justifying the decision to test in these 4 population groups. For further results and discussion of the original economic analysis see Chapter 6 and Appendix N.
Quality of evidence	**Risk tools** No clinical evidence was identified. **Risk factors** The majority of the evidence was of Low or Very Low quality. The main reasons for downgrading the evidence were risk of bias and risk of imprecision. The aspects contributing to the risk of bias are discussed below. All studies conducted a multivariable analysis and reported an adjusted risk which accounted for the influence of confounding factors, but different variables were included in the analysis between the different studies. Some studies did not account for confounding factors which are specific to cirrhosis (such as hepatitis status) as they were not initially designed to investigate cirrhosis as an outcome. These limitations were taken into account when assessing the risk of bias. Studies reported the risk of cirrhosis taken from death certificates or records of hospital admissions, with cirrhosis as the underlying cause. Therefore, the outcomes used by the studies were death or hospitalisation due to cirrhosis, rather than a diagnosis of cirrhosis. This might be confounded by the fact that people may change their alcohol consumption or diet following a diagnosis, and the risk of diagnosis with cirrhosis may differ from the risk of death or hospitalisation due to cirrhosis. For alcohol consumption as a risk factor, some studies looked at alcohol consumption at baseline and then assessed the risk over many years during which no further information on drinking behaviour was collected. Collecting alcohol intake data only once ignores the fact that people change their drinking habits over time, so longitudinal studies that assess alcohol consumption only at baseline may result in a misclassification. In the baseline questionnaire, some studies did not question subjects on their past drinking habits, meaning that ex-drinkers who were currently abstaining would have been included in their non-drinking group, therefore biasing the results. These limitations were taken into account when assessing the risk of bias. For alcohol consumption as a risk factor, the GDG noted that one study which compared 2 models of alcohol abuse (Model 1 and Model 2) had very large confidence intervals associated with the odds ratio (OR) of death/hospitalisation, which accounted for the Very Low quality evidence. Another study reported hazard ratios but without confidence intervals and therefore was included in the narrative only. For BMI as a risk factor, again the GDG noted the very large confidence intervals associated with the HR or OR, which accounted for the Very Low quality evidence. The GDG discussed the applicability of the evidence to a UK population. The GDG noted that the Liu 2010 study was a large 1.3 million population of NHS patients with a large number of events.
Other considerations	**Research recommendation** The GDG agreed that development of a risk tool was a high priority research recommendation. For much of the time, until presentation with jaundice or decompensation, liver

	disease may remain asymptomatic and silent. The earlier liver disease and even cirrhosis is diagnosed, the better the opportunity to treat, limiting disease progression but in many cases offering a cure. The prevention of progression to end-stage liver disease, avoiding complications, reducing the need for investigation, hospitalisation and intervention would have the potential for very large savings for the NHS. The earlier the diagnosis, the greater the potential patient and financial benefit. This is why GPs need a guide or 'toolkit' to identify people who are at high risk of having, or developing, advanced liver fibrosis or cirrhosis. One approach would be to identify a retrospective cohort of people with cirrhosis, and to look at their cirrhosis risk factors. The proposed study should use a multivariate analysis to find the risk factors associated with the outcome of cirrhosis. By weighting the risk factors according to their association with the outcome, a risk tool should be developed to predict a person's risk of developing cirrhosis.

6 Diagnostic tests

6.1 Introduction

Clinical evaluation can identify patients with cirrhosis when these individuals exhibit clinical signs of decompensated liver disease such as jaundice, ascites or hepatic encephalopathy. However, it is recognised that cirrhosis is not always clinically apparent, even to an experienced hepatologist[18], because people with compensated cirrhosis may experience few or no symptoms or signs of liver disease, Consequently, people identified to be at risk of cirrhosis require a confirmatory test.

Liver biopsy is considered the 'gold standard' to assess the stage of liver fibrosis in people with chronic liver disease and is the definitive method for confirming a diagnosis of cirrhosis. However, liver biopsy is expensive, is not popular with patients, and is associated with a small risk of severe complications such as bleeding and death.[25] It requires skilled practitioners to perform the procedure and to interpret liver histology; consequently, the application of liver biopsy is confined to secondary and tertiary care settings. Sampling error reduces the precision of liver biopsy to assess fibrosis and a biopsy specimen shorter than 25 mm in length increases the risk of inaccurate categorisation of liver fibrosis.[16] Indeed, liver biopsy may fail to detect cirrhosis in up to 15% of cases.[170] Given the problems associated with liver biopsy, simple non-invasive tests are often the preferred option to assess whether a person has cirrhosis[76], especially tests that can be employed in primary as well as secondary care. Patients need to be fully informed of the potential pros and cons of invasive and non-invasive test options so that, with the support of their clinician, they can choose the best method for them.

Routine laboratory liver blood tests have been evaluated, as predictors of cirrhosis, but normal values of bilirubin, albumin, aspartate aminotransferase (AST), and alanine aminotransferase (ALT) do not exclude cirrhosis. Combinations of routine laboratory blood tests are used to predict cirrhosis, including AST/ALT ratio, AST to platelet ratio (APRI), and FIB-4. Proprietary test panels employing blood tests that are surrogate markers of fibrogenesis include FibroTest and Enhanced Liver Fibrosis panel (ELF). Since increased liver fibrosis is associated with a greater degree of 'stiffness' of the liver, recent work has focused on measuring liver elastography using transient elastography (TE) and Acoustic Radiation Force Impulse (ARFI) imaging to assess liver fibrosis. These tests are performed in an outpatient setting and the results are available immediately. Magnetic resonance elastography has also been used to asses liver fibrosis. The GDG decided to compare the clinical and cost-effectiveness of routine laboratory blood tests, the blood fibrosis tests and imaging tests, both individually and in combination, to determine their performance characteristics for the diagnosis of cirrhosis against the reference standard which is examination of liver histology.

6.2 Review question: In people with suspected (or under investigation for) cirrhosis:

a) What is the most accurate blood fibrosis test to identify whether cirrhosis is present?

b) What is the most accurate non-invasive imaging test to identify whether cirrhosis is present?

c) Is the most accurate blood fibrosis test more accurate compared to an individual blood test to identify whether cirrhosis is present?

d) Is a combination of 2 non-invasive tests more accurate compared to a blood fibrosis test alone or an imaging test alone to identify whether cirrhosis is present?

For full details see the review protocol in Appendix C.

Table 16: Characteristics of review question

Population	Adults and young people >16 years with suspected (or under investigation for) cirrhosis. **Stratify studies based on the underlying cause:** Alcohol misuse disorders Hepatitis C NAFLD People with multiple aetiologies Primary biliary cholangitis (PBC) or primary sclerosing cholangitis (PSC) (reported separately)
Target condition	Cirrhosis
Index test(s)/comparator(s)	An individual blood test (for list see full protocol in Appendix C) A blood fibrosis test (for list see full protocol in Appendix C) A non-invasive imaging test (for list see full protocol in Appendix C) A combination of 2 non-invasive tests (for list see full protocol in Appendix C)
Reference standard(s)	Cirrhosis diagnosed by liver biopsy using one of the following scoring systems: Knodell score (F4), Ishak fibrosis score (F5 or F6), METAVIR (F4), Kleiner 2005 (for NAFLD) and Brunt 2001 (for NAFLD).
Statistical measure/outcomes	**Critical outcomes:** Specificity Sensitivity **Important outcomes:** ROC curve or area under curve
Study design	Cross-sectional studies, cohort studies, case series (including both retrospective and prospective analyses)

6.3 Clinical evidence

Fifty- three studies were included in the review[10,14,22,23,29-32,39,40,52,55,57,58,72,73,75,77,78,80,81,94,96,105,108,109,116,120,128-130,132,138,164,173,186,202,205,210-212,215,243-245,247,24823,24,62,69,70,136,200,249]. Study characteristics are summarised in Table 18 and evidence from these studies is summarised in the clinical evidence profiles below (Table 19, Table 20, Table 21, Table 22, Table 23, Table 24, Table 25, Table 26, Table 27, Table 28). See also the study selection flow chart in Appendix E, sensitivity/specificity forest plots, summary receiver operating characteristics (sROC) curves and ROC AUC plots in Appendix K, study evidence tables in Appendix H and exclusion list in Appendix L. Prospective and retrospective cohort studies in which the index test(s) and the reference standard test were applied to the same patients in a cross-sectional design were included in the review. The included population was those people suspected of having cirrhosis due to 1 of the specified risk factors. Two-gate study designs (sometimes referred to as case-control) are cross-sectional studies which compare the results of the index test in patients with an established diagnosis of cirrhosis with the results from healthy controls in order to assess the diagnostic accuracy of a test. This study design was excluded as it is unrepresentative of practice and is unlikely to contain the full spectrum of health and disease over which the test would be used. Studies of this design may lead to the selective inclusion of cases with more advanced disease and overestimations of sensitivity. The inclusion of healthy controls is likely to lead to overestimations of specificity.

The reference standard for all included studies was liver biopsy, with the level of fibrosis (and therefore the diagnosis of cirrhosis) scored using one of the fibrosis scoring systems specified in the protocol. There are known to be limitations with using liver biopsy for the diagnosis of cirrhosis. For example, the accuracy of liver biopsy can be affected by sampling errors and fibrosis heterogeneity within the liver itself. These inaccuracies are accentuated in biopsy samples of inadequate size. The UK standard criteria for an adequate biopsy length is ≥25 mm and containing at least 10 portal tracts. The GDG were aware that many studies fall below this operational definition. Studies including biopsies below this standard were not automatically excluded, but were downgraded in the quality of the evidence, as the accuracy of the reference standard will be compromised. The GDG also set a lower limit for the size of the biopsy, at which any studies including all or a proportion of biopsies below this lower limit would be excluded. This lower limit was set at 15 mm and 6 portal tracts, as the GDG felt that below this level the accuracy of the biopsy would be severely compromised and an accurate level of fibrosis would not be possible to assess. Studies including all or a proportion of biopsies below this level (or not stating the biopsy length) were excluded. The GDG discussed that setting these lower and upper limits would give the right balance between only including the higher quality evidence, without excluding a high proportion of the available studies and making conclusions on only a small proportion of the evidence. If a study reported that the biopsy was at least 15 mm ‘or’ 6 portal tracts, then this study was included, even if the other measure fell below the lower limit. The GDG thought that accurate staging of fibrosis and a diagnosis of cirrhosis would be possible as long as 1 of these parameters was met.

The fibrosis scoring using these systems is normally performed by an experienced histopathologist and has an element of subjectivity in the diagnosis. It is also a process that is subject to intra- and inter-observer variability. The ideal reference standard is a diagnosis of cirrhosis scored by a single pathologist, blinded to the patient’s clinical data and blinded to the diagnosis made using the index test. Therefore, when assessing the risk of bias for each study, the evidence quality was downgraded if the assessor was not blinded to patient clinical information or results of the index test, or if different people assessed biopsies from different patients, possibly introducing inter-observer variability.

Population strata for the different underlying aetiologies of liver disease which were pre-specified in the protocol were not combined in the analyses. These pre-specified strata were separated as there

are known to be factors, distinct to each aetiology, which will interfere with the results of the non-invasive tests (for example, alcohol consumption, portal hypertension, hepatic inflammation and obesity). Therefore, for the non-invasive tests there will be variation in the diagnostic accuracies and optimal cut-off thresholds for a positive result between aetiological groups. For this reason, studies reporting the diagnostic accuracy of index tests in mixed populations (without subgroup analysis by aetiology) were excluded from the review. The following population strata were separated within the analysis:

- People with hepatitis C virus infection (HCV or chronic hepatitis C, CHC): evidence summarised in Table 19 and Table 20.
- NAFLD: evidence summarised in Table 21 and Table 22.
- ALD: evidence summarised in Table 23 and Table 24.
- Primary biliary cirrhosis (PBC) or Primary sclerosing cholangitis (PSC): evidence summarised in Table 25 and Table 26.
- Multiple aetiologies: evidence summarised in Table 27 and Table 28.

For the multiple aetiologies stratum, evidence was only identified for people with HIV/HCV co-infection. No evidence was identified for the population stratum of primary sclerosing cholangitis specified in the protocol.

Forest plots showing the sensitivity and specificity values from the individual studies for each index test at relevant cut-off thresholds are summarised in Appendix K. The corresponding pooled sensitivity and specificity values of each index test at relevant cut-off thresholds are summarised in the clinical evidence profiles in Table 19, Table 21, Table 23, Table 25 and Table 27 (1 table for each population strata, sectioned into individual blood tests, blood fibrosis tests, imaging tests, and combinations of tests). Where evidence was available from 3 or more studies for an index test at the specified cut-off threshold, a diagnostic meta-analysis was performed and the pooled sensitivity and specificity value presented in the clinical evidence profile (along with the summary sensitivity and specificity value displayed in ROC space in Appendix K). Where evidence was available from fewer than 3 studies or a single study, the median sensitivity value is presented along with the corresponding specificity value from the same study, and the range of sensitivity and specificity values (no diagrams of the sensitivity and specificity values within ROC space were presented in Appendix K as a meta-analysis was not performed).

Studies may report sensitivity and specificity values at a pre-specified published cut-off threshold, or they may determine the optimal threshold from an ROC analysis. This resulted in a range of thresholds being reported for some index tests. If all the sensitivity and specificity values from the range of cut-off thresholds are pooled together, this can result in an overestimation of the diagnostic accuracy in comparison to another index test where sensitivity and specificity values are only reported for one cut-off threshold. For the below tests, the range of thresholds was considered too wide to pool the studies together in the analysis, and the cut-off thresholds were separated into the below categories prior to analysis:

- Transient elastography: low (9 to <13 kPa), medium (13 to <15 kPa), high (≥15 kPa)
- APRI: low (0.5 to <1.5), high (1.5 to 2.5)

For the following tests, the range of thresholds was considered narrow enough to pool together in the analysis:

- FibroTest: range 0.56–0.75
- ELF: range 9.3–10.44
- ARFI: range 1.55–2.0 m/s

In addition to reporting the sensitivity and specificity of a test at a particular cut-off threshold, some individual studies also report the AUC from an ROC analysis for each index test investigated. Where

available, this mean AUC value with its 95% CI was plotted on a graph for each index test. The AUC value and its 95% CI from the median study (along with the range of AUC values from all the studies) for each index test was summarised in the clinical evidence profiles in Table 20, Table 22, Table 24, Table 26 and Table 28 (1 table for each population stratum, sectioned into individual blood tests, blood fibrosis tests, imaging tests, and combinations of tests).

Some studies reported AST and ALT results on a transformed scale whereas other studies did not. This was to account for a change in laboratory reference levels introduced in 2003. Detection of enzyme activity is dependent on temperature, requiring all ALT and AST assays to be performed at 37°C. This resulted in a change in the upper limit of normal (ULN) level for both enzymes. Studies which were performed during the changeover period in 2003 may not always report whether they took into account this change. However, ratio measures such as the AST/ALT ratio would not be affected as both measures would be expected to increase by the same proportion. All studies were included even if they did not transform the data, as this was either normalised to the ULN or a ratio measure.

Table 17: Summary of index tests: components of non-invasive tests and applicable aetiology

Index Test	Components	Applicable aetiology
Individual blood tests		
Platelet count	–	HCV, NAFLD
Gamma-glutamyl transpeptidase (γGT)	–	ALD, PSC
Albumin	–	No population-specific data
Prothrombin time (INR)	–	No population-specific data
Aspartate transaminase (AST)	–	No population-specific data
Alanine transaminase (ALT)	–	No population-specific data
Bilirubin	–	No population-specific data
Blood fibrosis tests		
AST to Platelet Ratio Index (APRI)	–	HCV, ALD, NAFLD
AST/ALT ratio	–	HCV, NAFLD
FIB-4	Age, AST, ALT, platelet count	HCV, NAFLD
Enhanced liver fibrosis (ELF) test	PIIINP, hyaluronate, TIMP-1	HCV, NAFLD
FibroTest	γ-GT, haptoglobin, bilirubin, A1 apolipopotein, alpha2-macroglobulin	HCV, ALD, NAFLD
Imaging tests		
Acoustic radiation force impulse imaging (ARFI)	–	HCV, NAFLD
Real time elastography	–	HCV, NAFLD
Point shear wave elastography (pSWE)	–	HCV
Transient elastography (TE,	–	HCV, ALD, NAFLD

Index Test	Components	Applicable aetiology
Fibroscan)		
MR elastography	–	All
Ultrasound	–	All
Combinations of tests		
SAFE algorithm	Based on sequential use of APRI, FibroTest and liver biopsy. APRI as the initial screening test with a low and high cut-off and FibroTest as a second step	HCV
Castera algorithm	Combination of TE and FibroTest. When TE and FibroTest agree no biopsy is performed whereas when they disagree, liver biopsy is needed	HCV

Table 18: Summary of studies included in the review

Study	Population	Index tests reported	Cut-off threshold (if reported)	Reference standard scoring system
ARENA 2008[10]	HCV (n=150)	Transient elastography	14.8 kPa	METAVIR
AYKUT 2014[14]	NAFLD (n=88)	Transient elastography	Not reported	Kleiner
BORRONI 2006[22]	HCV (n=228)	APRI AST/ALT ratio	2 1	Knodell
BOTA 2011A[23]	HCV (n=212)	Transient elastography APRI	13.3 1	METAVIR
BOTA 2015[24]	HCV (n=117)	ARFI	1.87 m/s	METAVIR
CARDOSO 2012[29]	HCV (n=363)	Transient elastography	12.5 kPa	METAVIR
CASTERA 2010A[30]	HCV (n=302)	SAFE algorithm (based on sequential use of APRI, FibroTest and liver biopsy. APRI as the initial screening test with a low and high cut-off and FibroTest as a second step.) Castera algorithm (combination of TE and FibroTest. When TE and FibroTest agree no biopsy is performed whereas when they disagree, liver biopsy is needed.)	If APRI lower than low cut-off (1.0) then cirrhosis absent, if higher than 1.0 then FibroTest performed. FibroTest ≤0.48 (cirrhosis absent), FibroTest 0.49–0.74 (liver biopsy needed) and ≥0.75 (cirrhosis present) TE ≥12.5 and FT <0.75 (disagree), TE <12.5 and FT ≥0.75 (disagree), TE failure (disagree), TE <12.5 and FT <0.75 (agree cirrhosis absent), TE ≥12.5 and FT ≥0.75 (agree cirrhosis present)	METAVIR

Study	Population	Index tests reported	Cut-off threshold (if reported)	Reference standard scoring system
CATANZARO 2013[31]	HCV (n=162)	ELF APRI	9.3 1.19	METAVIR
CAVIGLIA 2013[32]	HCV (n=57)	Transient elastography	13.8	METAVIR
CHEN 2012[39]	HCV (n=127)	FibroTest ARFI	Not reported 1.98 m/s	METAVIR
CHRYSANTHOS 2006[40]	HCV (n=284)	APRI	1.0 2.0	Ishak
DE 2006[52]	HCV/HIV co-infected (n=72)	Transient elastography Platelet count APRI AST/ALT ratio FIB-4	11.8 kPa 140 g/litre 2 1 3.25	METAVIR
ESMAT 2013[57]	HCV (n=164)	Transient elastography	12.5 kPa	METAVIR
FAHMY 2011[58]	HCV (n=110)	Transient elastography	16.5 kPa	METAVIR
FERNANDES 2015[62]	HCV (n=120)	Transient elastography ELF	Results not reported 10.44	METAVIR
FERRAIOLI 2014[70]	HCV (n=102)	pSWE Transient elastography	7.2 kPa 9.3 kPa	METAVIR
FIERBINTEANU	NAFLD (n=64)	ARFI	1.636 m/s	Kleiner

Study	Population	Index tests reported	Cut-off threshold (if reported)	Reference standard scoring system
BRATICEVICI 2013[72]				
FLOREANI 2011[75]	PBC (n=114)	Transient elastography APRI FIB-4 AST/ALT ratio Combination of TE with each marker	11.4 Not reported Not reported Not reported Not reported	METAVIR
FRIEDRICH-RUST 2010[78]	HCV (n=36)	FibroTest ELF Transient elastography	0.73 10.31 12.5	METAVIR
FRIEDRICH-RUST 2010A[77]	NAFLD/NASH (non-alcoholic steatohepatitis; n=50)	Transient elastography M probe Transient elastography XL probe	Not reported Not reported	Kleiner
FUJII 2009[80]	NASH (n=50)	AST/ALT ratio APRI	Not reported Not reported	Brunt
GAIA 2011[81]	HCV (n=77)	Transient elastography	11.5 kPa	METAVIR (for HCV population)

Study	Population	Index tests reported	Cut-off threshold (if reported)	Reference standard scoring system
	NAFLD (n=72)		10.5 kPa	Brunt (for NAFLD population)
GUECHOT 2012[94]	HCV (n=512)	ELF	9.35	METAVIR
HALFON 2007[96]	HCV (n=356)	FibroTest APRI	0.56 0.83	METAVIR
JANSSENS 2010[105]	ALD (n=49)	APRI Transient elastography	2.0 19.6 kPa 21.1 kPa 23.5 kPa	METAVIR
KAYADIBI 2014[108]	HCV (n=202)	FIB-4 APRI AST/ALT ratio AST ALT Platelet count	Not reported Not reported Not reported Not reported Not reported Not reported	METAVIR
KETTANEH 2007[109]	HCV (n=560)	Transient elastography	Not reported	METAVIR
LACKNER 2005[116]	HCV (n=194)	AST/ALT ratio APRI Platelet count	1.0 1.0 2.0 130 x 109 litres 150 x 109 litres	Ishak
LEROY 2014[120]	HCV (n=255)	FibroTest	0.63 0.74	METAVIR
LUPSORPLANTON	HCV (n=1202)	Transient elastography	13.2 kPa	METAVIR

Study	Population	Index tests reported	Cut-off threshold (if reported)	Reference standard scoring system
2013[128]				
MACIAS 2006[130]	HCV HIV co-infected (n=263)	APRI AST/ALT ratio	1 2 0.6	Knodell
MARTINEZ 2011A[132]	HCV (n=340)	APRI FIB-4 ELF	1 2 Not reported 0.06 1.73	METAVIR
MUELLER 2010[136]	ALD (n=101)	Transient elastography	11.5 kPa 12.5 kPa	Kleiner
MYERS 2012B[138]	NAFLD, BMI≥28 (n=127)	Transient elastography M probe Transient elastography XL probe	22.3 kPa 16.0 kPa	METAVIR
RIZZO 2011[173]	HCV (n=139)	Transient elastography ARFI	11 kPa 2 m/s	METAVIR
SANCHEZ-CONDE 2010[186]	HCV HIV co-infected (n=100)	Transient elastography	14 kPa	METAVIR
SHEHAB 2014[200]	HCV (n=842)	APRI FIB-4	0.5 and 2.0 3.25	METAVIR

Study	Population	Index tests reported	Cut-off threshold (if reported)	Reference standard scoring system
SILVA JUNIOR 2014[202]	HCV (n=51)	ARFI APRI FIB-4	1.95 m/s 1.71 Not reported	METAVIR
SIRLI 2010[205]	HCV (n=150)	Transient elastography APRI FIB-4 Platelet count	13.3 kPa 1.38 2.3122 155,000/mm^3	METAVIR
SPOREA 2011A[212]	HCV (n=197)	Transient elastography ARFI Combination of Transient elastography and ARFI	12.2 kPa 1.8 m/s As above	METAVIR
SPOREA 2012A[211] Includes data from: LUPSOR 2009[129], EBINUMA 2011[55], FIERBINTEANUBRATI CEVICI 2009[73], SPOREA 2011D[210], PISCAGLIA[164]	HCV (n=914)	ARFI Transient elastography	1.55 m/s 11.9 kPa 13.1 kPa	METAVIR
STIBBE 2011[215]	HCV (n-41)	FibroTest Transient elastography	0.75 14 kPa	METAVIR
WONG 2010B[244]	NAFLD (n=246)	Transient elastography	10.3 kPa	Kleiner

Study	Population	Index tests reported	Cut-off threshold (if reported)	Reference standard scoring system
			11.5 kPa	
		APRI	Not reported	
		AST/ALT ratio	Not reported	
		FIB-4	Not reported	
WONG 2012[243]	NAFLD (n=193)	Transient elastography M probe	10.3 kPa 11.5 kPa	Kleiner
		Transient elastography XL probe	7.2 kPa 7.9 kPa 11.0 kPa	
YAMADA 2006[245]	HCV (n=44)	Ultrasound	Not reported	METAVIR
YONENDA 2008[247]	NASH (n=97)	Transient elastography	17.5 kPa	Brunt
YONENDA 2010[248]	NAFLD (n=54)	Transient elastography	16 kPa	Brunt
		ARFI	1.9 m/s	
ZARSKI 2012[249]	HCV (n=436)	Transient elastography	12.9 kPa	METAVIR
		FibroTest	0.74	
		APRI	2.0	
		FIB-4	0.84	

6.3.1 Hepatitis C

Table 19: Clinical evidence profile (sensitivity and specificity): HCV population

Index Test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
Individual blood tests									
Albumin	0					–			
Platelet count	2	344	Serious[a]	None[b]	None[c]	None[d]	76 (55, 91) 87 (60, 98)[e]	88 (82, 93) 84 (76, 89)[e]	MODERATE
Prothrombin Time (INR)	0					–			
AST	0					–			
ALT	0					–			
Bilirubin	0					–			
γGT	0					–			
Blood fibrosis tests									
FibroTest (cut-off range 0.56–0.75)[f]	4	1289	Very serious[a]	Serious[b]	None[c]	Very serious[d]	Pooled[g] 80.3 (39.5, 98.9)	Pooled[g] 69.3 (26.7, 94.7)	VERY LOW
ELF (cut-off range 9.3–10.44)[h][j]	3	794	Very serious[a]	None[b]	None[c]	None[d]	Pooled 81.4 (56.9, 95.7)	Pooled 80.0 (51.4, 95.0)	LOW
APRI (low cut-off 0.5 to <1.5)[k]	7	1699	Very serious[a]	Serious[b]	None[c]	None[d]	Pooled 83.8 (70.6, 94.0)	Pooled 77.8 (68.6, 85.4)	VERY LOW
APRI (high cut-off 1.5–2.5)[l]	5	1285	Very serious[a]	Serious[b]	None	Serious[d]	Pooled 36.5 (12.3, 67.8)	Pooled 94.4 (83.5, 99.1)	VERY LOW

Index Test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
FIB-4 (cut-off 2.3122)[(m)]	1	150	Very serious[(a)]	None[(b)]	None[(c)]	Serious[(d)]	80 (52, 96)	78 (70, 84)	VERY LOW
AST/ALT ratio (cut-off 1.0)	2	421	Very serious[(a)]	None[(b)]	None[(c)]	None[(d)]	30 (15, 49) 35 (19, 55) [(e)]	97 (94, 99) 90 (84, 94) [(e)]	LOW
Imaging tests									
Transient elastography (low cut-off 9.0 to <13.0 kPa)[(i)]	7	1424	Very serious[(a)]	Serious[(b)]	None[(c)]	None[(d)]	Pooled 81.5 (68.4, 91.2)	Pooled 90.4 (84.9, 94.4)	LOW
Transient elastography (medium cut-off 13.0 to <15.0 kPa)	7	1923	Very serious[(a)]	None[(b)]	None[(c)]	Serious[(d)]	Pooled 93.4 (87.9, 97.0)	Pooled 92.9 (86.5, 97.0)	VERY LOW
Transient elastography (high cut-off 15+ kPa)	1	110	Very serious[(a)]	None[(b)]	None[(c)]	Serious[(d)]	86 (65, 97)	91 (83, 96)	VERY LOW
ARFI (cut-off range 1.55– 2.0 m/s)	6	1541	Very serious[(a)]	Serious[(b)]	None[(c)]	Serious[(d)]	Pooled 88.1 (78.5, 95.1)	Pooled 84.3 (74.9, 91.7)	VERY LOW
pSWE	1	102	Serious[(a)]	None[(b)]	None[(c)]	Serious[(d)]	90 (55, 100)	89 (81, 95)	LOW
Ultrasound	0	–							
MR elastography	0	–							
Combinations of non-invasive tests									
Transient elastography + ARFI (12.2 kPa and 1.8 m/s)	1	197	Very serious[(a)]	None[(b)]	None[(c)]	None[(d)]	85 (72, 93)	94 (89, 98)	LOW
Transient elastography or ARFI (12.2 kPa or 1.8 m/s)	1	197	Very serious[(a)]	None[(b)]	None[(c)]	Serious[(d)]	96 (87, 100)	83 (76, 89)	VERY LOW

Index Test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
SAFE algorithm	1	302	Serious[a]	None[b]	None[c]	None[d]	87 (77, 94)	90 (85, 93)	MODERATE
Castera algorithm	1	302	Serious[a]	None[b]	None[c]	None[d]	89 (80, 95)	98 (96, 100)	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity, as this was the primary measure for decision-making pre-specified in the protocol. If data were available from 3 or more studies, a diagnostic meta-analysis was performed and the pooled sensitivity and specificity result presented. Otherwise, sensitivity and specificity values were presented from the individual studies.

a) Risk of bias was assessed using the QUADAS-II checklist for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b) Inconsistency was assessed by inspection of the sensitivity forest plots, or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

c) Indirectness was assessed using the QUADAS-II checklist items referring to applicability for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at serious indirectness, and downgraded by 2 increments if the majority of the evidence was at very serious indirectness

d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis, if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted, imprecision was assessed according to the sensitivity confidence region of the largest study. Imprecision was assessed on the sensitivity confidence region as the primary measure for decision-making. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

e) Data available from 2 studies. A meta-analysis was not performed if only 2 studies were available, therefore the sensitivity and corresponding specificity values for both studies are presented

f) Data from 2 studies could not be combined in the analysis as the studies (Friedrich-rust 2010 and Leroy 2014)[78,120] only reported the sensitivity and specificity values and not the 2x2 table values or prevalence. Friedrich-rust 2010: cut-off 0.73, sensitivity: 67%, specificity: 81%; Leroy 2014: cut-off 0.74, sensitivity: 59%, specificity: 91%

g) Model did not converge due to limited data. Pooled result presented is using available results from the model as the statistics suggest the model is the best fit available

h) Data from 1 study could not be combined in the analysis as the study (Friedrich-rust 2010)[78] only reported the sensitivity and specificity values and not the 2x2 table values or prevalence. Friedrich-rust 2010: cut-off 10.31, sensitivity: 89%, specificity: 63%

i) Data from 1 study could not be combined in the analysis as the study (Friedrich-rust 2010)[78] only reported the sensitivity and specificity values and not the 2x2 table values or prevalence. Friedrich-rust 2010: cut-off 12.5 kPa, sensitivity: 78%, specificity: 84%

j) Martinez study[132] excluded from the analysis as the ELF cut-off thresholds of 0.06 and 0.82 did not match other studies (presumed to use an older version of ELF)

k) Data from 1 study could not be combined in the analysis as the study (Shehab 2014)[200] only reported the sensitivity and specificity values and not the 2x2 table values or prevalence (for the 842 patients included in the final analysis). Shehab 2014: APRI cut-off 0.5, sensitivity: 100%, specificity: 12.8%

l) Data from 1 study could not be combined in the analysis as the study (Shehab 2014)[200] only reported the sensitivity and specificity values and not the 2x2 table values or prevalence (for the 842 patients included in the final analysis). Shehab 2014: APRI cut-off 2.0, sensitivity: 15.4%, specificity: 96%
m) Data from 1 study could not be combined in the analysis as the study (Shehab 2014)[200] only reported the sensitivity and specificity values and not the 2x2 table values or prevalence (for the 842 patients included in the final analysis). Shehab 2014: FIB-4 cut-off 3.25, sensitivity: 28.2%, specificity: 93.5%

Table 20: Clinical evidence profile (AUC): HCV population

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (CIs), [range]	Quality
Individual blood tests								
Albumin	0					–		
Platelet count	3	546	Serious[a]	None[b]	None[c]	None[d]	89.0 (83–94) [82.7–89.9]	MODERATE
Prothrombin time (INR)	0					–		
AST	1	202	Serious[a]	None[b]	None[c]	None[d]	75.2 (67.1– 83.2)	MODERATE
ALT	1	202	Serious[a]	None[b]	None[c]	None[d]	62.6 (53.4–71.7)	MODERATE
Bilirubin	0					–		
γGT	0					–		
Blood fibrosis tests								
FibroTest	4	1375	Serious[a]	None[b]	None[c]	None[d]	86.5 [75.7–87.0]	MODERATE
ELF	4	1134	Very serious[a]	None[b]	None[c]	None[d]	83.5 [72–94]	LOW
APRI	8	2015	Very serious[a]	None[b]	None[c]	None[d]	88.0 [84.7-92.0]	LOW
FIB-4	4	1074	Very serious[a]	None[b]	None[c]	None[d]	84.8 [84–89]	LOW
AST/ALT ratio	3	623	Serious[a]	None[b]	None[c]	None[d]	73 (63, 83) [61–76]	MODERATE
Imaging tests								
Transient elastography	10	3158	Very serious[a]	None[b]	None[c]	None[d]	92.6 [75.7–97.9]	LOW

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (CIs), [range]	Quality
ARFI	2	266	Very serious[(a)]	None[(b)]	None[(c)]	None[(d)]	Mean 86.1 [83.1–89]	LOW
pSWE	1	102	Serious[(a)]	None[(b)]	None[(c)]	Serious[(d)]	95 (89, 99)	LOW
Ultrasound	0	–						
MR elastography	0	–						
Combinations of non-invasive tests								
Transient elastography + ARFI	0	–						
Transient elastography or ARFI	0	–						
SAFE algorithm	1	197	Serious[(a)]	None[(b)]	None[(c)]	None[(d)]	87 (84, 90)	MODERATE
Castera algorithm	1	197	Serious[(a)]	None[(b)]	None[(c)]	Serious[(d)]	93 (90, 96)	LOW

a) Risk of bias was assessed using the QUADAS-II checklist for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b) Inconsistency was assessed by inspection of the AUC values across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

c) Indirectness was assessed using the QUADAS-II checklist items referring to applicability for each individual study for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at serious indirectness, and downgraded by 2 increments if the majority of the evidence was at very serious indirectness

d) The judgement of precision was based on the median AUC value and 95% CI. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

6.3.2 NAFLD

Table 21: Clinical evidence profile (sensitivity and specificity): NAFLD population

Index Test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
Individual blood tests									
Albumin	0					–			
Platelet count	0					–			
Prothrombin time (INR)	0					–			
AST	0					–			
ALT	0					–			
Bilirubin	0					–			
γGT	0					–			
Blood fibrosis tests									
FibroTest	0					–			
ELF	0					–			
APRI (low cut-off 0.5 to <1.5)	0					–			
APRI (high cut-off 1.5–2.5)	0					–			
FIB-4	0					–			
AST/ALT ratio	0					–			
Imaging tests									
Transient elastography (low cut-off 10.0 to <13.0)	2	318	Very serious[a]	None[b]	None[c]	Serious[d]	78 (40, 97) 76 (55, 91)[e]	95 (87, 99) 91 (86, 94)[e]	VERY LOW
Transient elastography (medium cut-off 13.0 to <15.0)	0					–			
Transient elastography high cut-off >15[f]	2	151	Very serious[a]	None[b]	None[c]	Serious[d]	100 (66, 100) 100 (54, 100)[e]	97 (90, 99) 98 (89, 100)[e]	VERY LOW

Index Test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
ARFI (cut-off range 1.636–1.9)	2	118	Very serious[a]	Serious[b]	None[c]	Serious[d]	92 (62, 100) 100 (54, 100)[e]	92 (81, 98) 96 (86, 99)[e]	VERY LOW
pSWE	0	–							
Ultrasound	0	–							
MR elastography	0	–							
Combinations of non-invasive tests									
	0	–							

The assessment of the evidence quality was conducted with emphasis on test sensitivity, as this was the primary measure for decision-making pre-specified in the protocol. If data were available from 3 or more studies, a diagnostic meta-analysis was performed and the pooled sensitivity and specificity result presented. Otherwise, sensitivity and specificity values were presented from the individual studies.

a) Risk of bias was assessed using the QUADAS-II checklist. Evidence quality was downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b) Inconsistency was assessed by inspection of the sensitivity forest plots, or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

c) Indirectness was assessed using the QUADAS-II checklist items referring to applicability. Evidence quality was downgraded by 1 increment if the majority of the evidence was at serious indirectness, and downgraded by 2 increments if the majority of the evidence was at very serious indirectness

d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis, if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted, imprecision was assessed according to the sensitivity confidence region of the largest study. Imprecision was assessed on the sensitivity confidence region as the primary measure for decision-making. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

e) Data available from 2 studies. A meta-analysis was not performed if only 2 studies were available, therefore the sensitivity and corresponding specificity values for both studies are presented

f) Data from 1 study could not be combined in the analysis as the study (Myers 2012[138]) only reported the sensitivity and specificity values and not the 2x2 table values or prevalence. Myers 2012: M probe high cut-off 22.3 kPa, sensitivity: 80%, specificity: 91%

g) Data from 1 study could not be combined in the analysis as the study (Myers 2012[138]) only reported the sensitivity and specificity values and not the 2x2 table values or prevalence. Myers 2012: XL probe high cut-off 16kPa, sensitivity: 100%, specificity: 91%

Table 22: Clinical evidence profile (AUC): NAFLD population

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (CIs), [range]	Quality
Individual blood tests								
Albumin	0	–						
Platelet count	0	–						
Prothrombin time (INR)	0	–						
AST	0	–						
ALT	0	–						
Bilirubin	0	–						
γGT	0	–						
Blood fibrosis tests								
FibroTest	0	–						
ELF	0	–						
APRI	2	296	Very serious[(a)]	None[(b)]	None[(c)]	None[(d)]	Mean 76.8 [75–78.6]	LOW
FIB-4	1	246	Very serious[(a)]	None[(b)]	None[(c)]	None[(d)]	81 (73, 89)	LOW
AST/ALT ratio	2	296	Very serious[(a)]	None[(b)]	None[(c)]	None[(d)]	Mean 73.7 [66–81.3]	LOW
Imaging tests								
Transient elastography	3	406	Very serious[(a)]	None[(b)]	None[(c)]	Serious[(d)]	94 (88–100) [91–95]	VERY LOW
ARFI	1	64	Very serious[(a)]	None[(b)]	None[(c)]	None[(d)]	98.4 (95.8,100)	LOW
pSWE	0	–						

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (CIs), [range]	Quality
Ultrasound	0	–						
MR elastography	0	–						
Combinations of non-invasive tests								
	0	–						

a) Risk of bias was assessed using the QUADAS-II checklist for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
b) Inconsistency was assessed by inspection of the AUC values across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–95% and 95–100%)
c) Indirectness was assessed using the QUADAS-II checklist items referring to applicability for each individual study for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at serious indirectness, and downgraded by 2 increments if the majority of the evidence was at very serious indirectness
d) The judgement of precision was based on the median AUC value and 95% CI. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

6.3.3 ALD

Table 23: Clinical evidence profile (sensitivity and specificity): ALD population

Index test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
Individual blood tests									
Albumin	0	–							
Platelet count	0	–							
Prothrombin time (INR)	0	–							
AST	0	–							

Index test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
ALT	0					–			
Bilirubin	0					–			
γGT	0					–			
Blood fibrosis tests									
FibroTest	0					–			
ELF	0					–			
APRI (low cut-off 0.5 to <1.5)	0					–			
APRI (high cut-off 1.5–2.5)	1	48	Very serious[a]	None[b]	Serious[c]	Serious[d]	40 (19, 64)	61 (41, 78)	VERY LOW
FIB-4	0					–			
AST/ALT ratio	0					–			
Imaging tests									
Transient elastography (low cut-off 11.0 to <13.0)	1	25	Very serious[a]	None[b]	None[c]	Serious[d]	100 (54, 100)	79 (54, 94)	VERY LOW
Transient elastography (medium cut-off 13.0 to <15.0)	0					–			
Transient elastography (high cut-off 15+)	1	49	Very serious[a]	None[b]	Serious[c]	None[d]	80 (56, 94)	76 (56, 90)	VERY LOW
ARFI	0					–			
pSWE	0					–			
Ultrasound	0					–			
MR elastography	0					–			
Combinations of non-invasive tests									
	0					–			

The assessment of the evidence quality was conducted with emphasis on test sensitivity, as this was the primary measure for decision-making pre-specified in the protocol. If data were available from 3 or more studies, a diagnostic meta-analysis was performed and the pooled sensitivity and specificity result presented. Otherwise, sensitivity and specificity values were presented from the individual studies.

a) Risk of bias was assessed using the QUADAS-II checklist. Evidence quality was downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b) Inconsistency was assessed by inspection of the sensitivity forest plots, or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

c) Indirectness was assessed using the QUADAS-II checklist items referring to applicability. Evidence quality was downgraded by 1 increment if the majority of the evidence was at serious indirectness, and downgraded by 2 increments if the majority of the evidence was at very serious indirectness. Population of included study may be preselected to have more severe fibrosis

d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted, imprecision was assessed according to the sensitivity confidence region of the largest study. Imprecision was assessed on the sensitivity confidence region as the primary measure for decision-making. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

e) Data available from 2 studies. A meta-analysis was not performed if only 2 studies were available, therefore the sensitivity and corresponding specificity values for both studies are presented

Table 24: Clinical evidence profile (AUC): ALD population

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (CIs), [range]	Quality
Individual blood tests								
Albumin	0					–		
Platelet count	0					–		
Prothrombin time (INR)	0					–		
AST	0					–		
ALT	0					–		
Bilirubin	0					–		
γGT	0					–		
Blood fibrosis tests								
FibroTest	0					–		

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (CIs), [range]	Quality
ELF	0					–		
APRI	0					–		
FIB-4	0					–		
AST/ALT ratio	0					–		
Imaging tests								
Transient elastography	1	25	Very serious[a]	None[b]	None[c]	Serious[d]	92.1 (87, 97)	VERY LOW
ARFI	0					–		
pSWE	0					–		
Ultrasound	0					–		
MR elastography	0					–		
Combinations of non-invasive tests								
	0					–		

a) Risk of bias was assessed using the QUADAS-II checklist for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b) Inconsistency was assessed by inspection of the AUC values across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

c) Indirectness was assessed using the QUADAS-II checklist items referring to applicability for each individual study for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at serious indirectness, and downgraded by 2 increments if the majority of the evidence was at very serious indirectness

d) The judgement of precision was based on the median AUC value and 95% CI. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

6.3.4 Primary biliary cholangitis (PBC)

Table 25: Clinical evidence profile (sensitivity and specificity): PBC population

Index test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
Individual blood tests									
Albumin	0					–			
Platelet count	0					–			
Prothrombin time (INR)	0					–			
AST	0					–			
ALT	0					–			
Bilirubin	0					–			
γGT	0					–			
Blood fibrosis tests									
FibroTest	0					–			
ELF	0					–			
APRI	0					–			
FIB-4	0					–			
AST/ALT ratio	0					–			
Imaging tests									
Transient elastography (cut-off 11.4)	1	114	Serious[a]	None[b]	None[c]	Serious[d]	100 (80, 100)	94 (87, 98)	LOW
ARFI	0					–			
pSWE	0					–			
Ultrasound	0					–			
MR elastography	0					–			
Combinations of non-invasive tests									

Index test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
	0	–							

The assessment of the evidence quality was conducted with emphasis on test sensitivity, as this was the primary measure for decision-making pre-specified in the protocol. If data were available from 3 or more studies, a diagnostic meta-analysis was performed and the pooled sensitivity and specificity result presented. Otherwise, sensitivity and specificity values were presented from the individual studies.

a) Risk of bias was assessed using the QUADAS-II checklist. Evidence quality was downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b) Inconsistency was assessed by inspection of the sensitivity forest plots, or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

c) Indirectness was assessed using the QUADAS-II checklist items referring to applicability. Evidence quality was downgraded by 1 increment if the majority of the evidence was at serious indirectness, and downgraded by 2 increments if the majority of the evidence was at very serious indirectness

d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted, imprecision was assessed according to the sensitivity confidence region of the largest study. Imprecision was assessed on the sensitivity confidence region as the primary measure for decision-making. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

e) Data available from 2 studies. A meta-analysis was not performed if only 2 studies were available, therefore the sensitivity and corresponding specificity values for both studies are presented

Table 26: Clinical evidence profile (AUC): PBC population

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (CIs), [range]	Quality
Individual blood tests								
Albumin	0	–						
Platelet count	0	–						
Prothrombin time (INR)	0	–						
AST	0	–						

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (CIs), [range]	Quality
ALT	0					–		
Bilirubin	0					–		
γGT	0					–		
Blood fibrosis tests								
FibroTest	0					–		
ELF	0					–		
APRI	1	114	Serious[a]	None[b]	None[c]	Serious[d]	84 (74, 97)	LOW
FIB-4	1	114	Serious[a]	None[b]	None[c]	None[d]	74 (58, 88)	MODERATE
AST/ALT ratio	1	114	Serious[a]	None[b]	None[c]	Serious[d]	58 (42, 74)	LOW
Imaging tests								
Transient elastography	1	114	Serious[a]	None[b]	None[c]	Serious[d]	99 (94, 100)	LOW
ARFI	0					–		
pSWE	0					–		
Ultrasound	0					–		
MR elastography	0					–		
Combinations of non-invasive tests								
Transient elastography + APRI	1	114	Serious[a]	None[b]	None[c]	Serious[d]	99 (94, 100)	LOW
Transient elastography + FIB-4	1	114	Serious[a]	None[b]	None[c]	Serious[d]	99 (94, 100)	LOW
Transient elastography + AST/ALT ratio	1	114	Serious[a]	None[b]	None[c]	Serious[d]	99 (94, 100)	LOW

a) Risk of bias was assessed using the QUADAS-II checklist for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b) Inconsistency was assessed by inspection of the AUC values across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

c) Indirectness was assessed using the QUADAS-II checklist items referring to applicability for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at serious indirectness, and downgraded by 2 increments if the majority of the evidence was at very serious indirectness

d) The judgement of precision was based on the median AUC value and 95% CI. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

6.3.5 Multiple aetiologies

Table 27: Clinical evidence profile (sensitivity and specificity): HIV/HCV population

Index test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
Individual blood tests									
Albumin	0					–			
Platelet count (150x109/litre)	1	263	Very serious[a]	None[b]	None[c]	Serious[d]	63 (46, 77)	77 (71, 82)	VERY LOW
Prothrombin time (INR)	0					–			
AST	0					–			
ALT	0					–			
Bilirubin	0					–			
γGT	0					–			
Blood fibrosis tests									
FibroTest	0					–			
ELF	0					–			
APRI (low cut-off 0.5 to <1.5)	1	263	Very serious[a]	None[b]	None[c]	None[d]	78 (62, 89)	57 (50, 63)	LOW
APRI (high cut-off 1.5–2.5)	1	263	Very serious[a]	None[b]	None[c]	Serious[d]	53 (36, 68)	89 (84, 93)	VERY LOW
FIB-4	0								
AST/ALT ratio (cut-off 1.0)	1	263	Very serious[a]	None[b]	None[c]	Serious[d]	38 (23, 54)	77 (71, 82)	VERY LOW

Index test (threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (95% CI)	Specificity % (95% CI)	Quality
Imaging tests									
Transient elastography (low threshold 11.8)	1	72	Serious	None[b]	None[c]	Serious[d]	100 (80, 100)	93 (82, 98)	LOW
Transient elastography (medium cut-off range 14.0–14.5)	2	172	Very serious[a]	Serious[b]	None[c]	Serious[d]	88 (64, 99) 100 (63, 100)[e]	96 (87, 100) 93 (86, 98)[e]	VERY LOW
ARFI	0	–							
pSWE	0	–							
Ultrasound	0	–							
MR elastography	0	–							
Combinations of non-invasive tests									
	0	–							

The assessment of the evidence quality was conducted with emphasis on test sensitivity, as this was the primary measure for decision-making pre-specified in the protocol. If data were available from 3 or more studies, a diagnostic meta-analysis was performed and the pooled sensitivity and specificity result presented. Otherwise, sensitivity and specificity values were presented from the individual studies.

a) Risk of bias was assessed using the QUADAS-II checklist. Evidence quality was downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b) Inconsistency was assessed by inspection of the sensitivity forest plots, or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

c) Indirectness was assessed using the QUADAS-II checklist items referring to applicability. Evidence quality was downgraded by 1 increment if the majority of the evidence was at serious indirectness, and downgraded by 2 increments if the majority of the evidence was at very serious indirectness

d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis, if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted, imprecision was assessed according to the sensitivity confidence region of the largest study. Imprecision was assessed on the sensitivity confidence region as the primary measure for decision-making. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold

above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%)
e) Data available from 2 studies. A meta-analysis was not performed if only 2 studies were available, therefore the sensitivity and corresponding specificity values for both studies are presented

Table 28: Clinical evidence profile (AUC): HIV/HCV population

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (CIs), [range]	Quality
Individual blood tests								
Albumin	0					–		
Platelet count	2	335	Very serious[a]	None[b]	None[c]	None[d]	Mean 79.5 [79–80]	LOW
Prothrombin time (INR)	0					–		
AST	0					–		
ALT	0					–		
Bilirubin	0					–		
γGT	0					–		
Blood fibrosis tests								
FibroTest	0					–		
ELF	0					–		
APRI	2	335	Very serious[a]	None[b]	None[c]	None[d]	Mean 77.5 [76–79]	LOW
FIB-4	1	72	Serious[a]	None[b]	None[c]	None[d]	73 (57, 89)	MODERATE
AST/ALT ratio	2	335	Very serious[a]	Serious[b]	None[c]	None[d]	Mean 52.5 [45–60]	VERY LOW
Imaging tests								
Transient elastography	2	172	Very serious[a]	None[b]	None[c]	None[d]	Mean 80 [97–99]	LOW
ARFI	0					–		

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (CIs), [range]	Quality
pSWE	0	–						
Ultrasound	0	–						
MR elastography	0	–						
Combinations of non-invasive tests								
	0	–						

a) Risk of bias was assessed using the QUADAS-II checklist for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b) Inconsistency was assessed by inspection of the AUC values across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

c) Indirectness was assessed using the QUADAS-II checklist items referring to applicability for each individual study for each individual study. Evidence quality was downgraded by 1 increment if the majority of the evidence was at serious indirectness, and downgraded by 2 increments if the majority of the evidence was at very serious indirectness

d) The judgement of precision was based on the median AUC value and 95% CI. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the 95% threshold set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the confidence interval varied across 2 areas (for example 50–95% and 95–100%) and by 2 increments if the confidence interval varied across 3 areas (for example 0–50%, 50–95% and 95–100%)

6.4 Economic evidence

6.4.1 Published literature

One economic evaluation was identified that compared a 'no testing' strategy with liver biopsy and transient elastography in chronic hepatitis C patients with no fibrosis.[27]

One economic evaluation was identified that compared liver biopsy and transient elastography in 3 relevant patient subgroups: hepatitis B, hepatitis C and NAFLD.[213]

One economic evaluation was identified that compared liver biopsy, transient elastography, ELF and FibroTest in patients with suspected liver fibrosis related to alcohol consumption.[214]

These are summarised in the economic evidence profile below (Table 29) and the economic evidence tables in Appendix I.

One economic evaluation relating to this review question was identified but was excluded due to limited applicability.[43] This is listed in Appendix M, with reasons for exclusion given.

See also the economic article selection flow chart in Appendix F.

Table 29: Economic evidence profile: Comparisons of diagnostic tests for cirrhosis

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost-effectiveness	Uncertainty
Canavan 2013[27]	Directly applicable	Potentially serious limitations[(a)]	• Interventions considered: no testing, annual biopsy, annual TE • Study considered chronic hepatitis C patients with no fibrosis • Cost-utility analysis, Markov decision tree	Annual biopsy – no testing: £11,750 Annual TE – no testing: £3,500 Annual TE – annual biopsy: –£8,250	Annual biopsy – no testing: –1.00 Annual TE – no testing: 0.55 Annual TE – annual biopsy: 1.55	• Annual liver biopsy is dominated by 'no testing' and annual TE (more expensive and less effective) • ICER (annual TE versus no testing): £6,557 per QALY gained	Univariate sensitivity analysis; ICER most sensitive to rate of developing cirrhosis from METAVIR 3 but TE still considered cost-effective using a £30,000 threshold. Changes in other parameters do not change the cost-effectiveness

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost-effectiveness	Uncertainty
			model (3 month cycle length) • Model mainly focused on HCC as a complication				conclusions.
Steadman 2013[213]	Partially applicable[(b)]	Potentially serious limitations[(c)]	• Interventions considered: TE, liver biopsy • Three patient subgroups considered: HBV, HCV, NAFLD • Decision tree model based on the diagnostic accuracy of every test • Outcome is cost per correct diagnosis for cirrhosis	Liver biopsy versus TE: £205	Additional correct diagnoses per 1,000 people (liver biopsy versus TE): Hepatitis B: 180 Hepatitis C: 102 NAFLD: 53	Cost per additional correct diagnosis: Hep B: £1,136 (95% CI: £276–2927) Hep C: £2,001 (95% CI: £284–7317) NAFLD: £3,841 (95% CI: £288–N/A)	Changes in sensitivity, specificity and prevalence have a significant effect on the resulting cost per correct diagnosis
Stevenson 2012 [214]	Partially applicable[(d)]	Potentially serious limitations[(e)]	• Cost-utility analysis, discrete event simulation model • Model considered liver biopsy, TE, ELF and FibroTest in 6 different strategies • Study considered patients with suspected liver	Results summarised in Table 30 and Figure 3 below			There is high uncertainty in the results. This was explored with the identification of 36 scenarios for every strategy which were based on the combination of changes in 4 key parameters: liver biopsy diagnostic

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost-effectiveness	Uncertainty
			fibrosis related to alcohol consumption				accuracy, liver biopsy type (percutaneous or transjugular), NILT diagnostic accuracy, disutility level of liver biopsy

Abbreviations: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life years; TE: transient elastography; HCC: hepatocellular carcinoma

(a) Quality of life estimates do not come from a meta-analysis but from single studies. Liver biopsy unit costs low compared to other considered economic evaluations. Model did not include the polymerase inhibitor drug treatments as parameters

(b) Differences in healthcare system may make results less applicable to UK, no health outcomes were considered in the model

(c) Only cost per correct diagnosis was estimated, the time horizon is not long enough to capture all the effects

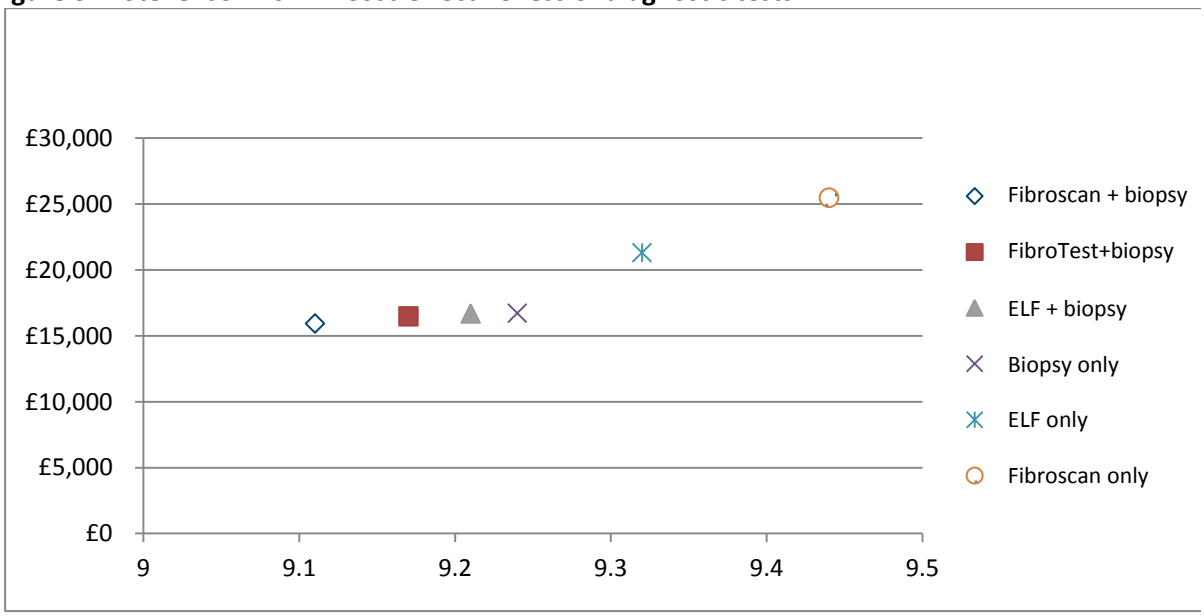
(d) Most of the quality of life values are taken from hepatitis C patients. For some health states, QALYs are based on assumptions

(e) Quality of life and test accuracy estimates do not come from a meta-analysis but from single studies, there is inconsistency between the trial data used in the model, for some tests small patient numbers lead to high uncertainty over the test accuracy, ELF did not report sensitivity and specificity for detecting only cirrhosis, results not subjected to probabilistic sensitivity analysis

Table 30: Stevenson 2012 – Cost-effectiveness results

Interventions	QALYs	Costs	ICER compared with previous best option
Fibroscan and biopsy	9.11	£15,952	–
FibroTest and biopsy	9.17	£16,472	Extendedly dominated
ELF and biopsy	9.21	£16,702	Extendedly dominated
Biopsy only	9.24	£16,730	£5,984.62
ELF only	9.32	£21,308	Extendedly dominated
Fibroscan only	9.44	£25,495	£43,825.00

Figure 3: Stevenson 2012 – Cost-effectiveness of diagnostic tests



6.4.2 Unit costs

See Table 57 in Appendix N.

6.4.3 New cost-effectiveness analysis

Original cost-effectiveness modelling was undertaken for this question using the NGC liver disease pathway model developed for this guideline. A summary is included here. Evidence statements summarising the results of the analysis can be found below. The full analysis can be found in Appendix N.

6.4.3.1 Aim and structure

The aim of the health economic modelling for this question was to determine the most cost-effective diagnostic test to detect cirrhosis in 4 aetiology groups: NAFLD, ALD, HBV and HCV. HBV patients were further separated in 2 cohorts: positive or negative hepatitis B e antigen (HBeAg). HCV patients were further separated by disease genotype (genotypes 1–4).

For these purposes a lifetime health state transition (Markov) model was constructed, following the NICE reference case,[146] which depicted the patient pathway from advanced fibrosis to liver transplantation.

The number of diagnostic strategies compared differed among the 4 examined cirrhosis aetiologies. This was related to the amount of evidence identified in the review of the diagnostic studies for each group.

Table 31: Tests included in the model by disease aetiology

Hepatitis B	Hepatitis C	ALD	NAFLD
FibroTest at 0.74	Platelet count	APRI at 1.5 – 2.5	TE at 10.0 – <13.0 kPa
Transient elastography (TE) at 11.kPa	FibroTest at a 0.56 – 0.75	TE at 11.0 – <13.0 kPa	TE at >15 kPa
APRI at 2.0	ELF at 9.3 – 10.44	TE at 15+ kPa	ARFI at 1.636 – 1.9
APRI at 1.0	APRI at 0.5 – <1.5		
	APRI at 1.5 – 2.5		
	FIB-4 at 2.3122		
	AST/ALT ratio at 1.0		
	TE at 9.0 – <13.0 kPa		
	TE at 13.0 – <15.0 kPa		
	TE at 15+ kPa		
	ARFI at 1.55 – 2.0		
	pSWE at optimal level		
	TE and ARFI (at 12.2 kPa and 1.8 m/s)		
	TE or ARFI (at 12.2 kPa and 1.8 m/s)		
	SAFE algorithm		
	Castera algorithm		

In each population group each of the diagnostic tests above were compared to the options of:

- Liver biopsy (reference standard)
- No test, monitor all patients in the relevant population assuming they have cirrhosis.
- No test, monitor no-one, assuming none have cirrhosis until later clinical presentation.

People testing negative with the test were retested using the same test every 2 years. The model used diagnostic accuracy data from studies identified in the present guideline review. When there were not enough studies (fewer than 3) around the diagnostic accuracy of a specific test for pooled sensitivity and specificity estimates, the corresponding 2×2 diagnostic table was selected from a single study that was believed to represent the best quality evidence. Test costs were obtained from published literature and GDG sources. Health states costs were constructed under GDG guidance specifically for the purposes of the model. Utilities and transition probabilities were mostly obtained from published literature and through extrapolations from other liver diseases where there was a lack of evidence (mainly in the NAFLD and ALD cohorts). The model was built probabilistically to take account of the uncertainty around input parameter point estimates.

Cost-effectiveness was defined by the value of the net monetary benefit (NMB) attributed to every test. The decision rule applied is that the comparator with the highest NMB is the most cost-effective option at the specified £20,000 per extra QALY threshold. For ALD, incremental cost-effectiveness

ratios (ICERs) comparing all strategies to 'no test – no monitoring' were also calculated due to the high uncertainty depicted in the confidence intervals.

6.4.3.2 Results

6.4.3.2.1 NAFLD cohort

Table 32: Test ranking in NAFLD cohort

Test	Mean cost (£)	Mean QALYs	NMB (£) at £20,000/QALY	Rank
TE (at 10.0 - <13.0)	19,237	9.22	165,163	3
TE (at >15.0)	19,229	9.22	165,224	1
ARFI at 1.636 - 1.9	19,275	9.22	165,212	2
Liver biopsy	22,087	9.19	161,811	6
No test – monitor all	19,929	9.23	164,614	5
No test – no monitoring	18,310	9.16	164,818	4

Across the different strategies compared, transient elastography at a threshold of 15.9 kPa ranked first mainly due to having the highest diagnostic accuracy among the non-invasive tests. ARFI followed second being slightly less accurate but also having lower test unit costs. Transient elastography at 10.0–<13.0 kPa ranked third having similar specificity to the other 2 tests but lower sensitivity. All 3 non-invasive tests had similarly wide confidence intervals (ranging from first to fourth place).

In the deterministic sensitivity analysis, rankings were sensitive to increases in the transient elastography and ARFI unit costs and in the decrease of the diagnostic accuracy of transient elastography. Therefore, no safe conclusion can be made over the most cost-effective option among the top 3 comparators.

6.4.3.2.2 ALD cohort

Table 33: Test ranking and ICERs in ALD cohort

Test	Mean cost (£)	Mean QALYs	NMB (£) at £20,000/QALY	Rank
APRI at 1.5 - 2.5	38,483	5.30	67,504	5
TE at 11.0 - <13.0	38,965	5.33	67,644	3
TE at 15+	38,947	5.33	67,616	4
Liver biopsy	42,562	5.42	65,870	6
No test – monitor all	31,163	4.97	68,321	2
No test – no monitoring	29,278	4.90	68,697	1

Testing people with alcohol-related liver disease for cirrhosis was not cost-effective compared to 'no test – no monitoring' and 'no test – monitor all' at a cost-effectiveness threshold of £20,000 per QALY gained. However, it was cost-effective at a threshold of £30,000 per QALY gained: the ICERs for the 3 non-invasive liver tests were £22,438–£22,977). All three non-invasive tests had similarly wide confidence intervals (from first or second to fifth place).

In none of the deterministic sensitivity analysis scenarios did a test strategy rank higher than third. Ranking among the 3 non-invasive liver tests slightly varied across the different scenarios with transient elastography at 11.0 - <13.0 remaining third in ranking for 9 out of the 10 tested scenarios.

6.4.3.2.3 HBV cohorts

Table 34: Test ranking in HBeAg negative cohort

Test	Mean cost (£)	Mean QALYs	NMB (£) at £20,000/QALY	Rank
FibroTest at 0.74	63,375	8.33	103,210	3
TE at 11.kPa	63,583	8.34	103,237	2
APRI at 2.0	63,317	8.32	103,070	4
APRI at 1.0	63,521	8.34	103,281	1
Liver biopsy	65,820	8.32	100,612	7
No test – monitor all	64,096	8.35	102,849	6
No test – no monitoring	62,552	8.27	102,904	5

For the HBeAg negative group, APRI at 1.0 ranked first, most probably due to its low test unit costs and its moderate diagnostic accuracy (second best after transient elastography). Transient elastography and FibroTest ranked second and third. APRI at 2.0 ranked last among the NILT mainly due to its considerably lower sensitivity. All non-invasive liver tests had similarly wide 95% confidence intervals.

Table 35: Test ranking in HBeAg positive cohort

Test	Mean cost (£)	Mean QALYs	NMB (£) at £20,000/QALY	Rank
FibroTest at 0.74	42,758	10.06	158,366	1
TE at 11.kPa	42,927	10.06	158,362	2
APRI at 2.0	42,827	10.05	158,155	5
APRI at 1.0	42,916	10.06	158,358	3
Liver biopsy	45,997	10.03	154,533	7
No test – monitor all	43,527	10.07	157,851	6
No test – no monitoring	42,013	10.02	158,328	4

In the HBeAg positive group, FibroTest ranked first with TE and APRI at 1.0 ranking second and third. All non-invasive liver tests had similarly wide 95% confidence intervals. In the probabilistic analysis, the 3 tests also shared similar probabilities ranking first (20–23%).

Deterministic sensitivity analyses were conducted for the HBeAg negative group. Rankings between the deterministic and the probabilistic analyses varied particularly for the FibroTest and transient elastography tests highlighting how incorporating the uncertainty of the input parameters in the model affects the cost-effectiveness results. APRI at 1.0 ranked first or second in all scenarios. FibroTest and transient elastography each ranked between first and fourth in each scenario. The cost-effectiveness of APRI at 1.0 was sensitive to the decrease of HBV prevalence, the presence of varices at the point of cirrhosis diagnosis and changes to the cost and the accuracy of transient elastography.

6.4.3.2.4 HCV cohorts

Results are only presented for genotypes 1 and 3 as the results for genotypes 2 and 4 were consistent with these. The rankings of the top 3 tests are presented for all 4 genotypes.

Table 36: Test ranking in HCV genotype 1 cohort

Test	Mean cost (£)	Mean QALYs	NMB (£) at £20,000/QALY	Rank
Platelet count	30,936	12.20	213,159	11

Test	Mean cost (£)	Mean QALYs	NMB (£) at £20,000/QALY	Rank
FibroTest at 0.56–0.75	32,666	12.17	210,760	15
ELF at 9.3–10.44	31,522	12.19	212,285	13
APRI at 0.5 – <1.5	31,589	12.20	212,445	12
APRI at 1.5–2.5	31,827	11.90	206,075	16
FIB-4 at 2.3122	31,877	12.17	211,589	14
AST/ALT ratio at 1.0	31,677	11.86	205,528	17
TE at 11.0 – <13.0	30,268	12.20	213,734	10
TE at 13.0 – <15.0	29,417	12.25	215,580	2
TE at 15+	30,170	12.20	213,822	8
ARFI at 1.55–2.0	30,737	12.23	213,769	9
pSWE (optimal cut-off)	30,348	12.21	213,868	7
TE+ARFI (12.2 kPa and 1.8 m/s)	29,747	12.20	214,276	5
TE or ARFI (12.2 kPa or 1.8 m/s)	30,589	12.25	214,444	4
SAFE algorithm	30,378	12.21	213,902	6
Castera algorithm	29,140	12.22	215,251	3
Liver biopsy	28,762	12.26	216,472	1
No testing – monitor all	39,699	12.17	203,774	18
No testing – no monitoring	32,505	11.44	196,274	19
No testing – no monitoring or treatment	18,149	8.36	149,055	20

Table 37: Test ranking in HCV genotype 3 cohort

Test	Mean cost (£)	Mean QALYs	NMB (£) at £20,000/QALY	Rank
Platelet count	17,990	11.84	218,732	11
FibroTest at 0.56–0.75	21,494	11.79	214,391	15
ELF at 9.3–10.44	19,120	11.82	217,245	13
APRI at 0.5 - <1.5	19,453	11.84	217,316	12
APRI at 1.5–2.5	18,782	11.40	209,184	16
FIB-4 at 2.3122	20,040	11.80	215,996	14
AST/ALT ratio at 1.0	18,219	11.33	208,441	17
TE at 11.0 – <13.0	16,320	11.82	220,067	8
TE at 13.0 – <15.0	14,334	11.88	223,199	3
TE at 15+	16,049	11.82	220,294	6
ARFI at 1.55–2.0	17,528	11.86	219,663	10
pSWE (optimal cut-off)	16,597	11.83	220,036	9
TE+ARFI (12.2 kPa and 1.8 m/s)	14,895	11.81	221,326	4
TE or ARFI (12.2 kPa or 1.8 m/s)	17,256	11.89	220,577	5
SAFE algorithm	16,546	11.84	220,227	7
Castera algorithm	13,110	11.82	223,277	2
Liver biopsy	11,759	11.87	225,611	1
No testing – monitor all	36,657	11.75	198,272	18
No testing – no monitoring	18,724	10.69	195,174	19

Test	Mean cost (£)	Mean QALYs	NMB (£) at £20,000/QALY	Rank
No testing – no monitoring or treatment	18,164	8.36	149,001	20

Table 38: HCV diagnostic tests – top 3 ranked tests in every genotype

Rank	Genotype 1	Genotype 2	Genotype 3 [(a)]	Genotype 4
First	Liver biopsy	Liver biopsy	Liver biopsy	Liver biopsy
Second	TE at 13.0<15.0	Castera algorithm	Castera algorithm	TE at 13.0<15.0
Third	Castera algorithm	TE at 13.0<15.0	TE at 13.0<15.0	TE or ARFI (12.2 kPa or 1.8 m/s)

(a) For genotype 3, the Castera algorithm and TE at 13.0<15.0 had almost identical NMBs

For all 4 genotypes, liver biopsy ranked first with substantially higher NMB values compared to the second options. This is mainly attributable to the fact that liver biopsy was assumed to have perfect sensitivity and specificity, and that cirrhosis misdiagnosis is associated with the incorrect administration of the highly costly polymerase inhibitor drugs. This led to the economic model particularly favouring the test with the highest diagnostic accuracy irrespective of its unit cost. In genotypes 1 and 3, liver biopsy ranked first in 90% and 97% of the simulations respectively. Transient elastography at 13.0–<15.0 ranked second or third in genotypes 1–4 and the 'TE or ARFI' strategy ranked third in genotype 4.

Deterministic sensitivity analyses were conducted for the genotype 3 group. Liver biopsy remained first in all but 2 scenarios. These were the 'no HCV treatment' and the 'diagnostic accuracy for transient elastography at 13.0 kPa at high 95% CI' scenarios, also highlighting how crucial the drug treatment element is for the HCV diagnostic model.

For more details on all the analyses see Appendix N.

6.5 Evidence statements

6.5.1 Clinical

- Fifty-three studies were included in the review covering 5 aetiologies of cirrhosis. Thirty-three studies looking at the hepatitis C population, 10 studies looked at the NAFLD or NASH population, 3 studies looked at the HCV/HIV co-infected population, 1 study looked at the PBC population and 2 studies looked at ALD. No evidence was identified for the population stratum of PSC specified in the protocol.
- Of the index tests listed in the protocol, no evidence was identified for albumin, prothrombin time (INR), bilirubin, γGT, ultrasound or MR elastography.
- Studies were identified relating to the accuracy of platelet count, AST, ALT, FibroTest, ELF, APRI, FIB-4, AST/ALT ratio, transient elastography, ARFI and combination of these tests in diagnosing cirrhosis. Data presented to the GDG were in the form of paired sensitivity and specificity values and AUC values. Data relating to transient elastography was reported at a range of thresholds: low 9-<13 kPa, medium 13 to <15 kPa, high ≥15 kPa. Similarly, data relating to APRI was divided into low (0.5 to <1.5) and high threshold ranges (1.5–2.5).

Hepatitis C

- In the hepatitis C population, Moderate quality evidence from 5 studies indicated a high sensitivity (76 and 87), a high specificity (84 and 88) and a high AUC (range 82.7–89.9) for platelet count.

- Moderate quality evidence from 1 study indicated a high AUC (75.2) for AST.
- Moderate quality evidence from 1 study indicated a moderate AUC (62.6) for ALT.
- Very Low quality evidence from 4 studies indicated a high sensitivity (80.3) and a moderate specificity (69.3) for FibroTest. Moderate quality evidence from 4 studies indicated a high AUC value (86.5) for FibroTest.
- Low quality evidence from 3 studies indicated a high sensitivity (83.0) and specificity (82.0) for ELF.
- Very Low quality evidence from 7 studies indicated a high sensitivity (83.8) and specificity (77.8) for APRI at low cut-offs. Very Low quality evidence from 5 studies indicated a low sensitivity (36.5) and high specificity (94.4) for APRI at high cut-offs. Low quality evidence from 8 studies indicated a high AUC 88.0 for APRI.
- Very Low quality evidence from 1 study indicated a high sensitivity (80) and high specificity (78) for FIB-4. Low quality evidence from 4 studies indicated a high AUC (84.8) for FIB-4.
- Low quality evidence from 2 studies indicated a low sensitivity (30 and 35) and high specificity (90 and 97) for AST/ALT ratio. Moderate quality evidence from 3 studies indicated a high AUC (73) for AST/ALT ratio.
- Low quality evidence from 7 studies indicated a high sensitivity (81.5) and specificity (90.4) for transient elastography at a low cut-off. Very Low quality evidence from 7 studies indicated a high sensitivity (93.4) and a high specificity (92.9) for transient elastography at medium thresholds. Very Low quality evidence from 1 study indicated high sensitivity (86) and specificity (91) of transient elastography at a high threshold. Low quality evidence from 10 studies indicated a high AUC (92.6) for transient elastography.
- Very Low quality evidence from 6 studies indicated a high sensitivity (88.1) and specificity (84.3) for ARFI. Low quality evidence from 2 studies indicated a high AUC (86.1) for ARFI.
- Low quality evidence from 1 study indicated a high sensitivity (90) and specificity (89) for pSWE. Low quality evidence from 1 study indicated a high AUC (95) for pSWE.
- Low quality evidence from 1 study indicated a high sensitivity (85) and specificity (94) for transient elastography plus ARFI. Very Low quality evidence from 1 study indicated a high sensitivity (96) and specificity (83) for transient elastography or ARFI.
- Moderate quality evidence indicated a high sensitivity (87), high specificity (90) and AUC (87) of the SAFE algorithm (sequential use of APRI, FibroTest and liver biopsy).
- Moderate quality evidence indicated a high sensitivity (89) and specificity (98) of the Castera algorithm (combination of transient elastography and FibroTest. When transient elastography and FibroTest agree no biopsy is performed whereas when they disagree, liver biopsy is needed). Low quality evidence from 1 study indicated a high AUC (93) for the Castera algorithm.

NAFLD

- In the NAFLD population, Very Low quality evidence from 2 studies indicated a moderate sensitivity (78 and 76) and high specificity (95 and 91) for transient elastography at a low cut-off, and a high sensitivity (100 and 100) and specificity (97 and 98) for transient elastography at a high cut-off.
- Very Low quality evidence from 2 studies indicated a high sensitivity (92 and 100) and specificity (92 and 96) for ARFI.
- Low quality evidence indicated a moderate AUC for APRI (76.8), FIB-4 (81) and AST/ALT ratio (73.7).
- There was a high AUC for transient elastography (94) and ARFI (98.4) from Very low and Low quality evidence, respectively.

ALD

- In the ALD population, Very Low quality evidence from 1 study indicated a low sensitivity (40) and moderate specificity (61) for APRI at a high cut-off.
- Very Low quality evidence from 1 study indicated a high sensitivity (100) and specificity (79) of transient elastography at a low cut-off. Very Low quality evidence from 1 study indicated a high sensitivity (80) and specificity (76) of transient elastography at a high cut-off. Very Low quality evidence from 1 study indicated a high AUC for transient elastography (92.1).

Primary biliary cholangitis (PBC)

- In the PBC population, Low quality evidence from 1 study indicated a high sensitivity (100) and specificity (94) of transient elastography at a low cut-off.
- Low quality evidence from 1 study indicated a high AUC (84) of APRI.
- Moderate quality evidence from 1 study indicated a high AUC (74) of FIB-4.
- Low quality evidence from 1 study indicated a moderate AUC (58) for AST/ALT ratio.
- Low quality evidence from 1 study indicated a good AUC (99) for transient elastography.
- Low quality evidence from 1 study indicated high AUC values (99) for 3 combinations of tests (TE plus APRI, TE plus FIB-4 and TE plus AST/ALT ratio).

HCV/HIV co-infection

- In the HCV/HIV co-infected population, Very Low quality evidence from 1 study indicated a moderate sensitivity (63) and a high specificity (77) for platelet count. Low quality evidence from 2 studies indicated a high AUC (79.5) for platelet count.
- Low quality evidence from 1 study indicated a high sensitivity (78) and a moderate specificity (57) for APRI at a low cut-off. Very Low quality evidence from 1 study indicated a low sensitivity (53) and a high specificity (89) for APRI at a high cut-off. Low quality evidence from 2 studies indicated a high AUC (77.5) for APRI.
- Moderate quality evidence from 1 study indicated a high AUC (73) for FIB-4.
- Very Low quality evidence from 1 study indicated a low sensitivity (38) and a high specificity (77) for AST/ALT ratio. Very Low quality evidence from 2 studies indicated a low AUC (52.5) for AST/ALT ratio.
- Low quality evidence from 1 study indicated a high sensitivity (100) and specificity (93) for transient elastography at a low cut-off. Very Low quality evidence from 2 studies indicated high (88 and 100) sensitivity and specificity (93 and 96) for transient elastography at medium thresholds. Low quality evidence from 2 studies indicated a high AUC (80) for transient elastography.

6.5.2 Economic

- One cost-utility analysis that compared annual liver biopsy, annual transient elastography and no testing for diagnosis of cirrhosis in chronic hepatitis C patients found that:
  - annual transient elastography was cost-effective compared to no testing (ICER: £6,557 per QALY gained)
  - annual liver biopsy was dominated by both alternatives (more expensive and less effective).

  This analysis was assessed as directly applicable with potentially serious limitations.
- One cost analysis that compared liver biopsy and transient elastography for diagnosis of cirrhosis in 3 relevant patient subgroups found that liver biopsy had additional costs of £1,136, £2,001 and £3,841 per additional correct diagnosis when compared to transient elastography for the HBV,

HCV and NAFLD subgroups respectively. This analysis was assessed as partially applicable with potentially serious limitations.

- One cost-utility analysis that compared 6 diagnostic strategies for diagnosis of cirrhosis in adults with ALD found that liver biopsy was cost-effective at a cost-effectiveness threshold of £20,000 per QALY gained compared to the following strategies:
  - triage with transient elastography, biopsy in those who tested positive with transient elastography
  - triage with FibroTest, biopsy in those who tested positive with FibroTest
  - triage with ELF, biopsy in those who tested positive with ELF
  - transient elastography alone
  - ELF alone.

  This analysis was assessed as partially applicable with potentially serious limitations.
- One original cost-utility analysis that compared 6 strategies to diagnose cirrhosis in people with NAFLD and advanced fibrosis with a retest frequency of 2 years found that transient elastography ranked first compared to the following diagnostic strategies, using relevant thresholds for each test, with reference to a cost-effectiveness threshold of £20,000 per QALY gained:
  - ARFI
  - transient elastography (lower threshold)
  - no test – no surveillance
  - no test – surveillance for all
  - liver biopsy.

  This analysis was assessed as directly applicable with minor limitations.
- One original cost-utility analysis that compared 6 strategies to diagnose cirrhosis in people with ALD, with a retest frequency of 2 years, found that:
  - The 'no test – no surveillance' strategy ranked first compared to the following diagnostic strategies, using relevant thresholds for each test, with reference to a cost-effectiveness threshold of £20,000 per QALY gained:
    - no test – surveillance for all
    - transient elastography (low threshold)
    - transient elastography (high threshold)
    - APRI
    - liver biopsy.
  - When compared to the 'no test – no monitor' strategy, the 3 non-invasive tests had ICERs between £22,438 and £22,977 per QALY gained.

  This analysis was assessed as directly applicable with minor limitations.
- One original cost-utility analysis that compared 7 strategies to diagnose cirrhosis in people with hepatitis B and HBeAg negative with a retest frequency of 2 years found that APRI ranked first compared to the following diagnostic strategies, using relevant thresholds for each test, with reference to a cost-effectiveness threshold of £20,000 per QALY gained:
  - transient elastography
  - FibroTest
  - APRI (higher threshold)
  - no test – no surveillance
  - no test – surveillance for all
  - liver biopsy.

  This analysis was assessed as directly applicable with minor limitations.

- One original cost-utility analysis that compared 7 strategies to diagnose cirrhosis in people with hepatitis B and HBeAg positive with a retest frequency of 2 years found that FibroTest ranked first compared to the following diagnostic strategies, using relevant thresholds for each test, with reference to a cost-effectiveness threshold of £20,000 per QALY gained:
  - transient elastography
  - APRI (low threshold)
  - no test – no surveillance
  - APRI (high threshold)
  - no test – surveillance for all
  - liver biopsy.

  This analysis was assessed as directly applicable with minor limitations.
- One original cost-utility analysis that compared 20 strategies to diagnose cirrhosis in people with hepatitis C with a retest frequency of 2 years found that liver biopsy ranked first compared to the following diagnostic strategies, using relevant thresholds for each test, with reference to a cost-effectiveness threshold of £20,000 per QALY gained:
  - Castera algorithm
  - transient elastography (medium threshold)
  - transient elastography and ARFI
  - transient elastography or ARFI
  - transient elastography (high threshold)
  - SAFE algorithm
  - point shear wave elastography
  - transient elastography (low threshold)
  - ARFI
  - platelet count
  - APRI
  - ELF
  - FIB-4
  - FibroTest
  - APRI
  - AST-ALT ratio
  - no testing – surveillance for all, treat HCV using medication for people with cirrhosis
  - no testing – no surveillance, treat HCV using medication for people with fibrosis
  - no testing – no surveillance, no treatment for HCV.

  This analysis was assessed as directly applicable with minor limitations.

6.6 Recommendations and link to evidence

Recommendations	**2. Discuss with the person the accuracy, limitations and risks of the different tests for diagnosing cirrhosis.** **3. Offer transient elastography to diagnose cirrhosis for:** • **people with hepatitis C virus infection** • **men who drink over 50 units of alcohol per week and women who drink over 35 units of alcohol per week and have done so for several**

	months • **people diagnosed with alcohol-related liver disease.** **4. Offer either transient elastography or acoustic radiation force impulse imaging (whichever is available) to diagnose cirrhosis for people with NAFLD and advanced liver fibrosis (as diagnosed by a score of 10.51 or above using the enhanced liver fibrosis [ELF] test). Also see the assessment for advanced liver fibrosis section in NICE's NAFLD guideline.** **5. Consider liver biopsy to diagnose cirrhosis in people for whom transient elastography is not suitable.** **6. For recommendations on diagnosing cirrhosis in people with hepatitis B virus infection, see the assessment of liver disease in secondary specialist care section in NICE's hepatitis B (chronic) guideline.** **7. Do not offer tests to diagnose cirrhosis for people who are obese (BMI of 30 kg/m^2 or higher) or who have type 2 diabetes, unless they have NAFLD and advanced liver fibrosis (as diagnosed by a score of 10.51 or above using the ELF test). Also see the assessment for advanced liver fibrosis section in NICE's NAFLD guideline.** **8. Ensure that healthcare professionals who perform or interpret non-invasive tests are trained to do so.** **9. Do not use routine laboratory liver blood tests to rule out cirrhosis.** **10. Refer people diagnosed with cirrhosis to a specialist in hepatology.** **11. Offer retesting for cirrhosis every 2 years for:** • **people diagnosed with alcohol-related liver disease** • **people with hepatitis C virus infection who have not shown a sustained virological response to antiviral therapy** • **people with NAFLD and advanced liver fibrosis.** **12. For recommendations on reassessing liver disease in hepatitis B virus infection, see the assessment of liver disease in secondary specialist care section in NICE's hepatitis B (chronic) guideline.**
Relative values of different outcomes	The GDG was interested in the performance of various blood or imaging tests in the diagnosis of cirrhosis in people with risk factors for cirrhosis even in the absence of signs and symptoms (for example HCV, NAFLD, alcohol misuse). The GDG did not consider the performance of these tests as screening tools in the general population. Therefore, test performance was assessed from studies matching the intended population for use of the test clinically. Studies including healthy populations without suspected chronic liver disease were not included. Due to existing NICE guidance for assessment of fibrosis and cirrhosis in people with HBV, new clinical evidence was not searched for. In order to assess test performance we first searched for any diagnostic RCTs assessing patient outcomes, from studies randomising patients to diagnoses using one test or another, followed by identical therapeutic interventions based on the results of the tests. This is seen as the gold standard study design as it assesses

	patient outcomes as clinically important consequences of diagnostic accuracy. No studies of this design were identified from the clinical evidence review. Therefore, the GDG reviewed evidence from diagnostic accuracy studies. The reference standard test used to define the presence or absence of cirrhosis was liver biopsy. The GDG specified in the protocol the most commonly used biopsy scoring systems for cirrhosis, including Knodell F4, Ishak F5/6, METAVIR F4 or, for NAFLD populations, the Kleiner or Brunt F4 scoring systems. For decision-making, the GDG focused on diagnostic accuracy measures including the sensitivity and specificity of the tests for a diagnosis of cirrhosis. It was noted that these data were used to inform the health economic model, in order to identify the most cost-effective test, or combination of tests, for the diagnosis of cirrhosis. The GDG discussed that, for a condition such as cirrhosis where early identification is essential for effective management (including treatment of the underlying cause or monitoring for life threatening complications), it is crucial to have a highly sensitive test, especially early on in the patient pathway if multiple tests are used. This is because a sensitive test will result in very few people with cirrhosis being missed (few false negative results). The GDG noted that the cut-off threshold used to define a positive test can vary and assessed the accuracy of the tests at a variety of published thresholds. A threshold set to increase the sensitivity of the test will consequently reduce the specificity. The GDG also discussed the importance of a test with high specificity, which would result in very few people without cirrhosis being incorrectly labelled with cirrhosis (false positive results). This is particularly important if the results of the test determine people who would then possibly undergo an invasive or costly intervention.
Trade-off between clinical benefits and harms	**Hepatitis C** The majority of available clinical evidence meeting the protocol criteria was in populations of people with chronic HCV infection. The only data available for individual blood tests were for platelets with a sensitivity ranging from 76% to 87% and a specificity ranging from 84% to 88%. The GDG noted the lack of any evidence to support ALT or AST as individual blood tests in the diagnosis of cirrhosis. This would support the fact that a diagnosis of HCV-related cirrhosis should not be discounted on the basis of these individual blood tests alone. Data were available for the accuracy of AST/ALT as a ratio measure, providing evidence of a very low sensitivity. Therefore, despite a high specificity, AST/ALT ratio would not be a very good first-line test as there would be a high number of false negative results and people with cirrhosis would be missed. However, the GDG discussed whether it would be an option to combine such a test with a highly sensitive test. For blood fibrosis tests there were data available for FibroTest, ELF, APRI and FIB-4. The GDG noted a relatively high sensitivity and specificity for all these blood fibrosis tests. For imaging tests, accuracy data were available for transient elastography and ARFI. Both had a relatively high sensitivity and specificity, with transient elastography at a cut-off threshold of between 13 kPa and 15 kPa performing the best (pooled sensitivity and specificity of 93.4% [95% CI 87.9, 97.0] and 92.9% [95% CI 86.5, 97.0], respectively). The GDG also noted that the tests using transient elastography and ARFI in combination (a positive result on both, or a positive result on one or the other) also gave relatively high sensitivities and specificities. One study also assessed transient elastography within an algorithm of tests, the best performing of which was the Castera algorithm consisting of a combination of transient elastography and FibroTest with a sensitivity of 89% and a specificity of 98%. **NAFLD** No relevant studies were identified looking at either individual blood tests or blood fibrosis tests in a NAFLD population. For imaging tests, accuracy data were available for transient elastography and ARFI. The GDG noted the anomalous results seen for

transient elastography in the NAFLD population. Normally, increasing the threshold of a test will result in a decrease in the sensitivity and an increase in the specificity. This pattern was not always observed in the evidence. This may be due to the differing aims of the studies included in the clinical evidence review. Some studies aimed to assess the diagnostic accuracy of transient elastography in a NAFLD population, using the most appropriate probe size (M or XL probe) for each individual's BMI, as per the manufacturer's instructions. Other studies aimed to compare the accuracy of the M probe with the XL probe, assessing both probes in each individual regardless of their BMI. It was agreed that the latter type of study was not appropriate for the clinical evidence review, as the GDG were interested in the overall accuracy of transient elastography in this population, with the assumption that the test is performed according to the manufacturer's instructions using the most appropriate probe in each patient. Therefore, the following studies were removed from the analysis: Wong 2012, Myers 2012 and Friedrichrust 2010A.

For the overall accuracy of transient elastography in the NAFLD population, Wong 2010B and Gaia 2011 were available for the lower threshold range and showed a sensitivity of 76–78% and a specificity of 91–95%. Yoneda 2008 and Yoneda 2010 were available for the higher threshold range and showed a sensitivity of 100% and a specificity of 97–98%. No studies were available for the medium threshold range. The GDG noted that the higher thresholds had an unexpectedly high sensitivity and discussed that this may be due to the smaller study size in comparison to the studies included for the lower threshold range. There was also a low prevalence of cirrhosis in these studies and it was noted that even a single diagnosis in either direction would impact considerably on the performance variables. In practice, the manufacturer suggest a threshold of around 10–11 kPa for diagnosis of cirrhosis in this population. The GDG discussed that it is often difficult to interpret the transient elastography reading in this population, but that the introduction of the XL probe has helped.

Data were available from 2 studies for the accuracy of ARFI and both studies showed similarly high sensitivity and specificity of ARFI. Overall, this was higher than that of transient elastography in the NAFLD population.

ALD

No relevant studies were identified looking at individual blood tests in a population with ALD. The GDG discussed the limitations of tests using AST in this population, as people with high alcohol consumption may have a raised AST. For blood fibrosis tests, there were only data available for APRI which showed a low sensitivity of 40% and specificity of 61% in this population. Transient elastography proved to be accurate in this population at both the lower threshold range and the higher threshold range. The lower threshold range gave a very high sensitivity of 100%, but the GDG noted the wide confidence intervals for all the accuracy measures in this population, perhaps due to the very small sample sizes of the studies. The GDG agreed that transient elastography should be included in the modelling for this population, but that the wide confidence intervals in the estimate should be reflected in the sensitivity analyses. It was discussed that for both blood fibrosis tests and imaging tests, care needs to be taken when interpreting results in people who are actively drinking. Alcohol consumption per se may alter the circulating blood levels of the individual markers irrespective of the degree of hepatic fibrosis. In addition, active alcohol consumption causes swelling and protein retention in liver cells which increases liver stiffness and so imaging tests may overestimate the degree of fibrosis. The presence of steatosis and inflammation in the liver has a similar effect and consequences. Although diagnostic tests should be performed at the point of first contact, the tests should be repeated after a period of abstinence in this population subgroup.

	Primary biliary cholangitis (PBC) In the PBC population, data were only available for the accuracy of transient elastography, which showed a high sensitivity and specificity. The GDG noted that this population would not be modelled, due to the limited data available. **HIV and HCV** The only evidence identified for the accuracy of the tests in people with multiple aetiologies was in people with HIV and HCV co-infection. The only data available for individual blood tests were for platelets which had a sensitivity of 63% and a specificity of 77%. For blood fibrosis tests there were data available for APRI and AST/ALT ratio. The GDG noted the poor sensitivity of both AST/ALT ratio and APRI using the high cut-off threshold, but the improved performance of APRI using a lower cut-off threshold. Transient elastography showed a very high sensitivity and specificity at both the low and medium threshold ranges. The limited number of patients and the low prevalence of cirrhosis were noted as these will affect the accuracy data.
Trade-off between net clinical effects and costs	Three relevant published economic evaluations were identified for this review. Canavan 2013 was a cost-utility analysis that compared annual liver biopsy, annual transient elastography and a no testing strategy in a cohort of patients with chronic hepatitis C. It found that liver biopsy was dominated by both alternatives (more expensive and less effective) and that annual transient elastography was cost-effective (ICER: £6,557 per QALY gained) when compared to no testing. The GDG noted that although the model structure was considered representative of the condition it lacked the inclusion of the new polymerase inhibitor treatments as a model parameter. Steadman 2013 was a cost analysis that compared liver biopsy and transient elastography in 3 relevant patient subgroups. It found that liver biopsy had additional costs of £1,136, £2,001 and £3,841 per additional correct diagnosis when compared to transient elastography for the HBV, HCV and NAFLD groups respectively. The GDG could not reach a conclusion on whether these additional costs are a cost-effective price per correct diagnosis as the study did not take into account any further follow-up health costs or savings related to every test result. Additional limitations of this study included the use of observational studies to determine the accuracy of transient elastography. Stevenson 2012 was a cost-utility analysis that compared 10 strategies (of which 6 were relevant to this review) in patients with alcohol-related liver disease. The model comparators included triage with transient elastography, FibroTest or ELF with a liver biopsy as confirmation test to those positive in the first test, and transient elastography, ELF or liver biopsy in single-test strategies. The study found that only liver biopsy was cost-effective at a £20,000 per QALY gained threshold. The GDG noted the presence of multiple limitations in this analysis since most quality of life values were obtained from a HCV cohort and some QALYs were based on assumptions. In addition, results were not subjected to probabilistic sensitivity analysis. Original cost-effectiveness analysis was conducted for this guideline to address the cost-effectiveness of diagnostic tests for cirrhosis in adults with HBV, HCV, NAFLD and ALD. **Hepatitis C** This analysis found that liver biopsy was the most cost-effective test of the 17 tests compared in all 4 genotypes at a cost-effectiveness threshold of £20,000 per QALY gained. In second and third places were transient elastography at 13.0<15.0

threshold, the Castera algorithm or a combination of transient elastography (at 12.2 kPa) and ARFI (at 1.8 m/s) (testing positive for either test), depending on the genotype. There was minimal uncertainty in the ranking of liver biopsy with it ranking first in more than 90% of the simulations. There was moderate to high uncertainty in the rankings of the remaining tests.

The GDG acknowledged that this result is mainly attributed to the fact that cirrhosis misdiagnosis is associated with the incorrect administration of the highly costly polymerase inhibitor drugs. As a result, the economic model seemed to particularly favour the test with the highest diagnostic accuracy irrespective of its unit cost. The GDG also noted that the diagnostic accuracy of liver biopsy used in the model was set to 100% sensitivity and specificity since it served as the reference standard for the test comparisons. Therefore any model bias regarding the diagnostic accuracies of the tests is in favour of liver biopsy.

Hepatitis B

This analysis, using clinical effectiveness data from the clinical review conducted for NICE CG165,[141] found that in HBeAg negative patients, APRI at 1.0 was the most cost-effective test of the 5 tests compared at a cost-effectiveness threshold of £20,000 per QALY gained, with transient elastography at 11.0 and FibroTest at 0.74 very close behind in second and third place respectively. For HBeAg positive patients, FibroTest at 0.74, transient elastography at 11.0 and APRI at 1.0 ranked in the first 3 positions with very similar cost-effectiveness figures at a cost-effectiveness threshold of £20,000 per QALY gained. There was considerable uncertainty in the results with all the non-invasive tests having wide confidence intervals and the 3 strategies listed all having the first ranking place within their confidence intervals.

NAFLD

This analysis found that in patients with NAFLD, transient elastography at >15.0 was the most cost-effective test of the 4 tests compared at a cost-effectiveness threshold of £20,000 per QALY gained, with ARFI at 1.636–1.9 in second place. There was considerable uncertainty in the results with the 3 non-invasive test strategies having similarly wide confidence intervals (first to fourth place).

ALD

This analysis found that in patients with ALD, none of the diagnostic tests was cost-effective at a cost-effectiveness threshold of £20,000 per QALY gained, with 'no test – no monitoring' and 'no test – monitor all' ranking first and second. Incremental cost-effectiveness ratios that compared the 3 non-invasive tests with the 'no test – no monitor' strategy were, however, only just beyond the £20,000 threshold (£22,438–22,977), meaning that testing was cost-effective at a threshold of £30,000 per QALY gained. There was considerable uncertainty in the cost-effectiveness rankings, with all strategies but liver biopsy (consistently ranked last) having similarly wide confidence intervals (first or second to fifth place).

Conclusions

After taking into account the overall cost-effectiveness results of the original analysis for all 4 examined populations, the GDG acknowledged that there is significant variation in the cost-effectiveness of the diagnostic tests across the different aetiologies. The economic model suggested the use of a non-invasive test for NAFLD and HBV, the use of liver biopsy for patients with HCV and no testing for ALD patients. However, for the ALD cohort, the GDG concluded that testing is an appropriate strategy, as this cohort comprises the largest group of people with

	cirrhosis, and the population which has the highest risk of dying from cirrhosis. The GDG noted that defining who is most at risk of cirrhosis due to alcohol use is difficult due to the lack of a universally agreed definition, but that the criterion of exceeding 50 units per week for men or 35 units for women, constituting 'harmful drinking', was the highest threshold that could be set and thus the population selected will be at higher risk of cirrhosis than if a lower threshold of alcohol use had been chosen. The base case ICER for transient elastography compared to no testing was £22,438 per QALY gained, which is only slightly above a cost-effectiveness threshold of £20,000 per QALY. The range of the confidence interval around this base case showed that testing could be below £20,000 per QALY within the range of uncertainty, but could not be above £30,000 per QALY. One source of uncertainty is the effect of cirrhosis diagnosis on drinking behaviour. There are no clear data in this area, but if the diagnosis of cirrhosis has the effect of substantially increasing a person's likelihood of abstaining from or reducing consumption of alcohol, in addition to the positive effects on health of treating the cirrhosis, then this would make testing people with ALD for cirrhosis substantially more cost-effective. As for the selection of the most appropriate non-invasive cirrhosis test, the GDG noted the practicality of recommending a common test for all aetiologies, and that there is an existing recommendation for people with hepatitis B in CG165. Taking these factors into account, the GDG recognised that there was adequate evidence across all aetiologies to conclude that transient elastography (at the appropriate threshold for each aetiology) is a cost-effective option for the diagnosis of cirrhosis irrespective of the underlying cause since, after taking combined model parameter uncertainty into consideration, transient elastography could rank first in 3 out of the 4 examined cirrhosis aetiologies. For the hepatitis B cohort the GDG noted the similarcost-effectiveness of APRI and transient elastography. Transient elastography has been recommended in CG165 for assessing the stage of liver disease in people with hepatitis B, and these results are consistent with that recommendation with regard to testing for cirrhosis. For the NAFLD cohort, the GDG felt there was too much uncertainty in the model results to exclusively recommend transient elastography since ARFI exhibited similar cost-effectiveness and is similar in its availability, cost and ease of use. Some centres currently have access to transient elastography and others to ARFI, and there is no reason why whichever technique is most easily available should not be used. The GDG acknowledged that liver biopsy ranked as the most cost-effective option modelled for people with hepatitis C, principally due to the very high price of drugs to treat hepatitis C and thus the high cost of misdiagnosis. However, this assumes that liver biopsy has 100% sensitivity and specificity. Although regarded as the reference standard, it is acknowledged that liver biopsy does not have perfect sensitivity and specificity and does misclassify some people. Without a more objective test to compare it against this misclassification cannot be quantified, but the lower the quality of the liver biopsy (determined by the number and length of the samples taken), the greater the risk of misclassification. This does however mean that the results of the economic model are biased slightly in favour of liver biopsy. In addition, the invasive nature of liver biopsy means that it causes adverse events, including a small risk of death. It is also considered unpleasant by patients, leading to a very low acceptability among patients. The GDG considered that if liver biopsy was the only option offered to people with hepatitis C then a large majority of patients would refuse any testing and so not be diagnosed with (or without) cirrhosis. Diagnosis of cirrhosis status is also required to determine the correct drugs to use to treat hepatitis C itself, and therefore if cirrhosis status cannot be determined then hepatitis C cannot be treated. Recommending only liver biopsy would therefore be likely to cause a severe negative impact on people with hepatitis C. The GDG also highlighted the fact that many people have multiple aetiologies (for example hepatitis C and ALD) and that recommendations should have some consistency across the different aetiologies for this reason.

	However, liver biopsy should be permitted if people with hepatitis C wish to choose it, aware of the risks and benefits. Therefore the GDG recommended that liver biopsy should be considered in hepatitis C patients when transient elastography is not suitable, for example if the person opts for liver biopsy given an informed choice. For people with NAFLD and for people suspected for ALD or drinking at harmful levels, liver biopsy is not the most cost-effective option. It is expected that most of these people will prefer and choose transient elastography. However, liver biopsy remains the most accurate and authoritative test and may be appropriate in some cases, including when transient elastography cannot be successfully used. Therefore, as with the hepatitis C population, the GDG recommended that liver biopsy should be considered in these people when transient elastography is not suitable, for example in someone who has not abstained from alcohol for at least 6 weeks prior to testing. Again, this must be an informed choice by the patient. Further results of the original economic model showed that retesting annually was not cost-effective compared to retesting every 2 years in any of the populations. The GDG hence advised that people with hepatitis C, ALD or NAFLD with advanced fibrosis found to be negative for cirrhosis should be retested for cirrhosis using the same tests every 2 years if they still have the underlying condition. People whose hepatitis C has been cured in the meantime, people who have stopped drinking or whose NAFLD has improved, do not require retesting. The GDG noted the recommendation of CG165[141] that people with hepatitis B not undergoing antiviral treatment should be retested annually with transient elastography for assessment of progression of liver disease.
Quality of evidence	The majority of the evidence was of Very low or Low quality, with some exceptions. The main reason for downgrading the quality of the evidence was the risk of bias. The majority of the included studies were at high risk of bias because of perceived inadequacy of the liver biopsy. The GDG felt that the variation in biopsy length of the reference standard would dramatically impact on the accuracy of the index test, and any heterogeneity between studies might be attributed to this. The other reason for downgrading the quality of the evidence was the imprecision around the effect estimates as seen by wide confidence intervals. Of the 252 full text articles ordered, the main reason that evidence was excluded was because the criteria for the biopsy length did not match the review protocol. The GDG discussed that it was a balance between excluding too many studies and only including studies where they could be confident that the evidence represented the true accuracy of the test. The GDG agreed that the biopsy standard was important and if this were relaxed further then it would reduce their confidence in the evidence. The current recommended biopsy length in the UK is 25 mm containing at least 10 portal tracts. It was agreed that including all evidence, even from studies using biopsies smaller than 15 mm or 6 portal tracts, or not stating the biopsy criteria, would have a profound effect of the accuracies of the diagnostic tests, as they would be compared to a reference standard of lower accuracy and one that does not reflect the measure of adequacy of liver biopsy used in the UK today. The GDG also reviewed the number of studies that had been excluded for this reason, and whether a different standard would dramatically increase the available evidence. They found that the evidence base for HCV would be increased, but there was not a dramatic effect for the other aetiologies where evidence is lacking. Therefore, it was agreed that it was not worth compromising the quality of the evidence. The other main exclusion reason was studies including populations with cirrhosis of varying aetiology. The GDG confirmed these studies should be excluded due to different test performance for each aetiology, as outlined above.
Other considerations	In determining who should be tested using the diagnostic tests investigated, the GDG had recourse to recommendation 1 (see Section 5.7). They noted that: • All people diagnosed with hepatitis B are recommended by the NICE Hepatitis B

guideline (CG165)[141] to be tested for stage of liver disease using transient elastography.

- All people diagnosed with hepatitis C require knowledge of their cirrhosis status to determine the appropriate drugs to use to treat their hepatitis C.
- There is no widely accepted definition of who has or should be suspected of alcohol-related liver disease short of histopathological diagnosis using liver biopsy following taking a history of alcohol use. The GDG agreed that those who should be suspected as at high risk of cirrhosis due to alcohol misuse are men drinking more than 50 units per week and women drinking more than 35 units per week over a period of at least several months. In addition, anyone already diagnosed with alcohol-related liver disease by a specialist should also be tested for cirrhosis. Retesting for people who drink excessively but have not been diagnosed with alcohol related liver disease was not recommended as no adequate evidence was identified to support it. Therefore people receiving an initial negative cirrhosis diagnosis should be monitored for alcohol related liver disease according to existing NICE guidance (CG100){National Clinical Guideline Centre, 2010 CG100 /id}.
- People with obesity or type 2 diabetes are only at risk of cirrhosis through developing NAFLD. People with obesity or type 2 diabetes without NAFLD are not at immediate risk of cirrhosis and do not need to be tested.
- 'Non-alcoholic fatty liver disease (NAFLD): assessment and management'[142] recommends testing those diagnosed with NAFLD for advanced fibrosis. The guideline recommends using the ELF test, with a threshold of 10.51 to test for advanced fibrosis, as it was found to be the most diagnostically accurate test, and to be cost-effective compared to all other testing and non-testing strategies at a cost-effectiveness threshold of £20,000 per QALY gained. As all those who will go on to develop cirrhosis will first develop advanced fibrosis it is sufficient to test those with both NAFLD and advanced fibrosis for cirrhosis; people with NAFLD but without advanced fibrosis do not need to be tested for cirrhosis. The GDG therefore adopted the subgroup of people with NAFLD and advanced fibrosis (as determined by testing using ELF) as the population of interest in testing people with NAFLD for cirrhosis.

The GDG discussed in more detail the use of these tests and their applicability. It was agreed that the combination of a blood fibrosis test and an imaging test would be theoretically beneficial, as they measure different biological aspects of the disease, therefore they should give independent results and complement each other. Whilst imaging tests assess the current level of fibrosis in the liver but give little idea of whether the process of fibrosis is ongoing, blood fibrosis tests are more dynamic. The GDG discussed that another benefit of performing a combination of tests in people with ALD is that seeing multiple test results could encourage people to make necessary changes in their lifestyle, for example abstain from alcohol or at least reduce their alcohol consumption. A positive response may also be seen in patients with hepatitis B or C, as they may be encouraged to be concordant with antiviral treatment. For it to lead to greater accuracy than a single test alone, a combination of 2 tests would need to include a first test with high sensitivity and a second test with high specificity. It would be preferable for the first test to be a blood test (cheap and easy to conduct on a large number of people) with the second test being an imaging test, conducted on the smaller group testing positive for the first test. However, in practice no such combinations of tests were available given the diagnostic accuracy results of the individual tests, and so no combinations were considered apart from those algorithms tested within a single study included in the clinical review.

There was a general discussion that many of the papers the GDG were familiar with were not included in this review. It was noted that the current review looked at the accuracy of the tests specifically for the diagnosis of cirrhosis (for example F4 stage only when the METAVIR scoring system was used). This excluded a number of papers

	which grouped F3 and F4 together. In particular, the GDG noted that there is an evidence base looking at the accuracy of ELF for the diagnosis of 'advanced fibrosis' (F3 and F4 grouped together), but not for cirrhosis alone. The GDG discussed the tests that should be considered in the economic modelling. It was noted that in HCV there was no need for a diagnostic test of fibrosis in primary care because all patients will be referred to secondary care for an assessment and treatment. The GDG felt that the preferred diagnostic test for a HCV population should be transient elastography or ARFI. The GDG considered 2 additional methods of scoring the liver biopsy: the Batts-Ludwig and the Scheuer scoring systems, and questioned whether these should have been included in the protocol for the review, as another option for the reference standard. However, the GDG agreed that evidence from studies using these reference standard measures are in the minority compared to METAVIR. Also, it noted that changes in the reference standard can alter the sensitivity and specificity values of the tests. The GDG agreed that these reference standard methods should remain excluded from the review. It was agreed that transient elastography needs to be performed according to the manufacturer's instructions and operators need to be fully trained. It was also a concern that with the introduction of non-invasive tests, skills amongst new clinicians in performing liver biopsies may be lost, before the standard of the non-invasive tests have been developed to an acceptable level and are widely available. The GDG agreed that anyone diagnosed with cirrhosis should be referred to a hepatologist (or someone with more than 50% work time commitment to hepatology) for initial assessment. In some situations it may be appropriate for the patient's clinical and supportive care to be offered within primary care or by means of a shared-care model.

7 Severity risk tools

7.1 Introduction

The natural history of cirrhosis is characterised by a variable period, often of several years, during which the person affected remain well with few if any clinical symptoms and signs. At some stage, often determined by the passage of time but in some instances relating to lifestyle issues or intercurrent illness, complications develop which relate to the development of either portal hypertension or hepatocellular failure, or both. These complications include jaundice, ascites, variceal haemorrhage, or hepatic encephalopathy, and define the transition from compensated to decompensated cirrhosis. The course of cirrhosis varies considerably from person to person related to several factors, including the aetiology of the cirrhosis, changes in lifestyle for example abstinence from alcohol in people with alcohol-related cirrhosis, treatment for the underlying cause of the liver injury for example antiviral agents for people with chronic hepatitis C infection, and the development of hepatocellular carcinoma (HCC). The development of decompensation is associated with reduction in survival but this is not inevitable.

Since the course of cirrhosis is variable and because it is recognised that clinical evaluation alone does not accurately predict outcome, there is a clear need for a cirrhosis risk assessment tool to assist in the identification of people who are at high risk of liver decompensation before they experience a defining event. Such a risk prediction tool would allow patients with compensated cirrhosis to be optimally managed by providing information on the timing for referral to specialist hepatology services.

Several scoring systems have been developed to either assess the prognosis of cirrhosis or prioritise candidates for transplantation, including the Child-Pugh score[167], the model for end-stage liver disease (MELD) score[131] and the UK end-stage liver disease (UKELD) score. Other tests such as transient elastography have also been proposed for the assessment of prognosis of cirrhosis. The GDG decided to compare the clinical and cost-effectiveness of these risk assessment tools for predicting the risk of mortality and liver-related morbidity in people with compensated cirrhosis. The use of these tools to prioritise patients on a liver transplant waiting list is not included in this review.

7.2 Review question 1: Which risk assessment tool is the most accurate and cost-effective for predicting the risk of morbidity and mortality in people with compensated cirrhosis?

Review question 2: When (at what severity score on the risk assessment tool) should people with cirrhosis be referred to specialist care?

For full details see review protocol in Appendix C.

Table 39: Characteristics of review question

Population	Adults and young people >16 years with compensated cirrhosis (no prior decompensating event)
Target condition	Cirrhosis
Index test(s)/comparator(s)	Model for end-stage liver disease (MELD) Child-Pugh (Child-Turcotte-Pugh)

	UK model for end-stage liver disease (UKELD) Transient elastography (transient elastography) Modified risk tools by the addition of the following risk factors: • Hepatovenous portal pressure gradient (HVPG) • Sodium (Na) (for example MELD-Na) • Delta-MELD • MELD-EEG • Transient elastography • Nutrition
Event	• Survival (all-cause mortality) • A decompensating event (hepatic encephalopathy; ascites; spontaneous bacterial peritonitis (SBP); variceal bleeding; hepatorenal syndrome; jaundice) or hepatocellular carcinoma (HCC)
Statistical measure/outcomes	**Critical outcomes:** • ROC area under the curve (of each risk tool for each outcome)/concordance c-statistic • Sensitivity, specificity, predictive values • Predicted risk, observed risk/calibration plot (reproduced with author permissions) (that is, predicted x-year mean risk % verses Kaplan-Meier x-year event rate). Narrative of agreement between observed and predicted risk and whether underestimation/overestimation of predicted risk) • D statistics, R^2 statistic and Brier score
Study design	Cohort (prospective or retrospective). Only external validation studies (not the development/derivation or internal validation studies).

7.3 Clinical evidence

We searched for prospective and retrospective cohort studies assessing the accuracy of severity risk tools to predict the risk of mortality and liver-related morbidity in people with compensated cirrhosis. Only data relating to individuals with compensated cirrhosis at baseline were included. This population was specified as the aim was to find the most accurate risk tool for the prediction of mortality or decompensation in people who are currently compensated, but may require referral for specialist care because they are at higher risk of having one of these future events. Ten studies were included in the review.[9,61,74,111,112,114,162,174,183,242] Evidence from these are summarised in the clinical evidence profile below. See also the study selection flow chart in Appendix E, sensitivity/specificity forest plots in Appendix K, study evidence tables in Appendix H and exclusion list in Appendix L.

Seven studies looked at the prognostic accuracy of transient elastography at a variety of thresholds[61,111,112,114,162,174,242]; 3 studies looked at the MELD score[9,74,183] and 1 study looked at Child-Pugh.[183] The components of these 3 tools can be found in Table 40, Table 41 and Table 42. No studies were identified that looked at the prognostic accuracy of the UKELD score in the prediction of mortality or decompensation in people with compensated cirrhosis. No studies were identified that looked at the prognostic accuracy of a modified risk tool in prediction of mortality or decompensation in people with compensated cirrhosis (modified by addition of one of the following factors to the risk tool: HVPG, Na, delta-MELD, EEG, transient elastography or nutrition).

Table 40: MELD components

Variables/risk factors included in MELD	
MELD	
Serum bilirubin	3.78×ln[serum bilirubin (mg/dl)]

Variables/risk factors included in MELD	
International normalised ratio of prothrombin time	11.2×ln[INR]
Serum creatinine	9.57×ln[serum creatinine (mg/dl)]
Aetiology	6.43 × (0: cholestatic or alcohol-related, 1-otherwise)
MELD-Na (updated formula including serum sodium)	
Expressed as a function of MELD	MELD + 1.32 x (137-Na)– [0.033x MELD*(137-Na)]

Table 41: Child-Pugh components

Variables/risk factors included in Child-Pugh	
Total bilirubin (µmol/litre)	<34=1 point 34–50=2 points >50=3 points
Serum albumin (g/dl)	>3.5=1 point 2.8–3.5=2 points <2.8=3 points
International normalised ratio of prothrombin time	<1.7=1 point 1.71–2.3=2 points >2.3=3 points
Ascites	None=1 point Mild=2 points Moderate to severe=3 points
Hepatic encephalopathy	None=1 point Grade I-II (or suppressed with medication)=2 points Grade III–IV (or refractory)=3 points

Table 42: Transient elastography component

Variables/risk factors included in transient elastography	
Liver stiffness (kPa)	Standard procedure is that transient elastography is performed on the right lobe of the liver through the intercostal spaces in patients lying in the dorsal decubitus position with the right arm in maximal abduction. A liver portion ≥6 cm thick and free of large vascular structures is located. 10 acquisitions with a success rate ≥60% and an interquartile range to ratio <30% of the median value are classed as representative measurements.

The outcomes of interest were all-cause mortality and decompensation. The outcome of mortality was only reported as overall mortality and no studies reported liver-related mortality specifically. The decompensating events included under the outcome of decompensation differed slightly between individual studies and this information is summarised in Table 43. Six studies looked at a composite outcome of death or decompensation.[9,61,112,114,242] This outcome was analysed separately. One study also looked at a composite outcome of hepatic decompensation, varices development and variceal growth and a single outcome of varices progression.[242] Varices development, growth or progression

were not considered decompensating events in the review protocol, therefore these two outcomes were analysed separately.

Assessing the performance of a risk tool

Evaluating the performance of a prediction model is typically done by examining discrimination and calibration. Discrimination refers to the ability of the prediction model to distinguish between those who do or do not experience the event of interest (decompensation). Calibration concerns how well the predicted risks compare to observed risks. A model is well calibrated if, for every 100 patients given a prediction of p%, the observed number of events is close to p%. Discrimination is typically assessed by calculating the area under the receiver operating characteristic curve (ROC AUC or c-statistic), where a value of 0.5 implies the model is no better than flipping a coin. However, there are limitations in the usefulness and interpretation of the area under the receiver operating characteristic curve to conclude whether the model is of any use. Calibration is evaluated either by calculating the Hosmer-Lemeshow test statistic, or preferably by plotting predicted risks against observed risks (calibration plot). The resulting calibration plot, if there is close agreement, will have points lying on or around a line of 45° with a slope value of approximately 1.0.

Predictive test accuracy and discrimination

We wished to know how accurate the risk stratification tools are in predicting mortality or decompensation. This means we want to know across a population if:

- a high risk score in an individual is reflected in a future event (mortality or decompensation);
- a low risk score in an individual is reflected in freedom from a future event (mortality or decompensation).

This is very similar, in principle, to how we look at diagnostic test accuracy (for diagnosis) and we take an analogous approach here – and use the term 'predictive test accuracy'. Accordingly, we can use similar methods to determine predictive test accuracy statistics and similar quality assessments to diagnostic test accuracy. There are however some important differences, mainly related to the time dependence of prognosis, including the play of chance (that is, the fact that the event is yet to happen when we measure risk) and this means we have to modify our quality assessment and carry out additional analyses to truly answer these types of question (see below).

By analogy with diagnostic test accuracy, we considered the risk stratification tool as the 'index test' and the outcome (observed mortality or decompensation) as the 'reference standard'. To calculate the sensitivity and specificity we have to define the cut-off threshold for high and low risk. The area under the receiver operating characteristics (ROC) curve, abbreviated to area under the curve (AUC) can also be calculated. The ROC curve is a curve fitted to the set of combinations of sensitivity and (1-specificity), across all possible (theoretical) cut-off points. The AUC gives an overall measure of accuracy of the risk tool across a range of thresholds. An AUC of 1 would indicate a perfect risk tool that can discriminate between people who will and will not have the event.

AUC on its own is not a good method of discriminating between risk stratification tools because the statistics are very insensitive even to major changes in the algorithm, and we also investigated calibration and reclassification, where reported.

Differences between prognostic tests are best determined by both discrimination and calibration

Outcomes reported

All the studies reported outcomes relating to the discriminative ability of the prognostic risk tools. For each outcome, ROC AUC values, as reported by the individual studies, are summarised in Table 50, Table 51, Table 52, Table 53, Table 54, Table 55 and Table 56. The GDG agreed on the following

criteria for AUC: 90–100% indicates perfect discrimination; 70–89% indicates moderate discrimination; 50–69% indicates poor discrimination and <50% not discriminatory at all. Data other than AUC (for example sensitivity/specificity for certain thresholds, R^2, D statistics, Brier score) were also presented if given.

In addition to identifying the most accurate risk tool, the aim was also to identify a risk threshold at which people with compensated cirrhosis should be referred for specialist care. Coupled sensitivity and specificity values at given cut-off thresholds were reported for each risk tool. This information was used to determine the risk tool (and threshold) with the highest sensitivity, without the expense of losing specificity. The ideal threshold would have a high sensitivity, so that people who will have a future event are defined as high risk by the tool and are referred. A high sensitivity would mean fewer false negatives (people who will have a future event, but are defined as low risk because they fall below the chosen threshold on the risk tool and therefore are not referred). Lower thresholds will give a high sensitivity, however this would be at the expense of specificity. A specific risk tool would mean fewer false positives (people who will not have a future event, but are defined as high risk by the tool) and therefore, fewer referrals of people who are not at risk.

Unlike discrimination outcomes, only 1 study was identified which reported outcomes related to the calibration of the risk tools.[74] This study reported calibration for the MELD score. No studies were identified reporting calibration for the other risk tools.

No study reported reclassification of the risk tools. Reclassification is used to examine the added-value of new risk factors that have been proposed to improve the risk tool. No studies were identified that looked at the prognostic accuracy of a modified risk tool in prediction of mortality or decompensation in people with compensated cirrhosis (modified by addition of one of the following factors to the risk tool: HVPG, Na, delta-MELD, EEG, transient elastography or nutrition).

Table 43: Summary of studies included in the review

Study	Severity risk tool(s)	Number of patients	Aetiology	Median length of follow-up	Outcomes
Aravinthan 2013[9]	MELD	n=77	Alcohol	Median follow-up 57 months (1–120)	Adverse liver-related outcome (composite outcome): liver-related death, decompensation, variceal bleed, ALD and sepsis, liver transplantation, HCC
Ferlitsch 2012[61]	Transient elastography	n=145	Not reported	Maximum 64 months	Decompensation (ascites, jaundice, grade 3/4 hepatic encephalopathy, variceal bleeding, death, liver transplantation)
Finkenstedt 2012[74]	MELD	n=429	Alcohol-related/NASH (58.7%), hepatitis (25.6%), cryptogenic (5.7%), other (9.8%)	Median 1.3 years (IQR 0.6-3.5)	90-day mortality
Kim 2012[111]	Transient elastography	n=217	Hepatitis B	Median 42.1 months (range 6.1-	Hepatic decompensation

Study	Severity risk tool(s)	Number of patients	Aetiology	Median length of follow-up	Outcomes
				58.4 months)	events (ascites development, hepatic encephalopathy, variceal haemorrhage, deterioration of liver function to Child-Pugh class B or C)
Kim 2014[112]	Transient elastography	n=69	Hepatitis B	2 years	Hepatic decompensation (variceal bleeding, ascites, hepatic encephalopathy, SBP, HCC and hepatorenal syndrome)
Klibansky 2012[114]	Transient elastography	n=160	Hepatitis C (67.5%), NAFLD/NASH (13.5%), hepatitis B (8.2%), alcohol-related (1.9%), autoimmune hepatitis (1.8%), PBC (1.3%), hemochromatosis (0.6%), sarcoid (0.6%), PSC (0.4%), DILI (0.4%), cryptogenic (0.3%), other (3.3%)	Median 854 days	Composite outcome: death from any cause, first variceal bleed, new-onset ascites, new-onset encephalopathy, increase in Child-Pugh score by 2 or more, HCC or listing for liver transplant
Perez-latorre 2014[162]	Transient elastography	n=60	Hepatitis C	Median follow-up 42 months	Liver decompensation (ascites, hepatic encephalopathy, variceal bleeding, jaundice, HCC) HCC Liver-related events (composite outcome): decompensation or HCC, whichever occurred first
Robic 2011[174]	Transient elastography	n=65	Alcohol (38%), hepatitis B or C (28%), NASH (8%), auto-immune hepatitis (8%),	Mean follow-up 491 days	Clinical decompensation (defined as portal hypertension-related bleeding,

Study	Severity risk tool(s)	Number of patients	Aetiology	Median length of follow-up	Outcomes
			cholestatic liver disease (3%), other (9%)		ascites, hepatorenal syndrome, hepatic encephalopathy, HCC, and/or sepsis)
Said 2004[183]	MELD Child-Pugh	n=204	Alcohol (29.9%) ALD and HCV (14.5%) HCV (21.8%)	Median 24 months (1-72)	1-year mortality
Wang 2014b[242]	Transient elastography	n=220	Hepatitis B (54.1%), hepatitis C (30.5%), hepatitis B and hepatitis C (8.6%), other cause (6.8%)	36.9 months	Hepatic decompensation (variceal bleeding, development or growth; ascites; SBP; hepatic encephalopathy; HCC; liver-related death) Portal hypertension Progression (composite outcome): hepatic decompensation, varices development and varices growth Clinical disease Progression (composite outcome): portal hypertension progression, HCC development and liver-related death HCC Varices progression

Table 44: Clinical evidence profile: 90-day mortality

Index Test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (median/ range/ 95% CI)	Specificity % (median/ range/ 95% CI)	Quality
Index Test									
MELD (≥16) (range 6–40, lower values better)	1	429	No risk of bias[a]	No inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	0.85 (0.76, 0.92)	0.83 (0.79, 0.87)[e]	Moderate

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making

a) Risk of bias was assessed using the PROBAST checklist

b) Inconsistency was assessed by inspection of the sensitivity/ (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold of 90% set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

c) Indirectness was assessed using the PROBAST checklist items referring to applicability

d) The judgement of precision was based on visual inspection of the confidence region of the sensitivity value. As a rule of thumb (after discussion with the GDG) a range of 0–20% of differences in point estimates of sensitivity was considered not imprecise, 20–40% serious, and >40% very serious imprecision. Imprecision was assessed on the primary measure for decision-making (test specificity)

e) The quoted specificity value is the value associated with the median sensitivity (the primary measure) in order to maintain paired values; sensitivity was the primary measure discussed in decision-making

Table 45: Clinical evidence profile: Composite of death and other clinical events

Index Test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (median/ range/ 95% CI)	Specificity % (median/ range/ 95% CI)	Quality
Index Test									
Transient elastography (8.0 kPa)	1	160	Serious risk of bias[a]	No inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	1.00 (0.91, 1.00)	0.06 (0.02, 0.12)	Moderate
Transient elastography (10.5	1	160	Serious risk	No	Serious indirectness	No serious	0.97 (0.87, 1.00)	0.10 (0.05, 0.17)	Moderate

Index Test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (median/ range/ 95% CI)	Specificity % (median/ range/ 95% CI)	Quality
kPa)			of bias[a]	inconsistency[b]	[c]	imprecision[d]			
Transient elastography (12.5 kPa)	1	160	Serious risk of bias[a]	No inconsistency[b]	Serious indirectness [c]	No serious imprecision[d]	0.93 (0.80, 0.98)	0.16 (0.10, 0.24)	Moderate
Transient elastography (14.0 kPa)	1	220	No risk of bias[a]	No inconsistency[b]	Serious indirectness [c]	Serious imprecision[d]	0.57 (0.43–0.7)	0.68 (0.61–0.75)	Moderate
Transient elastography (15.0 kPa)	1	160	Serious risk of bias[a]	No inconsistency[b]	Serious indirectness [c]	Serious imprecision[d]	0.85 (0.70, 0.94)	0.27 (0.19, 0.36)	Low
Transient elastography (20.0 kPa)	1	160	Serious risk of bias[a]	No inconsistency[b]	Serious indirectness [c]	Serious imprecision[d]	0.80 (0.64, 0.91)	0.39 (0.30, 0.48)	Low
Transient elastography (30.0 kPa)	1	160	Serious risk of bias[a]	No inconsistency[b]	Serious indirectness [c]	Serious imprecision[d]	0.30 (0.17, 0.47)	0.53 (0.44, 0.62)	Low
Transient elastography (50.0 kPa)	1	160	Serious risk of bias[a]	No inconsistency[b]	Serious indirectness [c]	No serious imprecision[d]	0.05 (0.01, 0.17)	0.93 (0.87, 0.97)	Moderate
Transient elastography (70.0 kPa)	1	160	Serious risk of bias[a]	No inconsistency[b]	Serious indirectness [c]	No serious imprecision[d]	0.03 (0.00, 0.13)	0.98 (0.94, 1.00)	Moderate

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making

a) Risk of bias was assessed using the PROBAST checklist

b) Inconsistency was assessed by inspection of the sensitivity/ (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold of 90% set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

c) Indirectness was assessed using the PROBAST checklist items referring to applicability

d) The judgement of precision was based on visual inspection of the confidence region of the sensitivity value. As a rule of thumb (after discussion with the GDG) a range of 0–20% of differences in point estimates of sensitivity was considered not imprecise, 20–40% serious, and >40% very serious imprecision. Imprecision was assessed on the primary measure for decision-making (test specificity)

Table 46: Clinical evidence profile: Decompensation

Index Test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (median/ range/ 95% CI)	Specificity % (median/ range/ 95% CI)	Quality
Index Test									
Transient elastography (19.0 kPa)	1	69	No risk of bias[a]	No inconsistency[b]	No indirectness[c]	Serious imprecision[d]	0.92 (0.62, 1.00	0.42 (0.29, 0.56)	Moderate
Transient elastography (21.1 kPa)[e]	1	65	No risk of bias[a]	No inconsistency[b]	No indirectness[c]	No serious imprecision[d]	1.00 (0.81, 1.00)	0.40 (0.26, 0.56)	High
Transient elastography (21.1 kPa)	1	220	Serious risk of bias[a]	No inconsistency[b]	Serious indirectness[c]	Very serious imprecision[d]	0.78 (0.48-0.95)	0.84 (0.79-0.89)	Low
Transient elastography (<25.0 kPa)	1	60	No risk of bias[a]	No inconsistency[b]	No indirectness[c]	Serious imprecision[d]	0.92 (0.72-1.0)	0.65 (0.50-0.79)	Moderate
Transient elastography (<40.0 kPa)	1	60	No risk of bias[a]	No inconsistency[b]	No indirectness[c]	Very serious imprecision[d]	0.67 (0.36-0.98)	0.90 (0.80-0.99)	Moderate

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making

a) Risk of bias was assessed using the PROBAST checklist

b) Inconsistency was assessed by inspection of the sensitivity/ (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold of 90% set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

c) Indirectness was assessed using the PROBAST checklist items referring to applicability

d) The judgement of precision was based on visual inspection of the confidence region of the sensitivity value. As a rule of thumb (after discussion with the GDG) a range of 0–20% of differences in point estimates of sensitivity was considered not imprecise, 20–40% serious, and >40% very serious imprecision. Imprecision was assessed on the primary measure for decision-making (test specificity)

e) Variceal bleeding and/or ascites only

One study [61] reported that transient elastography predicted decompensation with a 20.3% sensitivity and an 88.2% specificity in 145 patients, however the threshold used was not reported.

One study [111] reported the following data for transient elastography predicting decompensation:

Score on risk tool:	**Risk of event:**
<13 kPa	0.93, 0.9, 2.31 and 4.02% at 1, 2, 3 and 4 years
13-18 kPa	5.88, 10.54, 132.74 and 23.10% at 1, 2, 3 and 4 years
≥18 kPa	13.38, 23.21, 30.5 and 55.32% at 1, 2, 3 and 4 years

One study [174] reported the following data for transient elastography predicting variceal bleeding and/or ascites:

Score on risk tool:	**Risk of event:**
<21.1 kPa	47%
≥21.1 kPa	100%

Table 47: Clinical evidence profile: HCC

Index Test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (median/ range/ 95% CI)	Specificity % (median/ range/ 95% CI)	Quality
Index Test									
Transient elastography (11.5 kPa)	1	220	No risk of bias[a]	No inconsistency[b]	No indirectness[c]	Very serious imprecision[d]	0.53 (0.32-0.73)	0.52 (0.45-0.59)	Low

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making

a) Risk of bias was assessed using the PROBAST checklist

b) Inconsistency was assessed by inspection of the sensitivity/ (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold of 90% set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

c) Indirectness was assessed using the PROBAST checklist items referring to applicability

d) The judgement of precision was based on visual inspection of the confidence region of the sensitivity value. As a rule of thumb (after discussion with the GDG) a range of 0–20% of differences in point estimates of sensitivity was considered not imprecise, 20–40% serious, and >40% very serious imprecision. Imprecision was assessed on the primary measure for decision-making (test specificity)

Table 48: Clinical evidence profile: Composite of hepatic decompensation, varices development and variceal growth

Index Test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (median/ range/ 95% CI)	Specificity % (median/ range/ 95% CI)	Quality
Index Test									
Transient elastography (17.0 kPa)	1	220	No risk of bias[a]	No inconsistency[b]	Serious indirectness[c]	Serious imprecision[d]	0.57 (0.39–0.73)	0.78 (0.72–0.83)	Low

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making
a) Risk of bias was assessed using the PROBAST checklist
b) Inconsistency was assessed by inspection of the sensitivity/ (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold of 90% set by the GDG (the threshold above which would be acceptable to recommend a test). The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)
c) Indirectness was assessed using the PROBAST checklist items referring to applicability
d) The judgement of precision was based on visual inspection of the confidence region of the sensitivity value. As a rule of thumb (after discussion with the GDG) a range of 0–20% of differences in point estimates of sensitivity was considered not imprecise, 20–40% serious, and >40% very serious imprecision. Imprecision was assessed on the primary measure for decision-making (test specificity)

Table 49: Clinical evidence profile: Varices progression

Index Test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity % (median/ range/ 95% CI)	Specificity % (median/ range/ 95% CI)	Quality
Index Test									
Transient elastography (12.0 kPa)	1	220	No risk of bias[a]	No inconsistency[b]	Serious indirectness[c]	Very serious imprecision[d]	0.62 (0.38-0.82)	0.60 (0.53-0.67)	Moderate

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making
a) Risk of bias was assessed using the PROBAST checklist
b) Inconsistency was assessed by inspection of the sensitivity/ (based on the primary measure) forest plots, using the point estimates and confidence intervals. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold of 90% set by the GDG (the threshold above which would be acceptable to recommend a test). The

evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)
c) Indirectness was assessed using the PROBAST checklist items referring to applicability
d) The judgement of precision was based on visual inspection of the confidence region of the sensitivity value. As a rule of thumb (after discussion with the GDG) a range of 0–20% of differences in point estimates of sensitivity was considered not imprecise, 20–40% serious, and >40% very serious imprecision. Imprecision was assessed on the primary measure for decision-making (test specificity)

AUC data

Table 50: Clinical evidence profile: Mortality

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality
MELD (90 day)	1	429	No risk of bias[a]	No inconsistency	Serious indirectness[b]	Serious imprecision[c]	0.9 (0.84–0.96)	Low
MELD (1 year)	1	204	No risk of bias[a]	No inconsistency	No indirectness[b]	No serious imprecision[c]	0.75 (0.59–0.9)	High
Child-Pugh (1 year)	1	204	No risk of bias[a]	No inconsistency	No indirectness[b]	No serious imprecision[c]	0.66 (0.5–0.82)	High

a) Risk of bias was assessed using the PROBAST checklist
b) Indirectness was assessed using the PROBAST checklist items referring to applicability
c) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

Table 51: Clinical evidence profile: Composite of death and decompensation

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median [range]	Quality
MELD	1	77	No risk of bias[a]	No inconsistency	No indirectness[b]	Serious imprecision[c]	0.59 (0.47–0.72)	Moderate

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median [range]	Quality
Transient elastography	2	380	No risk of bias[a]	No inconsistency	Serious indirectness[b]	No serious imprecision[c]	0.59 (0.50–0.69), 0.668 (0.577–0.759)	Moderate

a) Risk of bias was assessed using the PROBAST checklist
b) Indirectness was assessed using the PROBAST checklist items referring to applicability
c) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

Table 52: Clinical evidence profile: Decompensation

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median [range]	Quality
Transient elastography	5	631	No risk of bias[a]	No inconsistency	Serious indirectness[b]	Serious imprecision[c]	0.793 (0.62–0.852) [0.734–0.929]	Low

a) Risk of bias was assessed using the PROBAST checklist
b) Indirectness was assessed using the PROBAST checklist items referring to applicability
c) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

Table 53: Clinical evidence profile: Decompensation or HCC

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median [range]	Quality
Transient elastography	1	60	No risk of bias[a]	No inconsistency	No indirectness[b]	Serious imprecision[c]	0.85 (0.73–0.97)	Moderate

a) Risk of bias was assessed using the PROBAST checklist
b) Indirectness was assessed using the PROBAST checklist items referring to applicability

c) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

Table 54: Clinical evidence profile: HCC

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality
Transient elastography	2	280	No risk of bias[a]	Serious inconsistency	No indirectness[b]	Very serious imprecision[c]	0.77 (0.59–0.95), 0.504 [0.358–0.651]	Low

a) Risk of bias was assessed using the PROBAST checklist
b) Indirectness was assessed using the PROBAST checklist items referring to applicability
c) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

Table 55: Clinical evidence profile: Composite of hepatic decompensation, varices development and variceal growth

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality
Transient elastography	1	220	No risk of bias[a]	No inconsistency	Serious indirectness[b]	No serious imprecision[c]	0.744 (0.65–0.838)	Moderate

a) Risk of bias was assessed using the PROBAST checklist
b) Indirectness was assessed using the PROBAST checklist items referring to applicability
c) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

Table 56: Clinical evidence profile: varices progression

Index Test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality
Transient elastography	1	220	No risk of bias[a]	No inconsistency	Serious indirectness[b]	No serious imprecision[c]	0.638 (0.525–0.75)	Moderate

a) Risk of bias was assessed using the PROBAST checklist
b) Indirectness was assessed using the PROBAST checklist items referring to applicability
c) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example 0–50%, 50–90% and 90–100%)

Calibration data

Calibration of MELD for 3-month mortality was poor for scores within the lower three quintiles but seemed to be fairly good in the fourth and fifth quintile of each score. The calibration of the scores for 1-year mortality was better but still remained imprecise within the lower quintiles.

7.4 Economic evidence

7.4.1 Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

7.4.2 Unit costs

See Table 91 in Appendix O.

7.5 Evidence statements

7.5.1 Clinical

- Low quality evidence from 5 studies (n=631) demonstrated a good AUC value for transient elastography in predicting decompensation (0.79) but Moderate quality evidence from 2 studies (n=380) demonstrated a lower accuracy for predicting a composite outcome of death and/or decompensation (AUC 0.59).
- Low quality evidence from 1 study (n=429) indicated a good AUC value (0.90) for MELD in predicting 90- day all-cause mortality.
- High quality evidence from 1 study (n=204) indicated a moderate AUC value (0.75) for MELD in predicting 1-year all-cause mortality.
- High quality evidence from 1 study (n=204) indicated a moderate AUC value (0.66) for Child- Pugh in predicting 1-year all-cause mortality.
- Moderate quality evidence from 1 study (n=77) indicated a moderate AUC value (0.59) for MELD in predicting a composite outcome of death and decompensation.
- There was Moderate to Low quality evidence demonstrating that with transient elastography; as the threshold increases, sensitivity decreases and specificity increases in predicting decompensation.
- Moderate quality evidence from 1 study (n=429) showed a high sensitivity and specificity of MELD in predicting 90-day all-cause mortality at a threshold of 16.

7.5.2 Economic

- No relevant economic evaluations were identified.

7.6 Recommendations and link to evidence

Recommendations	**13. Refer people who have, or are at high risk of, complications of cirrhosis to a specialist hepatology centre.** **14. Calculate the Model for End-Stage Liver Disease (MELD) score every 6 months for people with compensated cirrhosis.** **15. Consider using a MELD score of 12 or more as an indicator that the person is at high risk of complications of cirrhosis.**
Relative values of	The GDG was interested in the prognostic accuracy of severity risk tools to predict

different outcomes	the risk of mortality and liver-related morbidity in people with compensated cirrhosis. GDG members discussed that currently, in their opinion, there are a large number of patients with compensated cirrhosis who are not referred to specialist hepatology services until they have an episode of decompensation. The GDG aimed to identify a risk tool that would be able to predict both all-cause mortality, and liver-related complications in people with compensated cirrhosis (defined as any of hepatic encephalopathy, hepatocellular carcinoma, ascites, spontaneous bacterial peritonitis, variceal haemorrhage, hepatorenal syndrome and jaundice). The outcome of mortality was only reported as overall mortality and no studies reported liver-related mortality specifically (with the exception of some studies reporting composite outcomes of morbidity and mortality). The GDG felt that the ability to predict those at high risk of a decompensating event should be the priority as this would allow timely prevention and intervention in people with compensated cirrhosis. The GDG agreed that people at risk of decompensation are likely to benefit from specialist hepatologist care. A population with compensated cirrhosis at baseline was specified, as the aim was to find the most accurate risk tool for the prediction of all-cause mortality or decompensation in people who are currently compensated, but may require referral for specialist care because they are at higher risk of having one of these events in future. The GDG agreed that people with decompensated cirrhosis should have already been referred to specialist services, and the aim here is to intervene before decompensation or death occurs. Therefore, studies of prognostic accuracy in people with decompensated cirrhosis at baseline were not considered in this review. The GDG focused on the value of the ROC AUC for each risk tool, as reported by the studies. This gives an overall measure of the prognostic accuracy of the tool across a range of cut-off thresholds and was used to identify the most accurate risk prediction tool. In addition to identifying the most accurate tool to predict these future events, the GDG also wanted to recommend a cut-off threshold for referral. This is a trade-off between sensitivity and specificity. A high sensitivity was desirable so that people who are at higher risk are not missed (fewer false negatives), but the GDG did not want to compromise the specificity too much, as patients who are not at risk would be referred unnecessarily (false positives).
Trade-off between clinical benefits and harms	Evidence was identified for the Child-Pugh score, the MELD score and transient elastography. No studies were identified that looked at the risk prediction accuracy of the UKELD score or of a risk tool modified by addition of one of the following factors to the risk tool: HVPG, Na, delta-MELD, EEG, transient elastography or nutrition. The GDG discussed the evidence for transient elastography. The evidence demonstrated a good AUC value for transient elastography for predicting decompensation (0.79) but a lower accuracy for predicting a composite outcome of death or decompensation (AUC 0.59). There was no AUC evidence for transient elastography for the prediction of mortality alone. The GDG expressed concerns about the use of transient elastography for assessing prognosis in people with cirrhosis. Transient elastography is only a measure of the degree of fibrosis in the liver. The GDG agreed that more fibrosis would not necessarily mean an increased risk of complications, and that transient elastography alone should not be used for prediction of patients at high risk of decompensation. The GDG noted that aetiology plays an important role in the accuracy of transient elastography. It is most useful in hepatitis C but would not be as accurate in predicting risk in people with alcohol-related cirrhosis or obesity. In addition, transient elastography would need to be repeated regularly if it were to be used as a risk prediction tool, which is not in line with current clinical practice. MELD data were available for the prediction of 1-year mortality with an AUC of 0.75 (95% CI 0.59–0.9). This value was better than the AUC value of the Child-Pugh score for prediction of 1-year mortality (0.66, 95% CI 0.5–0.82). The GDG also discussed other benefits of MELD over the Child-Pugh score; for example, MELD is completely

	objective (not subjective on clinical assessment) and all the constituent parts of MELD have independently been associated with mortality and outcome in patients with cirrhosis. There was also evidence for MELD for the predication of 3-month mortality with an AUC of 0.9 (95% CI 0.84–0.96). However, the GDG had concerns about the applicability of this study (Finkenstedt 2012[74]). The GDG suspected that this study included people with decompensated cirrhosis at baseline, due to the high mortality rate at 3 months and the high cut-off value of 16 used to assess sensitivity and specificity. Unfortunately, this was the only evidence for the sensitivity and specificity of MELD at a defined cut-off threshold. No AUC evidence was available for MELD or Child-Pugh for the prediction of decompensation. However, evidence was available for MELD for the prediction of a composite outcome of decompensation or death, with an AUC of 0.59 (95% CI 0.47–0.72). This is comparable to the AUC value for the prediction of decompensation or death using transient elastography, but the evidence for MELD was higher quality. As there was no evidence on the prognostic accuracy of Child-Pugh for decompensation and this outcome was considered a priority for this review, the GDG agreed that overall the evidence, along with their other considerations, supports the use of MELD for predicting risk in people with compensated cirrhosis. The GDG agreed it was important to recommend a cut-off threshold for referral to specialist hepatology services if they were to recommend the use of MELD as a severity risk tool. There is no standard threshold currently used for referral of people at high risk of decompensation to specialist care. The only commonly used threshold is a MELD score ≥16, which is used for assessment for liver transplantation. This cut-off is too high for the current recommendation as the people referred would most likely have decompensated disease. Thus the GDG, based on their clinical knowledge and experience, felt that a threshold of 12 for referral for to a specialist service would seem reasonable. MELD scores of around 6–7 indicate the presence of well compensated disease while a MELD score between 8 and 11 would not indicate an immediate risk of decompensation. A MELD score of 12 or above would mean the person would have an abnormal level for at least one of the measure variables and the GDG agreed this would put them at risk of decompensation. The GDG also made the point that if the MELD score were below 12, this is no way precluded referral to a specialist service if there were concerns about the wellbeing of a patient with cirrhosis. All the data identified were AUC or sensitivity and specificity values: that is, the discriminative ability of the risk tools. Data on calibration was only available from 1 study for MELD scores. The study concluded that calibration of MELD for 3-month mortality was poor for scores within the lower 3 quintiles but seemed to be fairly good in the fourth and fifth quintile of each score. The calibration of the scores for 1-year mortality was better but still remained imprecise within the lower quintiles. However, this evidence was from the Finkenstedt 2012 study[74] which, as discussed above, was in a population with an unclear status in relation to their functional hepatic reserve at baseline. The GDG did not feel this evidence could be considered due to the limitations of the study.
Trade-off between net clinical effects and costs	No relevant published economic evidence was identified. The GDG reviewed the unit costs of transient elastography, MELD and Child-Pugh score. The costs of calculating MELD and Child-Pugh are low as they only require the results of inexpensive blood tests, which are performed as part of routine clinical assessment. Transient elastography is more expensive; in particular the GDG noted that this would need to be repeated at regular intervals, leading to substantial costs over time. Although it is current practice to conduct the blood tests on a regular basis, it is not routine practice to automatically calculate MELD or Child-Pugh scores from them. The GDG considered the possibility of having laboratories routinely calculate and

	present the MELD score, in the same way as eGRF is routinely calculated from the serum creatinine as a measure of renal function. The GDG concluded that the routine report of the MELD score would be beneficial in people with confirmed cirrhosis, but might mislead clinicians to suspect liver decompensation inappropriately in people without liver disease but with raised creatinine and bilirubin due to other causes. The GDG also considered the costs and benefits of consultation with a specialist at a hepatology centre. For people at high risk of developing decompensation, the GDG agreed that there would be considerable benefit in specialist assessment, and such a referral would very likely be cost-effective. However, if too low a threshold were used and people at low risk were also referred for specialist assessment this would lead to increased costs with little benefit, and this strategy would be unlikely to be cost-effective. Therefore, the GDG recommended that all people with cirrhosis with a MELD score of 12 or more should be referred to a specialist hepatology centre; those with a lower score should not be referred routinely but they could have other reasons to justify referral to a specialist centre apart from the MELD score.
Quality of evidence	The AUC data for transient elastography for prediction of decompensation or for predicting a composite measure of death or decompensation was of Low quality, adding to the GDG's uncertainty about the use of transient elastography as a prognostic risk tool in people with cirrhosis. There was no AUC evidence for transient elastography for the prediction of mortality. Both High and Low quality evidence was available for the MELD AUC data for the prediction of 1-year and 3-month mortality, respectively. The GDG noted this High quality evidence for the MELD score for prediction of 1-year mortality, which showed a good AUC value as described above. There was also High quality evidence for the Child-Pugh AUC score for the prediction of 1-year mortality. However, as described above, the MELD score had a higher AUC than Child-Pugh for prediction of 1-year mortality (AUC 0.75, 95% CI 0.59–0.9; and 0.66, 95% CI 0.5–0.82, respectively, both High quality). No AUC evidence was available for MELD for the prediction of decompensation. However, the evidence for the MELD AUC data for the prediction of a composite outcome of decompensation or death was Moderate quality. Overall, the GDG agreed the AUC evidence for the MELD score was better quality evidence. However, data were only available from 1 study for the sensitivity and specificity of MELD at a particular cut-off threshold. This was Moderate quality evidence due to the indirect population. The GDG did not consider the population to be appropriate to the review question and recommendations, as it suspected people had decompensated disease at baseline. Therefore, the GDG used clinical judgement and experience to determine the optimal cut-off threshold for referral to specialist hepatology services. There was no evidence for the sensitivity and specificity of Child-Pugh for the prediction of mortality or decompensation. The GDG noted that some of the papers reviewed included decompensating events, which were not included in our protocol (for example, deterioration of liver function to a lower Child-Pugh class). This evidence was included, but downgraded for indirectness. The GDG focused on the protocol-specified outcomes of decompensation and mortality for decision-making.
Other considerations	The GDG discussed that the population in question are most likely to be cared for in secondary care or in shared primary care, and the recommendation is meant for those people who should be referred for specialist hepatology care, in either a secondary or tertiary centre. The GDG unanimously agreed that anyone with cirrhosis who has had a decompensating event prior to being referred to a specialist hepatology centre should be referred immediately. The GDG also noted that there can be considerable improvement in liver function with treatment of the underlying cause of liver disease, for example, abstinence from alcohol or treatment for hepatitis C infection. This should be taken into account when considering referral to specialist services. This in turn could influence the frequency of assessment which

	should be designed not only to detect deterioration, but also improvement. The GDG were aware of a number of papers assessing MELD which were not included in the review. It was noted that these papers had been excluded as they predicted future risk in people who were decompensated at baseline. These studies are more abundant than studies in people with compensated cirrhosis, however they are not relevant for the recommendations on referral to specialist care. In general, the GDG considered that MELD is a robust prognostic marker in people with compensated cirrhosis. The GDG felt that MELD is easy to calculate using the results of blood tests undertaken as part of standard practice, but noted that it would be useful if laboratories were encouraged to generate a MELD score automatically on liver blood tests, which could be used easily by clinicians. The GDG agreed that this recommendation is largely aimed at secondary care clinicians, as people with a diagnosis of cirrhosis are routinely seen in secondary care. The GDG discussed the limitations of using MELD in primary care. Unless someone has a definitive diagnosis of cirrhosis, the MELD score can be misleading as abnormal results can occur in other conditions such as kidney disease. The GDG acknowledged that although no evidence was identified on the UKELD score, this is routinely used in many specialised centres. However, no further research on this risk prediction tool was recommended since the GDG was aware of an ongoing project led by NHS Blood & Transplant that aims to replace UKELD and modify the system of organ allocation in the UK.

8 Surveillance for the early detection of hepatocellular carcinoma (HCC)

8.1 Introduction

People with cirrhosis are at high risk of developing hepatocellular carcinoma (HCC) with an annual cancer risk of between 1% and 6% depending on the aetiology of the liver disease, for example HCV OR=13.4 (95% CI, 4.1 to 43.5), HBV OR=9.1 (2.1 to 39.5), and alcohol misuse OR=1.8 (0.7 to 4.3). Around 3,900 people are diagnosed with HCC each year in the UK, which accounts for about 1% of all cancers. However, HCC incidence rates are rising and over the next 20 years mortality is predicted to increase by 39% (from 4.2 to 5.9 per 100,000).[56]

The prognosis in people with HCC critically depends on tumour stage at the time of diagnosis. For example those with early HCC, defined as one nodule less than 5 cm or three nodules each less than 3 cm in diameter, can achieve 5-year survival rates near 70% with surgical resection or liver transplantation.[102] In contrast, the survival rate in people who present with symptoms associated with a large tumour is poor, with median survival less than 6 months.[123]

Regular surveillance for people with cirrhosis, using liver ultrasound with or without serum alpha-fetoprotein (AFP) testing at 3- to 12-month intervals, endeavours to detect a tumour at an early stage when potentially curative treatment can be offered. Therefore, the GDG decided to compare the clinical and cost-effectiveness of ultrasound surveillance (with or without serum AFP testing) compared to no surveillance, and surveillance at difference frequencies of 3-, 6- and 12-month intervals for the detection of HCC in people with cirrhosis due to HCV, alcohol and non-alcohol-related fatty liver disease.

8.2 Review question: When and how frequently should surveillance testing be offered for the early detection of hepatocellular carcinoma (HCC) in people with cirrhosis?

For full details see review protocol in Appendix C.

Table 57: PICO characteristics of review question

Population	Adults and young people (16 and over) with confirmed cirrhosis, without HCC at the start of surveillance, or with a history of HCC prior to surveillance.

Intervention	Intervention:
	• No surveillance
	• Surveillance with ultrasound, with or without serum alpha-fetoprotein (AFP) testing:
	  ○ yearly
	  ○ 6-monthly
	  ○ 3-monthly

Comparison	No surveillance versus surveillance Different frequencies of surveillance
Outcomes	**Critical outcomes:** • Transplant-free survival (time-to-event) or mortality at 5 years • Health-related quality of life **Important outcomes:** • HCC occurrence

	• Lesion of HCC less than or equal to 3 cm, greater than 3 cm • Number of lesions (if multiple lesions) • Liver cancer staging (according to Barcelona Clinic Liver Cancer [BCLC] system) • Liver transplant
Study design	RCTs, systematic reviews of RCTs, observational studies, systematic reviews of observational studies

Randomised and observational studies comparing ultrasound surveillance (with or without serum AFP testing) with no surveillance for the early detection of HCC in people with cirrhosis were searched for. Randomised and observational studies comparing the effectiveness of different frequencies of ultrasound surveillance were also searched for. Implicit in the investigation of surveillance or surveillance frequency is that patients in whom HCC is detected earlier can be treated earlier, and have potentially better patient outcomes and a better chance of survival. The guideline does not, however, cover how HCC should be managed after diagnosis.

The review population was limited to people with confirmed cirrhosis. Due to the known link between HBV or HCV and HCC, some studies in the literature assess the effectiveness of surveillance in people with HCV or HBV regardless of the level of fibrosis or presence of cirrhosis. Studies including these populations were not included in this systematic review, which focuses on the effectiveness of surveillance in people with a confirmed diagnosis of cirrhosis. Studies including people with a mix of fibrosis stages were only included in the systematic review when the proportion of people with cirrhosis was >85%.

There is existing NICE guidance on surveillance for HCC in people with HBV (with fibrosis at any stage, including cirrhosis) and therefore, this guideline does not cover surveillance in people with cirrhosis due to HBV. There is evidence in the literature to suggest that people with HBV are at higher risk of developing HCC, even if they do not have cirrhosis. Therefore, the effectiveness of HCC surveillance may be different in this population compared with other disease aetiologies. For studies including people with cirrhosis and mixed aetiologies, studies were only included in the systematic review when the proportion of people with HBV was ≤15% due to the difference in prognosis in this population, and therefore an expected difference in the effectiveness of surveillance. A second approach agreed in the protocol was to use studies in people with cirrhosis and a proportion of people with HBV >15% if no evidence was identified using the above criteria. Evidence was identified for the comparison of surveillance versus no surveillance and for the comparison of different frequencies, therefore studies in people with cirrhosis and a proportion of people with HBV >15% were not included.

Observational non-randomised studies were included in the absence of evidence from RCTs. As pre-specified in the review protocol, only observational studies which performed a multivariate analysis to adjust for confounding factors were included in the review. Non-randomised studies reporting the characteristics of HCC (such as lesion size or the number of people with an advanced HCC stage) without adjusting for confounders were excluded from this review. All the observational studies identified in this review were retrospective cohort studies in people with a diagnosis of HCC, retrospectively analysed by their previous surveillance status prior to the diagnosis of HCC. Therefore, only people who developed HCC were analysed. No prospective observational studies were identified in people without HCC, followed up to see if HCC developed.

Studies of screened people who develop HCC compared to unscreened people who develop HCC are subject to lead-time bias. Lead-time bias is the apparent increase in survival that comes exclusively from diagnosis at an earlier stage of disease. The duration of survival from diagnosis to death is increased, even if no intervention is applied and some of the survival benefit could be ascribed to earlier diagnosis. Some of the studies included in this review attempted to correct for this bias by calculating the lead time and adjusting for it. This was taken into account when assessing the risk of bias of the included studies. Studies are also subject to length bias. This arises from the fact that

surveillance is more likely to detect slow growing cancers than rapidly growing cancers, which might go from undetectable to death within the surveillance interval.

Following detection of lesions during ultrasound surveillance, all the included studies followed standardised protocols for the diagnosis and staging of HCC and treated HCC according to standard procedures. Details of the individual studies can be found in the clinical evidence tables in Appendix H.

8.3 Clinical evidence

8.3.1 Surveillance versus no surveillance

Six observational studies[83,135,160,219,228,231] were identified comparing surveillance with no surveillance in people with cirrhosis. All were retrospective observational studies comparing outcomes in people with confirmed HCC depending on previous surveillance status. No RCTs were identified for this comparison. The outcomes of survival and HCC stage were reported from multivariate analyses. No evidence was identified from a multivariate analysis for the outcomes of quality of life, HCC occurrence, number of lesions, liver transplantation or lesion size less than 3 cm. Evidence for this comparison is summarised in the clinical evidence summary in Table 59.

8.3.2 Different surveillance frequencies

Two included studies compared surveillance at different frequencies: yearly versus 6-monthly surveillance in one observational study[188] and 3-monthly versus 6-monthly surveillance in one RCT.[233] Five other observational studies retrospectively compared yearly surveillance with 6-monthly surveillance, but were excluded from the current review due to the inclusion of people with cirrhosis of mixed aetiologies and the high proportion of people with HBV,[187,229,230] or due to only reporting unadjusted results.[45,59]

8.3.2.1 Yearly versus 6-monthly surveillance

One observational study[188] was identified comparing yearly surveillance with 6-monthly surveillance in people with cirrhosis. This was a retrospective observational study comparing outcomes in people with confirmed HCC depending on previous surveillance status. No RCTs were identified for this comparison. The outcomes of survival and HCC stage were reported from multivariate analyses. No evidence was identified from a multivariate analysis for the outcomes of quality of life, HCC occurrence, number of lesions, liver transplantation or lesion size less than 3 cm. Evidence for this comparison is summarised in the clinical evidence summary in Table 60.

8.3.2.2 3-monthly versus 6-monthly surveillance

One RCT[233] was identified comparing 3-monthly surveillance with 6-monthly surveillance in people with cirrhosis. Evidence was available for all protocol outcomes with the exception of quality of life. Evidence for this comparison is summarised in the clinical evidence summary in Table 61.

Table 58: Summary of studies included in the review

Study, Study design	Population	Intervention and comparison	Follow-up	Outcomes
Surveillance versus no surveillance				
Giannini 2000[83]	n=61 HCC and HCV-related	Ultrasound+AFP 6-monthly versus	N/A	Survival

Study, Study design	Population	Intervention and comparison	Follow-up	Outcomes
Retrospective cohort	cirrhosis 100% HCV	HCC detected incidentally		
Miquel 2012[135] Retrospective cohort	n=110 HCC and cirrhosis HCV 56.1%; alcohol 25.1%; HBV 2%; HCV+alcohol 11.2%; cryptogenic 5.2%	Ultrasound+AFP 6-monthly versus HCC detected incidentally (liver lesions detected as a result of imaging explorations)	N/A	Survival
Pascual 2008[160] Retrospective cohort	n=290 HCC and cirrhosis Alcohol 29.3%; HCV 45.9%; HBV 4.8%; alcohol+virus 8.3%; other 11.7%	Ultrasound+AFP 6-monthly versus HCC detected incidentally or due to symptoms	N/A	Survival
Stroffolini 2011[219] Retrospective cohort	n=411 HCC and 94.7% cirrhosis Mixed aetiologies (HBsAg negative/HCV positive 56.1% (15% HBsAg positive or HBsAg positive and anti-HCV positive))	Ultrasound 6- to 12-monthly (unclear if also used AFP) versus no surveillance	N/A	HCC stage
Trevisani 2004[228] Retrospective cohort	n=363 HCC and 97.5% cirrhosis Mixed aetiologies (predominantly HCV: 79.6% HCV or HCV co-infection) 12.7% HBV (unclear how many of the people with multiple aetiologies had HBV)	ultrasound+AFP 6- to 12-monthly versus HCC detected incidentally versus HCC detected by symptoms	N/A	Survival (adjusted risk not reported as it was not found to be an independent prognostic factor) HCC stage
Trevisani 2007[231] Retrospective cohort	n=608 HCC and cirrhosis Mixed aetiologies (predominantly HCV) 10.4% HBV (unclear how many of the people with multiple aetiologies had	ultrasound+AFP 6 to 12-monthly versus HCC detected incidentally or by symptoms	N/A	Survival (adjusted risk not reported as it was not found to be an independent prognostic factor)

Study, Study design	Population	Intervention and comparison	Follow-up	Outcomes
	HBV)			
Yearly versus 3-monthly surveillance				
Santi 2010[188] Retrospective cohort	n=649 HCC and cirrhosis HCV 63.3%; HBV 9.1%; alcohol 7.9%; multiple 15.9%; other 3.9%	ultrasound+AFP yearly versus ultrasound+AFP 6-monthly	N/A	Survival HCC stage
3-monthly versus 6-monthly surveillance				
Trinchet 2011[233] RCT	n=1340 Cirrhosis and no previous HCC Alcohol 39.2%; HCV 44.1%; HBV 13.2%; Haemochromatosis 1.6%; other 2.5%	ultrasound 3-monthly (and either AFP assay every 6 months or no serum AFP assay) versus ultrasound 6-monthly (and either AFP assay every 6 months or no serum AFP assay)	Median 47 months	Survival HCC occurrence HCC stage HCC ≤3 mm Number of nodules Liver transplant

Table 59: Clinical evidence summary: Surveillance versus no surveillance

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with no surveillance	Risk difference with surveillance (95% CI)
Survival Adjusted hazard ratio (HR >1 indicates an advantage to the surveillance group)[(a)]	351 (2 studies) 9 months reported by one study, follow-up in other study not reported	VERY LOW[(b),(c)] due to risk of bias, imprecision	Median not calculated as 2 studies only (range of HRs 1.49–2.61)	–	–[(e)]
Survival Adjusted odds ratio (OR >1 indicates an advantage to the surveillance group)[(d)]	110 (1 study) 5–7 years from recruitment estimated	VERY LOW[(b),(c)] due to risk of bias, imprecision	OR 1.13 (0.64 to 2.01)	–	–[(e)]
Detection of HCC at a very early stage (single nodule ≤2 cm) Adjusted odds ratio (OR >1 indicates an advantage to the surveillance group)[(f)]	1729 (1 study)	LOW	OR 5.4 (2.35 to 12.4)	–	–[(e)]
Detection of HCC at a non-advanced stage (single nodule ≤5 cm or 3 nodules each ≤3 cm without vascular and lymph nodal invasion and metastases) Adjusted odds ratio (OR >1 indicates an advantage to the surveillance group)[(f)]	1729 (1 study)	LOW	OR 3.1 (1.85 to 5.2)	–	–[(e)]
Detection of HCC at an advanced stage (according to Milano criteria) - Surveillance versus incidental diagnosis Adjusted odds ratio (OR <1 indicates an advantage to the surveillance group)[(g)]	296 (1 study)	LOW	OR 0.29 (0.17 to 0.49)	–	–[(e)]
Detection of HCC at an advanced stage (according to Milano criteria) - Surveillance versus symptom diagnosis	296 (1 study)	LOW	OR 0.18 (0.09 to 0.37)	–	–[(e)]

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with no surveillance	Risk difference with surveillance (95% CI)
Adjusted odds ratio (OR <1 indicates an advantage to the surveillance group)[(g)]					

[a] Study 1 adjusted for the following confounders: gender, Child-Pugh score, number of tumoural nodules (1/>1), AFP value, AFP (normal/increased), type of treatment (treated/not treated) and modality of diagnosis (follow-up/incidental). Study 2 adjusted for the following confounders: Child-Pugh status, tumour characteristics, treatment applied for HCC
[b] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias (main reasons for risk of bias include no adjustment for lead-time bias or no adjustment for all the key confounders)
[c] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
[d] Adjusted for factors found to be statistically significant in the univariate analysis: degree of liver function, screening, tumour size, and curative versus palliative. In this analysis, screening was not statistically significant (not an independent predictor of survival)
[e] Control group risk not reported for calculation of absolute effect
[f] Adjustment for the confounding factors (age, gender, surveillance, aetiologies, AFP levels, cirrhosis)
[g] Adjusted for centre of enrolment, age, sex, aetiology of cirrhosis, Child-Pugh class, AFP level and type of diagnosis

Narrative information for surveillance versus no surveillance

Trevisani 2004[228] compared ultrasound surveillance every 6–12 months with both HCC detected by symptoms and HCC detected incidentally. For both comparisons, surveillance was not a statistically significant independent predictor of survival , so the adjusted relative risk for the effect of surveillance on survival from multivariate analysis was not reported. Trevisani 2007[231] compared ultrasound surveillance every 6–12 months with HCC detected by symptoms or incidentally, with data stratified into Child-Pugh B and Child-Pugh C. Again, surveillance was not a statistically significant independent predictor of survival , so the adjusted relative risk for the effect of surveillance on survival from multivariate analysis was not reported.

The non-randomised studies included in this comparison also reported the characteristics of HCC including outcomes such as tumour size. However, these outcomes were not adjusted for confounding factors in a multivariate analysis, therefore results have not been extracted.

Table 60: Clinical evidence summary: Yearly versus 6-monthly surveillance

Outcomes	Number of	Quality of	Relative	Anticipated absolute effects

	participants (studies) Follow-up	the evidence (GRADE)	effect (95% CI)	Risk with 6-monthly surveillance	Risk difference with yearly surveillance (95% CI)
Survival Adjusted hazard ratio (HR >1 indicates an advantage to the 6-monthly surveillance group)[a]	649 (1 study)	VERY LOW[c] due to imprecision	HR 1.39 (1.06 to 1.82)	–	-[b]
Detection of HCC beyond a very early stage (solitary nodule >2 cm or multinodular tumour with/without vascular invasion and/or metastases) Adjusted odds ratio (OR >1 indicates an advantage to the 6-monthly surveillance group)[d]	649 (1 study)	LOW	OR 5.99 (2.57 to 13.98)	–	-[b]

[a] Adjusted variables: age, platelet count, AFP, Child-Pugh class and oesophageal varices
[b] Control group risk not reported for calculation of absolute effect
[c] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
[d] Adjusted variables included those associated with a tumour beyond the very early stage: surveillance interval, sex, aetiology, ALT, AFP, and Child-Pugh class

Narrative information for yearly versus 6-monthly surveillance

The non-randomised study included in this comparison also reported the characteristics of HCC including outcomes such as tumour size. However, these outcomes were not adjusted for confounding factors in a multivariate analysis, therefore results have not been extracted.

Table 61: Clinical evidence summary: 3-monthly versus 6-monthly surveillance

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with 6-monthly surveillance	Risk difference with 3-monthly surveillance (95% CI)
Survival Hazard ratio (HR <1 indicates an advantage to the 3-monthly surveillance group)	1278 (1 study) 47 months	MODERATE[c] due to imprecision	HR 0.87 (0.64 to 1.19)	142 per 1000[a]	17 fewer per 1000 (from 49 fewer to 25 more)[b]
HCC occurrence	1278 (1 study) 47 months	MODERATE[c] due to imprecision	RR 0.75 (0.54 to 1.06)	110 per 1000	27 fewer per 1000 (from 51 fewer to 7 more)
Diameter of the largest HCC nodule ≤30 mm Positive outcome, RR<1 indicates an advantage to the 6-monthly	1278 (1 study)	LOW[c]	RR 0.85 (0.57 to	77 per 1000	12 fewer per 1000

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with 6-monthly surveillance	Risk difference with 3-monthly surveillance (95% CI)
group	47 months	due to imprecision	1.27)		(from 33 fewer to 21 more)
Diameter of the largest HCC nodule >30 mm Negative outcome, RR<1 indicates an advantage to the 3-monthly group	1278 (1 study) 47 months	MODERATE[(c)] due to imprecision	RR 0.52 (0.25 to 1.07)	33 per 1000	16 fewer per 1000 (from 25 fewer to 2 more)
Number of lesions – Uninodular RR<1 indicates fewer events in the 3-monthly group	1278 (1 study) 47 months	LOW[(c),(d)] due to risk of bias, imprecision	RR 0.75 (0.48 to 1.19)	64 per 1000	16 fewer per 1000 (from 33 fewer to 12 more)
Number of lesions – 2 or 3 nodules RR<1 indicates fewer events in the 3-monthly group	1278 (1 study) 47 months	VERY LOW[(d)] due to risk of bias, imprecision	RR 1.25 (0.59 to 2.64)	19 per 1000	5 more per 1000 (from 8 fewer to 31 more)
Number of lesions – >3 nodules RR<1 indicates fewer events in the 3-monthly group	1278 (1 study) 47 months	VERY LOW[(c),(d)] due to risk of bias, imprecision	RR 0.57 (0.17 to 1.94)	11 per 1000	5 fewer per 1000 (from 9 fewer to 10 more)
Number of lesions - Infiltrative RR<1 indicates fewer events in the 3-monthly group	1278 (1 study) 47 months	LOW[(c),(d)] due to risk of bias, imprecision	RR 0.3 (0.08 to 1.08)	16 per 1000	11 fewer per 1000 (from 15 fewer to 1 more)
HCC stage (within Milan criteria: one nodule ≤50 mm or 2 or 3 nodules ≤30 mm) Positive outcome, RR<1 indicates an advantage to the 6-monthly group	1278 (1 study) 47 months	MODERATE[(c)] due to imprecision	RR 0.84 (0.56 to 1.24)	78 per 1000	12 fewer per 1000 (from 34 fewer to 19 more)
HCC stage (beyond Milan criteria: one nodule ≤50 mm or 2 or 3 nodules ≤30 mm) Negative outcome, RR<1 indicates an advantage to the 3-monthly group	1278 (1 study) 47 months	LOW[(c),(d)] due to risk of bias, imprecision	RR 0.55 (0.26 to 1.13)	31 per 1000	14 fewer per 1000 (from 23 fewer to 4 more)
Liver transplant	1278 (1 study)	VERY LOW[(d)] due to risk of bias,	RR 1.3 (0.64 to	20 per 1000	6 more per 1000 (from 7 fewer to 33 more)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with 6-monthly surveillance	Risk difference with 3-monthly surveillance (95% CI)
	47 months	imprecision	2.66)		

[a] Survival at 60 months in the control group was 85.8%
[b] Based on survival rate of control group at 60 months
[c] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
[d] Downgraded by 1 increment if the majority of the evidence was at high risk of bias and by 2 increments if the majority of the evidence was at very high risk of bias

8.4 Economic evidence

8.4.1 Published literature

One economic evaluation was identified that compared annual surveillance and 6-monthly surveillance for HCC in people with cirrhosis of mixed aetiology.[44]

One economic evaluation was identified that compared no surveillance, annual AFP, annual ultrasound, annual AFP plus ultrasound, 6-monthly AFP, 6-monthly ultrasound, and 6-monthly AFP plus ultrasound for HCC in people with cirrhosis with either ALD or hepatitis C.[226]

These are summarised in the economic evidence profile below Table 62 and the economic evidence tables in Appendix I.

See also the economic article selection flow chart in Appendix F.

Table 62: Economic evidence profile: frequency of surveillance for HCC

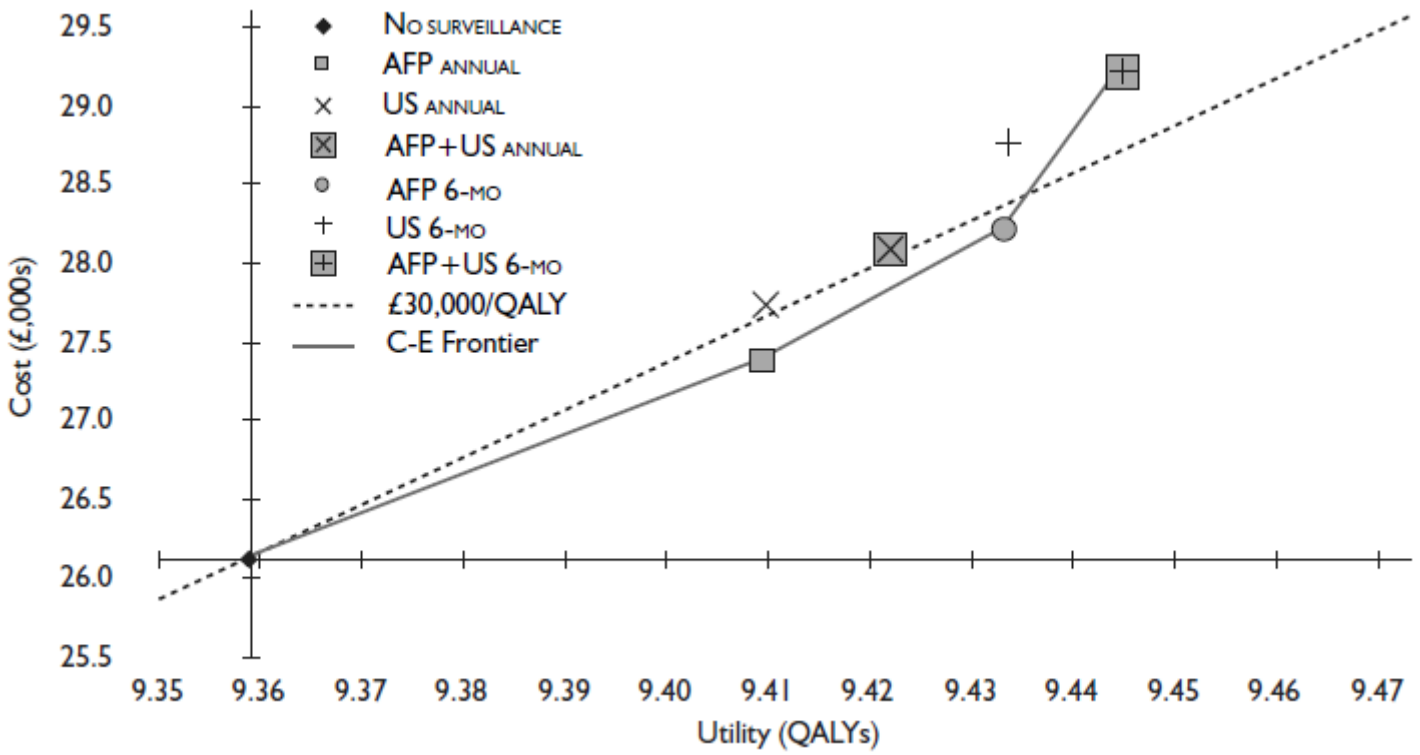
Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost-effectiveness	Uncertainty
Cucchetti 2012[44] (Italy)	Partially applicable[(a)]	Potentially serious limitations[(b)]	• Markov decision model with a 10-year time horizon • Model states: compensated cirrhosis, decompensated cirrhosis, surveillance, HCC diagnosis, HCC treatment, survival, death • Certain parameters related to a population including 18.1% hepatitis B patients	Compensated cirrhosis: £2,379 Decompensated cirrhosis: £2,462	Compensated cirrhosis: 0.11 Decompensated cirrhosis: 0.06	Compensated cirrhosis: £21,230 per QALY gained Decompensated cirrhosis: £40,540 per QALY gained	One-way sensitivity analyses: in patients with compensated cirrhosis changes in the annual HCC incidence or the risk ratio for survival gain make 6-monthly surveillance a cost-effective option at a threshold of £20,000 per QALY gained. In patients with decompensated cirrhosis no plausible variations in these 2 parameters made semi-annual treatment cost-effective compared with annual treatment.
Thompson Coon 2008 [225,226] (UK)	Directly applicable[(c)]	Minor limitations[(d)]	• Markov decision model with a lifetime horizon • Two relevant aetiologies: ALD and HCV • Model states include: No HCC, occult HCC (S,M,L), known HCC (S,M,L), transplant and resection in 4 discrete model sections: surveillance programme, transplant waiting list, curative treatment, palliative treatment	Details in Table 63, Table 64, Figure 4, and Figure 5.			ALD: At the £20,000 threshold, 'no surveillance' is likely to be the only cost-effective strategy (80% likelihood). At around £30,000 interventions 2,5 and 7 are all equally likely to be the preferable option HCV: At the £20,000 threshold, 'no surveillance' is likely to be considered cost-effective (75% likelihood). At £30,000 semi-annual AFP is preferred to no surveillance.

Abbreviations: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life years; RCT: randomised controlled trial
(a) Differences in healthcare system may make results less applicable to UK, a societal perspective was reported in terms of costs, no discounting applied to health effects
(b) Unclear source of resource use for every health state, unit cost data are extracted from Italian NHS data sources, only one-way and two-way sensitivity analyses were conducted
(c) Some quality of life values are based on authors' assumptions
(d) Only HCC-related costs are considered; not including costs related to other cirrhosis complications (such as ascites, hepatic encephalopathy)

Table 63: Thompson Coon 2008 – Cost-effectiveness results for ALD patients

Interventions	QALYs	Costs	ICER compared with previous best option
No surveillance	9.359	£26,100	N/A
AFP annual	9.410	£27,400	£25,490
Ultrasound annual	9.410	£27,700	Extendedly dominated
AFP plus ultrasound annual	9.422	£28,100	Extendedly dominated
AFP 6-monthly	9.433	£28,200	£34,783
Ultrasound 6-monthly	9.434	£28,800	Extendedly dominated
AFP plus ultrasound 6-monthly	9.445	£29,200	£83,333

Figure 4: Thompson Coon 2008 – Cost-effectiveness of HCC surveillance strategies for ALD patients



Source: Thompson Coon 2007 [225]

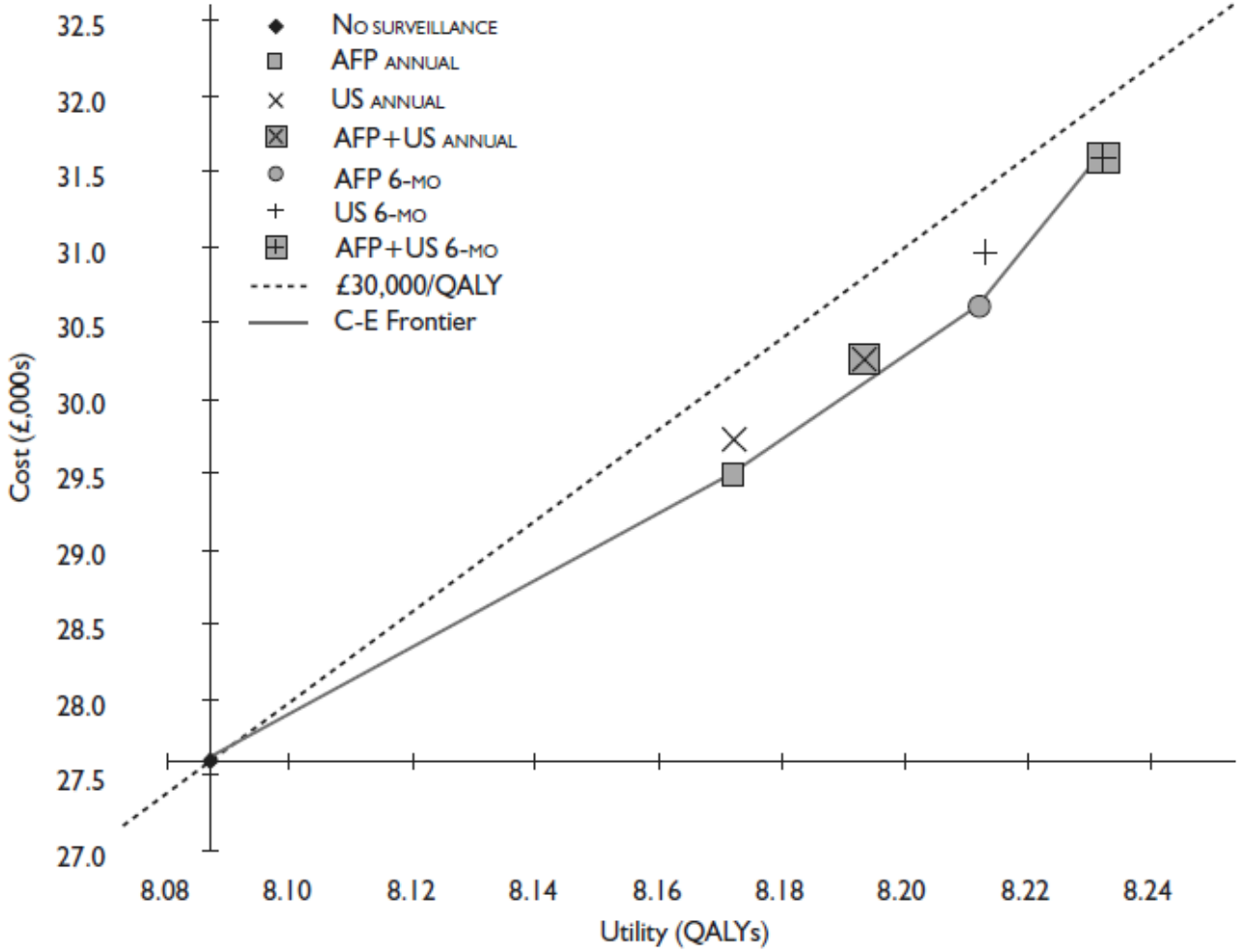
Note that the dotted line shown is for a cost-effectiveness threshold of £30,000 per QALY gained. A line for £20,000 per QALY gained would lie below all points on the graph

Table 64: Thompson Coon 2008 – Cost-effectiveness results for HCV patients

Interventions	QALYs	Costs	ICER compared with previous best option
No surveillance	8.087	£27,600	N/A
AFP annual	8.172	£29,500	£22,353

Interventions	QALYs	Costs	ICER compared with previous best option
Ultrasound annual	8.172	£29,700	Dominated
AFP plus ultrasound annual	8.193	£30,300	Extendedly dominated
AFP 6-monthly	8.212	£30,600	£27,500
Ultrasound 6-monthly	8.213	£31,000	Extendedly dominated
AFP plus ultrasound 6-monthly	8.232	£31,600	£50,000

Figure 5: Thompson Coon 2008 – Cost-effectiveness of HCC surveillance strategies for HCV patients



Source: Thompson Coon 2007 [225]

Note that the dotted line shown is for a cost-effectiveness threshold of £30,000 per QALY gained. A line for £20,000 per QALY gained would lie below all points on the graph

8.4.2 Unit costs

See Table 64 in Appendix N.

8.4.3 New cost-effectiveness analysis

Original cost-effectiveness modelling was undertaken for this question using the NGC liver disease pathway model developed for this guideline. A summary is included here. An evidence statement summarising the results of the analysis can be found below. The full analysis can be found in Appendix N.

8.4.3.1 Aim and structure

The aim of the health economic modelling for this question was to determine the optimal frequency of HCC surveillance. This was achieved by using the original lifetime diagnostic health state transition (Markov) model (see Section 6.4.3 and Appendix N) which followed the NICE reference case[146] and by comparing overall cost and QALYs of the cirrhosis tests of preference in 2 different scenarios; annual and semi-annual HCC surveillance.

To apply the clinical benefit of HCC surveillance, figures from 2 different sources, identified by the clinical review (1 included in the review – Santi 2010), were combined. A study by Zhang 2004 with a 5-year follow-up on 18,816 people with hepatitis B reported that 6-monthly surveillance (using alpha-fetoprotein [AFP] blood test plus ultrasound) was associated with a 37% reduction in HCC mortality in comparison to a no monitor control group. This number was combined with an increased risk of death figure (HR 1.39) for patients under annual surveillance using ultrasound plus AFP compared to a 6-monthly surveillance strategy reported by Santi 2010 (649 patients of mixed disease aetiology).

To determine the most cost-effective surveillance frequency, incremental cost-effectiveness ratios (ICERs) were calculated to compare the options in each population. Base case results below were obtained through the probabilistic analysis to take combined parameter uncertainty into account.

8.4.3.2 Results

Table 65: ICERs comparing 6-monthly surveillance against annual surveillance

Aetiology	ICER	Cirrhosis test used for the comparison
NAFLD	£23,220	TE at >15.0
ALD	£28,352	TE at 11.0 – <13.0
HBV – antigen	£26,063	TE at 11.0
HBV + antigen	£25,236	TE at 11.0
HCV genotype 1	£18,657	Liver biopsy
HCV genotype 3	£20,128	Liver biopsy

At a cost-effectiveness threshold of £20,000 per QALY gained, 6-monthly surveillance was cost-effective compared to annual surveillance only for the HCV genotype 1 group. Although this group had the fewest liver-related deaths, risk of HCC progression was particularly high in this group compared to other model cohorts, making more frequent surveillance cost-effective at the specified threshold. However, at a cost-effectiveness threshold of £30,000 per QALY gained, 6-monthly surveillance was cost-effective compared to annual surveillance for all groups: ICERs £18,657–28,352. Variation in the ICERs was mainly due differences in cirrhosis prevalence, risk of progression to HCC, and competing risks of other complications in each aetiology.

In the deterministic sensitivity analyses, reducing the surveillance costs or increasing the 6-monthly surveillance effectiveness reduced the ICERs for each group by up to £2,000 per QALY gained.

8.5 Evidence statements

8.5.1 Clinical

Surveillance versus no surveillance

- Very Low quality evidence from 2 studies (n=351) indicated a clinical benefit of surveillance for survival when analysed using time-to-event data. Low quality evidence from 1 study (n=1729) indicated a clinical benefit of surveillance for the detection of HCC at a non-advanced stage.

Yearly versus 6-monthly surveillance

- Very Low quality evidence from 1 study (n=649) indicated a clinical benefit of 6-monthly surveillance for survival. Low quality evidence from the same study indicated a clinical benefit of 6-monthly surveillance for the detection of HCC beyond a very early stage.

6-monthly versus 3-monthly surveillance

- Moderate quality evidence from 1 RCT (n=1278) indicated a clinical benefit of 3-monthly surveillance for survival and HCC occurrence. Evidence ranging from Very low to Moderate quality from the same RCT indicated no clinical difference in HCC diameter >30 mm at detection, the number of HCC nodules detected or the HCC stage at detection.

8.5.2 Economic

- One cost-utility analysis that compared 6-monthly versus annual surveillance for HCC in people with cirrhosis (mixed aetiology) found that 6-monthly surveillance for was not cost-effective for either compensated cirrhosis or decompensated cirrhosis (ICERs: £21,230 and £40,540 per QALY gained). This analysis was assessed as partially applicable with potentially serious limitations.
- One cost-utility analysis that compared 7 relevant strategies for surveillance of HCC in people with cirrhosis (including use or ultrasound, alpha-fetoprotein, both or neither, at 6-monthly or annual intervals) in people with ALD or HCV found that:
  - No other strategy was cost-effective compared to no surveillance at a cost-effectiveness threshold of £20,000 per QALY gained.
  - Annual surveillance using alpha-fetoprotein only had an ICER of £25,490 per QALY gained compared to no surveillance.

  This analysis was assessed as directly applicable with minor limitations.
- One original cost-utility analysis that compared 6-monthly with annual surveillance for HCC in people with cirrhosis at a cost-effectiveness threshold of £20,000 per QALY gained found that:
  - 6-monthly surveillance was cost-effective compared to annual surveillance for people with HCV genotype 1 (ICER: £18,657 per QALY gained).
  - 6-monthly surveillance was not cost-effective compared to annual surveillance for people with NAFLD, ALD, HBV or HCV genotype 3 (ICERs: £20,128–28,352).

  This analysis was assessed as directly applicable with minor limitations.

8.6 Recommendations and link to evidence

Recommendations	**16.Offer ultrasound (with or without measurement of serum alpha-fetoprotein) every 6 months as surveillance for hepatocellular**

	carcinoma (HCC) for people with cirrhosis who do not have hepatitis B virus infection. **17. For people with cirrhosis and hepatitis B virus infection, see the surveillance testing for hepatocellular carcinoma in adults with chronic hepatitis B section in NICE's hepatitis B (chronic) guideline.** **18. Do not offer surveillance for HCC for people who are receiving end of life care.**
Relative values of different outcomes	For the comparison of HCC surveillance with no surveillance, or different surveillance frequencies, the GDG chose transplant-free survival and quality of life as critically important outcomes. The GDG also considered HCC occurrence, the size of the HCC at the time of detection (≤3 cm, >3 cm), the number of lesions detected, the Barcelona Clinic Liver Cancer (BCLC) stage at the time of detection and the need for transplantation as important outcomes. All of the studies reported overall survival data rather than transplant-free survival. The GDG decided that overall survival was just as important and discussed this outcome when forming the recommendations. There is already existing NICE guidance on surveillance for HCC in people with HBV (with fibrosis at any stage, including cirrhosis) and therefore this guideline does not cover surveillance in people with cirrhosis due to HBV.
Trade-off between clinical benefits and harms	**Surveillance versus no surveillance** Six observational studies were identified comparing surveillance with no surveillance in people with cirrhosis. No RCTs were identified for this comparison. The outcomes of survival and HCC stage were reported from multivariate analyses. The GDG noted that there was a clinical benefit of surveillance on the outcome of survival from 2 studies reported as a hazard ratio, but a third study showed no clinical benefit of surveillance on survival, reported as an odds ratio. However, the GDG discussed that none of the studies adjusted for the lead-time bias. Two further studies reported in the narrative that surveillance was not found to be an independent predictor of survival, therefore the adjusted hazard ratios for the effect of surveillance on survival was not reported. One of these studies did adjust for possible lead-time bias. Evidence from 1 study showed an OR of 5.4 in favour of surveillance for detecting HCC at a very early stage (single nodule ≤2 cm). In addition, evidence from 2 studies showed a clinical benefit in favour of surveillance for detection of HCC at a non-advanced stage (according to the Milan criteria). No evidence was identified from a multivariate analysis for the outcomes of quality of life, HCC occurrence, number of lesions, liver transplant or lesion size less than 3 cm. **Yearly versus 6-monthly surveillance** One observational study was identified comparing yearly surveillance with 6-monthly surveillance in people with cirrhosis. No RCTs were identified for this comparison. The outcomes of survival and HCC stage were reported from multivariate analyses. The GDG noted that there was a clinical benefit of surveillance every 6 months compared to yearly surveillance on the outcome of survival, but the confidence intervals were wide. This study did adjust for the lead-time bias. There was also a clinical benefit of surveillance on the outcome of detection of HCC beyond a very early stage. No evidence was identified from a multivariate analysis for the outcomes of quality of life, HCC occurrence, number of lesions, liver transplant or lesion size less than or equal to 3 cm. **3-monthly versus 6-monthly surveillance**

One RCT was identified comparing 3-monthly surveillance with 6-monthly surveillance in people with cirrhosis. Evidence was available for all protocol outcomes with the exception of quality of life. There was a Hazard Ratio of 0.87 (95% CI 0.64 to 1.19) with 17 fewer patients per 1000 dying in the 3-monthly group versus the 6-monthly group. The GDG agreed this was a clinical benefit to the 3-monthly surveillance group. In the 3-monthly surveillance group the number of HCCs with a diameter ≤3 cm was less with a RR 0.85 (95% CI 0.57 to 1.27). There appeared to be no difference in the incidence of HCC >3 cm, the total number of nodules detected, or the number of infiltrative lesions in either the 3-monthly or the 6-monthly groups. The GDG noted that whilst there was no statistically significant difference between 3-monthly and 6-monthly surveillance, there was a trend towards a clinical benefit in favour of the shorter time interval. The GDG did note, however, that there were more patients who had HCC detected in the 6-monthly surveillance group so this may account for some of the variation in the stage of disease identified.

Overall considerations of the trade-off between clinical benefits and harms

The GDG felt that the fact that regular HCC surveillance is already common practice could not be overlooked. Existing guidance recommends that all people with hepatitis B and significant fibrosis should receive 6-monthly surveillance (NICE guideline CG165[141]) and this is established practice. Therefore, the GDG agreed that 6-monthly surveillance should be available to people with cirrhosis due to other underlying aetiologies. They also noted, however, that current practice is not supported by High quality evidence of improved survival. The GDG agreed that the decision to offer surveillance should be based on the ability to offer treatments for HCC. If the outcomes of surveillance are to be beneficial then effective treatments have to be available. The GDG discussed that new, more effective treatment options are available since the publication of some of the evidence included in this review and that this may enhance the benefit seen on patient survival. There are several effective treatment options available such as radiofrequency ablation, microwave ablation and chemoembolisation, as well as resection and liver transplantation. The GDG was concerned that there is some national variability in the availability of the treatment options offered to people with HCC. The management of HCC is not covered here as it is outside the scope of this guideline.

The GDG's consensus was that it would be a disservice to patients not to recommend surveillance for HCC. Overall the GDG felt that there was a clinical benefit from surveillance in the detection of tumours at an earlier stage and that surveillance should be offered to give patients the opportunity to receive potentially curative treatment. The GDG considered this to outweigh the potential harms of surveillance in terms of the over-investigation of false positive results. It was discussed that the doubling times of liver tumours are likely to be between 50 and 200 days with small primary HCC generally within the shortest timescales. Survival is strongly linked to size and stage of tumour so more regular screening was the preferred option of the GDG. There was a strong consensus among the GDG that the recommendation should be for a high quality, inclusive surveillance programme looking at all aspects of management and care provision. There was a strong representation from the patient perspective that people with cirrhosis would want to be monitored, especially with the advent of modern potentially curative treatment.

The GDG agreed that 6-monthly surveillance had clinical benefit over yearly surveillance. The GDG also discussed the potential economic and logistical considerations of recommending 3-monthly versus 6-monthly surveillance, and particularly the significant increased resource implication of 3-monthly surveillance in terms of radiology time. This would also increase the number of hospital visits required. Overall, the GDG agreed that 3-monthly surveillance was too frequent and that 6-monthly surveillance should be recommended.

Trade-off between net clinical effects and costs	The GDG noted that the costs of no surveillance take into account all of the treatment costs and the lifetime cost of care for patients who develop HCC. Two relevant economic evaluations were identified. The data from the paper by Cucchetti were based on 918 patients in 11 centres comparing surveillance with ultrasound and alpha-fetoprotein (with a CT scan arranged for all positive results). The results suggested that 6-monthly surveillance was not cost-effective compared to 12-monthly surveillance at a cost-effectiveness threshold of £20,000 per QALY gained (an ICER of only £21,230 for people with compensated cirrhosis but £40,540 per QALY for people with decompensated cirrhosis). A second paper by Thompson Coon assessed surveillance of patients with compensated cirrhosis under the age of 70 years. There were 7 different combinations of surveillance strategies proposed using ultrasound, alpha-fetoprotein or both, at intervals of 6 months or 12 months. They found that none of the surveillance strategies were cost-effective compared to no surveillance at a cost-effectiveness threshold of £20,000 per QALY gained, although the base case ICER for 12-monthly surveillance using alpha-fetoprotein only was £22,353 for people with hepatitis C. The GDG had a concern about the validity of the mortality rates associated with large tumours used as a basis of the economic calculations in this paper. The GDG acknowledged that the economic evidence presented did not support regular HCC surveillance at a threshold of £20,000 per QALY, however it highlighted that none of the presented models took into account the costs of non-HCC complications associated with cirrhosis (ascites, hepatic encephalopathy, variceal bleeding). In addition, in the case of Cucchetti, the GDG noted that differences in the healthcare system make the results less applicable to the UK. The GDG noted that the clinical evidence tended to be from more recent papers than the economic evidence. Although all the current interventions for HCC were available at the time the economic papers were published, the GDG agreed that these interventions may have been refined and the effectiveness of the interventions may have improved. The original economic modelling conducted for this guideline (see Appendix N) included 12-monthly surveillance using ultrasound with or without AFP in the base case, and investigated the cost-effectiveness of reducing this to every 6 months in a sensitivity analysis. Six-monthly surveillance was cost-effective at the threshold of £20,000 per QALY gained for the HCV genotype 1 cohort. For the other populations the ICERs for 6-monthly compared to annual surveillance varied between £20,128 and £28,352per QALY gained in each case. The second and third highest ICERs were for the HBV population (surveillance is already recommended for this group in the NICE Hepatitis B guideline CG165[141]).
Quality of evidence	Observational, non-randomised studies were included in the absence of evidence from RCTs. As pre-specified in the review protocol, only observational studies which performed a multivariate analysis to adjust for confounding factors were included. Non-randomised studies reporting the characteristics of HCC (such as lesion size or the number of people with an advanced HCC stage) without adjusting for confounders were excluded. All the observational studies identified were retrospective cohort studies in people with a diagnosis of HCC, analysed by their previous surveillance status prior to the diagnosis of HCC. Therefore, only people who developed HCC were analysed. No observational studies were identified in people without HCC, followed up to see if HCC developed. Studies of screened patients who develop HCC compared to unscreened patients who develop HCC are subject to lead-time bias. Lead-time bias is the apparent increase in survival that comes exclusively from diagnosis at an earlier stage of disease. The duration of survival from diagnosis to death is increased, even if no intervention is applied and some of the survival benefit could be ascribed to earlier

	diagnosis. Some of the included studies attempted to correct for this bias by calculating the lead time and adjusting for it. This was taken into account when assessing the risk of bias. Studies are also subject to length bias. This arises from the fact that surveillance is more likely to detect slow growing cancers than rapidly growing cancers, which might go from undetectable to death within the surveillance interval. As the studies only include patients with HCC detected with or without a surveillance program the magnitude and effect of false positives cannot be measured. **Surveillance versus no surveillance** Six observational studies were identified for this comparison. Evidence was available from a multivariate analysis for the outcomes of survival and HCC stage. However, all the evidence was graded Low or Very Low quality. The main reasons for downgrading the quality of the evidence for risk of bias and for imprecision. None of the studies adjusted for lead-time bias. **Yearly versus 6-monthly surveillance** One observational study was identified comparing yearly surveillance with 6-monthly surveillance in people with cirrhosis. No RCTs were identified for this comparison. Evidence was available from a multivariate analysis for the outcomes of survival and HCC stage. However, all the evidence was graded Low or Very Low quality. This study did adjust for the lead-time bias. **3-monthly versus 6-monthly surveillance** One RCT was identified comparing 3-monthly surveillance with 6-monthly surveillance in people with cirrhosis. Evidence was available for all protocol outcomes with the exception of quality of life. Evidence for the critical outcome of survival was of Moderate quality. Evidence for all other outcomes ranged from Moderate to Very Low quality. **Overall considerations on the quality of the evidence** There was a lack of High quality evidence and RCTs identified in this area to support recommendations made by the GDG. The GDG noted that very few RCTs are available in this area due to the ethical considerations of randomising patients to an arm without surveillance. Those few RCTs that do exist were excluded as they were in patients with HBV and not all the population had cirrhosis. The GDG chose to only include in the protocol studies in which more than 85% of the population investigated had cirrhosis. The GDG reviewed the search strategy and discussed the excluded papers. The main reasons for exclusion were the absence of, or small proportion of, patients with cirrhosis and/or the presence of more than 15% of patients with HBV. A number of studies were excluded if they had not adjusted their outcomes for other confounding factors (such as the severity of liver disease or the frequency of decompensations).
Other considerations	The GDG discussed that the evidence for the benefit of HCC surveillance is dependent on the effectiveness of current surveillance strategies. For example, the accuracy of ultrasound may be reduced in patients with NASH or obesity and will also be dependent on the extent of the cirrhosis. The quality of the ultrasound scan is also operator-dependent. However, the GDG agreed that ultrasound is still the favoured option for surveillance. Biomarkers such as AFP can aid diagnosis of HCC, but it is thought that only around 60% of HCCs are AFP-secreting. The accuracy of AFP would also be reduced in certain aetiologies such as alcohol-related cirrhosis. The GDG also felt that an important clinical aspect of ultrasound surveillance was

	not only the detection of HCC, but also the assessment for other complication of cirrhosis, such as portal hypertension, portal vein thrombus and ascites. It was discussed that surveillance for HCC could have further benefit if used as part of an integrated package of surveillance for other complications of cirrhosis. The GDG noted that recent changes in HCV treatment may have some impact on the necessity of surveillance in the population with HCV who have achieved viral clearance. At the current time the level of impact this may have on surveillance programmes was unknown. The GDG noted that the current system is an 'ad hoc' clinician-initiated surveillance programme. There was a general consensus that if surveillance is thought to be effective then it should be run in an organised fashion at an institutional level to ensure all patients are offered the opportunity to take part. The GDG noted recent statements from the UK Royal College of Radiologists and the United States National Cancer Institute highlighting the lack of High quality evidence to support routine HCC surveillance. However, it is very unlikely that an RCT of surveillance versus no surveillance will be undertaken on which to base recommendations. Overall, the GDG agreed that there was evidence that surveillance can detect HCC at an earlier stage and would give patients the opportunity to receive potentially curative treatment. The risks and benefits of surveillance should be discussed with the patient. The GDG noted that there were certain groups that would not benefit from surveillance as they are not eligible for treatment or for transplantation, such as those on an end of life strategy.

9 Surveillance for the detection of varices

9.1 Introduction

Variceal bleeding occurs in 25–40% of patients with cirrhosis and each bleeding episode is associated with a 10–30% mortality rate.[50] Consequently, prevention of variceal bleeding is an important goal in the management of patients with cirrhosis. Therefore, it is important that people with cirrhosis at risk for variceal bleeding should be identified as early as possible.

Clinical signs and symptoms such as ascites, thrombocytopenia, splenomegaly and Child-Pugh class do not adequately predict which patients will develop variceal bleeding.[35,147,168] Thus, the American College of Gastroenterology (ACG) and the American Association for the Study of Liver Disease (AASLD) have published guidelines recommending that all people with cirrhosis should be screened for the presence of varices using oesophagogastroduodenoscopy (OGD).[89,90]

Comparison of endoscopic surveillance versus no surveillance was excluded from the protocol as, due to the high incidence of oesophageal and gastric varices in people with cirrhosis and the subsequent high risk of bleeding and bleeding-related mortality, the GDG considered that all people with cirrhosis should undergo endoscopic surveillance. The question was therefore to find the most clinically and cost-effective frequency of endoscopic surveillance for the detection of the first occurrence of oesophageal or gastric varices in people with cirrhosis. Implicit in the investigation of surveillance frequency is that patients whose varices are detected earlier can be treated earlier with potentially better patient outcomes and a better chance of survival.

9.2 Review question: How frequently should surveillance testing using endoscopy be offered for the detection of oesophageal varices and isolated gastric varices in people with cirrhosis?

For full details see review protocol in Appendix C.

Table 66: PICO characteristics of review question

Population	Adults and young people (16 and over) with confirmed cirrhosis, without varices and who have not already been started on primary prophylactic therapy for the prevention of variceal bleeding.
Intervention	Intervention: endoscopy at: • Baseline only • Yearly • Every 2 years • Every 3 years
Comparison	Comparison: endoscopy at: • Baseline only • Yearly • Every 2 years • Every 3 years Exclusions: Surveillance endoscopy versus no surveillance endoscopy
Outcomes	**Critical outcomes:** • Survival (time-to-event) or mortality at 5 years

	• Free from variceal bleeding (time-to-event) or variceal bleeding at 5 years • Health-related quality of life **Important outcomes:** • Free from varices (time-to-event) • Occurrence of moderate or large varices • Size of varices • Number receiving prophylactic treatment (beta-blockers or EVL)
Study design	RCTs, systematic reviews of RCTs, observational studies, systematic reviews of observational studies

9.3 Clinical evidence

Randomised and observational studies comparing different intervals of surveillance testing in detecting varices in people with cirrhosis were searched for. No relevant clinical studies comparing different frequencies of endoscopic surveillance for the detection of varices were identified. For exclusion reasons see Appendix L.

9.4 Economic evidence

9.4.1 Unit costs

See Table 64 in Appendix N.

9.4.2 Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

9.4.3 New cost-effectiveness analysis

Original cost-effectiveness modelling was undertaken for this question using the NGC liver disease pathway model developed for this guideline. A summary is included here. An evidence statement summarising the results of the analysis can be found below. The full analysis can be found in Appendix N.

9.4.3.1 Aim and structure

The aim of the health economic modelling for this question was to determine the optimal frequency of oesophageal varices surveillance. This was achieved by using the original lifetime diagnostic health state transition (Markov) model (see Section 6.4.3 and Appendix N) which followed the NICE reference case[146] and by comparing overall cost and QALYs of the cirrhosis tests of preference in three different scenarios; annual, 2-yearly and 3-yearly varices surveillance.

Patients who had an endoscopy and were identified as having medium to large varices were immediately offered a band ligation procedure. Patients that received this procedure ran a lower risk of variceal bleeding. Therefore, in the economic model, the patient benefit from receiving more frequent endoscopies was the reduced time spent under the increased risk of bleeding from untreated varices.

To determine the most cost-effective surveillance frequency, incremental cost-effectiveness ratios (ICERs) were calculated across the available options. Base case results below were obtained through the probabilistic analysis to take combined parameter uncertainty into account.

9.4.3.2 Results

Table 67: ICERs comparing annual and 2-yearly surveillance against 3-yearly surveillance

Aetiology	Frequency	ICER	Cirrhosis test used for the comparison
NAFLD	2 years	Dominated	TE at >15.0
	1 year	£122,413	
ALD	2 years	£63,167	TE at 11.0 – <13.0
	1 year	£120,390	
HBV –antigen	2 years	£54,408	TE at 11.0
	1 year	Dominated	
HBV +antigen	2 years	£36,552	TE at 11.0
	1 year	£48,430	
HCV genotype 1	2 years	£75	Liver biopsy
	1 year	Dominated	
HCV genotype 3	2 years	Dominant	Liver biopsy
	1 year	Dominated	

Annual surveillance was not cost-effective compared to 3-yearly surveillance for any of the model cohorts with the ICERs either exceeding £45,000 per QALY gained or showing it being dominated by the 3-year frequency option (more costly and less effective). Two-yearly surveillance was cost-effective compared to 3-yearly surveillance at a cost-effectiveness threshold of £20,000 per QALY in the 2 HCV cohorts.

In the deterministic sensitivity analysis, changes in the surveillance costs or the RR applied on the bleeding probability had considerable effect on the ICERs of the higher frequencies. However with the base case ICERs of the deterministic analysis being far beyond the £20,000 threshold, any reductions in the ICERs made 2-yearly surveillance cost-effective only for the ALD cohort.

9.5 Evidence statements

9.5.1 Clinical

- No relevant clinical studies were identified.

9.5.2 Economic

- One original cost-utility analysis that compared annual, 2-yearly and 3-yearly surveillance for the detection of varices in people with cirrhosis at a cost-effectiveness threshold of £20,000 per QALY gained found that:
  - Annual surveillance was not cost-effective compared to 3-yearly surveillance (ICERs: £48,430–122,413 per QALY gained or dominated).
  - 2-yearly surveillance was cost-effective compared to 3-yearly surveillance in people with hepatitis C (ICERs: £75per QALY gained or dominant).
  - 2-yearly surveillance was not cost-effective compared to 3-yearly surveillance in people with NAFLD and advanced fibrosis, ALD or hepatitis B (ICERs: £36,552–63,167 per QALY gained or dominated).

This analysis was assessed as directly applicable with minor limitations.

9.6 Recommendations and link to evidence

Recommendations	**19. After a diagnosis of cirrhosis, offer upper gastrointestinal endoscopy to detect oesophageal varices.** **20. For people in whom no oesophageal varices have been detected, offer surveillance using upper gastrointestinal endoscopy every 3 years.**
Relative values of different outcomes	For comparison of surveillance frequency the GDG assessed the critical outcomes of all-cause mortality, freedom from variceal bleeding and health-related quality of life. The other important outcomes for decision-making were freedom from varices, the development of moderate or large varices, the size of varices and the number of patients receiving prophylactic treatment (beta-blockers or endoscopic variceal band ligation).
Trade-off between clinical benefits and harms	The GDG considered that all people with cirrhosis should be offered endoscopic surveillance, due to the high incidence of oesophageal and gastric varices in people with cirrhosis, and the subsequent high risk of bleeding and bleeding-related mortality. Therefore, the comparison of endoscopic surveillance versus no surveillance was excluded from the protocol. The question was therefore to find the most clinically and cost-effective frequency of endoscopic surveillance for the detection of the first occurrence of varices in people with cirrhosis. No clinical evidence was identified from RCTs or observational studies comparing different frequency of surveillance and the effect on patient outcomes. The GDG discussed the potential harms of monitoring for varices too infrequently, including not identifying people with varices soon enough to give prophylactic treatment for the prevention of variceal bleeding. There is also a high incidence of bleeding-related mortality in this population. As the GDG has also made a recommendation that people with medium or large oesophageal varices should be given prophylactic band ligation to prevent bleeding (see recommendation 22), it was agreed that people with cirrhosis should undergo endoscopy at the time of diagnosis and that those with no evident varices initially should undergo regular surveillance. The GDG also discussed the potential harms of undergoing endoscopy, including discomfort, aspiration, broken teeth and in rare instances perforation of the oesophagus and rupture of varices and, very rarely, death. The GDG agreed that the benefits of endoscopic surveillance outweighed the harms. A recommendation was made that people newly diagnosed with cirrhosis should undergo surveillance and that, in those with no evident varices, surveillance should be continued at 3-yearly intervals until detection of varices of any size. It was highlighted that if signs and symptoms indicating the presence of varices developed (for example, clinical or laboratory evidence suggestive of portal hypertension, such as ascites, splenomegaly or thrombocytopaenia), then endoscopy should be performed earlier.
Trade-off between net clinical effects and costs	No published economic evidence was identified. In the original economic modelling conducted for this guideline (see Appendix N), the impact of varying the frequency of testing people with cirrhosis but without varices from every 3 years to every 2 years or every year was investigated. This showed that the increased frequency had small health benefits with significant additional costs. Two-yearly testing was found to be cost-effective in the HCV cohorts at a cost-effectiveness threshold of £20,000 per QALY, but not for the other cohorts. Annual surveillance was not cost-effective for any of the populations. The GDG acknowledged that there was significant variation in the cost-effectiveness of more frequent testing between the population groups, with associated

	uncertainty, and that a consistent strategy for all groups would be beneficial. It therefore agreed not to recommend testing more frequently than every 3 years for any group.
Quality of evidence	No clinical evidence was identified from RCTs or observational studies for this review question.
Other considerations	The GDG noted that both the EASL and AASLD guidelines used consensus to make recommendations on surveillance for varices in people with cirrhosis. The GDG agreed that endoscopic surveillance should be performed by a person experienced in interventional endoscopy. Therefore, if medium or large varices are detected, the band ligation procedure can be performed at the same time to avoid the need for a second procedure. The GDG also discussed that there is a certain degree of inter-observer variability associated with the procedure[17,26,48], highlighting the need for endoscopists to be appropriately trained and experienced.

10 Prophylaxis of variceal haemorrhage

10.1 Introduction

Oesophageal varices, which develop as a result of portal hypertension, are found in approximately 30% of people with cirrhosis at the time of diagnosis. People with cirrhosis without varices at the time of their diagnosis develop them at a rate of 5% (95%CI: 0.8–8.2%) at 1 year and 28% (21.0–35.0%) at 3 years.[134] The factors precipitating variceal haemorrhage are still not clear but it is recognised that the risk of bleeding is related not only to the size of the varices (>5 mm) but also to the severity of liver disease and, in people with alcohol-related cirrhosis, whether or not they continue to drink. Once varices are present, they tend to enlarge; thus of people with small varices at the outset 12% (5.6–18.4%) will have large varices at 1 year and 31% (21.2–40.8%) at 3 years[134], resulting in a higher risk of bleeding. The estimated 2-year incidence of bleeding is approximately 24%[46] and most episodes of bleeding from varices (70%) occur within 2 years of diagnosis.

Although the in-hospital mortality associated with variceal bleeding has decreased in recent years due to improvements in endoscopic therapy and the use of antibiotic prophylaxis, the reported mortality rate, ranging from 12% to 44%, is still substantial. The risk of death within 6 weeks of the initial variceal haemorrhage is related closely to the severity of liver disease, as determined by the Child-Pugh grade: mortality is <10% in Child-Pugh class A compared to >32% in those in Child-Pugh class C.[28]

As approximately 30% of people with cirrhosis with oesophageal varices develop bleeding and 12–44% die as a result of the first bleed, prophylactic regimens to prevent bleeding have been developed. Non-selective beta-blocker therapy has been the main pharmacological approach for the primary prophylaxis of variceal haemorrhage because these drugs reduce azygos blood flow and variceal pressure.[46] Endoscopic variceal ligation (EVL) has been advocated as an option for primary prophylaxis.[100] Although EVL is a relatively simple endoscopic procedure, during which elastic bands are placed around the varices, repeated endoscopies are required both to achieve eradication of varices by EVL and for surveillance for variceal recurrence. Since there are 2 different treatment approaches, the GDG decided to examine the clinical and cost- effectiveness of non-selective beta-blockers and endoscopic band ligation both individually and head-to-head for the primary prevention of bleeding in patients with oesophageal varices due to cirrhosis.

10.2 Review question 1: What is the clinical and cost-effectiveness of non-selective beta-blockers for the primary prevention of bleeding in people with oesophageal varices due to cirrhosis?

Review question 2: What is the clinical and cost-effectiveness of endoscopic band ligation for the primary prevention of bleeding in people with oesophageal varices due to cirrhosis?

Review question 3: What is the clinical and cost-effectiveness of non-selective beta-blockers compared with endoscopic band ligation for the primary prevention of bleeding in people with oesophageal varices due to cirrhosis?

For full details see review protocol in Appendix C.

Table 68: PICO characteristics of review question

Population	Adults and young people (16 years and over) with endoscopically verified oesophageal varices that have never bled, with cirrhosis as the underlying cause. Population stratification • Size of varices (small) • Size of varices (medium or large)
Interventions	• Oral non-selective beta-blockers: carvedilol, propranolol • Endoscopic band ligation • Placebo or no intervention
Comparisons	• Oral non-selective beta-blockers versus placebo or no intervention • Endoscopic band ligation versus no intervention • Oral non-selective beta-blockers versus endoscopic band ligation
Outcomes	• Health-related quality of life • Survival (with or without transplant) • Free from primary variceal bleeding • Hospital admission • Hospital length of stay • Primary upper gastrointestinal bleeding (irrespective of bleeding source) • Bleeding-related mortality • Adverse events: fatigue
Study design	RCTs, systematic reviews of RCTs

10.3 Clinical evidence

We searched for randomised trials comparing the effectiveness of endoscopic band ligation or oral non-selective beta-blockers in the primary prophylaxis of variceal bleeding in people with oesophageal varices due to cirrhosis. The non-selective beta-blockers considered were propranolol and carvedilol. Propranolol is licensed in the UK for the prophylaxis of variceal bleeding in portal hypertension (up to a maximum of 160 mg twice daily). Carvedilol is not licensed in the UK for this particular indication however the GDG wanted to include this evidence in the review, as carvedilol is currently widely used for this indication. Only 2 studies in the comparison of endoscopic band ligation versus non-selective beta-blockers used carvedilol. As evidence for non-selective beta-blockers was combined within the same class, any recommendation made would be for non-selective beta-blockers as a class and not for either propranolol or carvedilol individually.

For comparison 1 (non-selective beta-blockers versus placebo or no intervention), 9 papers reporting 5 studies were included in the review.[8,42,93,153-155,158,159,193] All studies used propranolol as the intervention and the control group received placebo. Two of the studies[153,159] were in populations of people with medium or large oesophageal varices and were analysed in this stratum. One study[193] was in a population of people with small oesophageal varices and was analysed in this stratum. Two studies[8,42] were in people with varices of all sizes, however they provided a subgroup analysis of small versus medium and large varices for the following outcomes and these data were analysed within these strata (Andreani 1990: variceal bleeding and upper gastrointestinal bleeding; Conn 1991: variceal bleeding). One study was excluded because it only provided data for this comparison in people with varices of all sizes with no subgroup analysis[207] (see excluded studies list in Appendix L). The study characteristics are summarised in Table 69 below. Evidence from these studies is summarised in the clinical evidence summaries below (Table 72 and Table 73). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

For comparison 2 (endoscopic band ligation versus no intervention), 5 RCTs were included in the review.[118,125,191,221,232] There was some variation between the studies in the number of bands used in each ligation session, and the frequency of band ligation sessions which ranged from every 1–3 weeks (summarised in Table 70). In all studies, band ligation was performed until eradication of varices or until varices were too small to ligate. Sarin 1996[191] included 6 people with another underlying cause of portal hypertension and this study was downgraded for population indirectness. Four of the studies[118,125,191,221] were in populations of people with medium or large oesophageal varices and were analysed in this stratum. The final study[232] was in people with varices of all sizes, however this study did provide a subgroup analysis of small versus medium and large varices for the outcome of upper gastrointestinal bleeding, and this outcome was analysed in these separate strata. Evidence from these studies is summarised in the clinical evidence summaries below (Table 74 and Table 75). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

For comparison 3 (endoscopic band ligation versus non-selective beta-blockers), 1 Cochrane systematic review[88] and 2 RCTs[198,203] were included in the review. The Cochrane review included 7 conference abstracts using the published data only. Evidence from these 7 studies has been included here however conference abstracts have not been routinely included elsewhere in this review. In total, 25 papers reporting 20 studies were included for this comparison in this review.[1,4,6,38,51,53,54,82,88,106,107,117,127,149,161,166,192,194-196,198,203,206,227,234] All the included studies were analysed in the medium to large varices stratum, no studies were identified for the small varices stratum. The study characteristics are summarised in Table 71 below. Evidence from these studies is summarised in the clinical evidence summaries below (Table 76 and Table 77). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

The Cochrane review was partially included because the review did not include all the outcomes specified in our protocol. Consequently, the papers included in the Cochrane were examined individually to extract the additional outcomes: survival as a time-to-event outcome, freedom from variceal bleeding as a time-to-event outcome and hospital admissions. Where the individual studies reported survival and freedom from variceal bleeding as a time-to-event outcome, this was reported instead of mortality and variceal bleeding as dichotomous outcomes. For the outcome of upper gastrointestinal bleeding, if the individual study only reported variceal bleeding, this was also used for the upper gastrointestinal bleeding outcome in the Cochrane review. However, some studies report upper gastrointestinal bleeding (not including variceal bleeding) and some report upper gastrointestinal bleeding from varices and other sources and, where reported, these numbers were used. Analysis of the additional RCTs and of the evidence in comparisons 1 and 2 was also performed in this way.

The Cochrane review population comprised of people with oesophageal varices due to portal hypertension (not specifically portal hypertension due to cirrhosis as an inclusion criterion). All the included studies were check individually and all but 1 study only included people with cirrhosis as the underlying cause of portal hypertension. Sarin 1999[192] included 7 people with another underlying cause of portal hypertension and this study was downgraded for indirectness. The Cochrane review specified 'high risk' varices. In order to confirm all the studies fell into our predefined stratum of medium or large varices, studies were checked individually. All but 2 studies specifically mentioned the criteria which would fall into the category of medium or large varices. These 2 studies (Chen 1998 and Abdelfattah 2006) did not specify the size of the varices. Abdelfattah 2006 did specify high risk varices and was therefore included in the medium or large varices stratum. Chen 1998 did not specify the size or risk of varices and therefore and was included in the medium or large varices stratum but was downgraded for indirectness. Two studies specified that they included people with cirrhosis on the transplant waiting list (Gheorghe 2002)[82];(Norberto 2007)[149].

Two studies used carvedilol[198,234] and the remaining studies used propranolol. One study (Lo 2004)[124] was removed from the Cochrane review analysis as it used nadolol as the non-selective beta-blocker. Nadolol was excluded from the review protocol as it is not licensed or widely used in the UK for this indication.

No evidence was identified for the outcome of quality of life for any of the 3 comparisons.

Table 69: Summary of studies included in the review: non-selective beta-blockers versus placebo or no intervention

Study	Intervention and comparison	Population	Outcomes	Comments
Andreani 1990	Propranolol (n=43)/placebo (n=41) Propranolol twice daily. Dose titrated to achieve a 25% reduction in resting HR. Vitamin K (10 mg) twice daily	Cirrhosis, presence of oesophageal varices on endoscopy regardless of size; no history of gastrointestinal bleeding by rupture of oesophageal varices. Cirrhosis diagnosis proven by histological examination (or if unavailable, on the basis of clinical or lab test results, regardless of origin)	Variceal bleeding (D) gastrointestinal bleeding (D)	Inclusion: all sizes of varices with subgroup analysis for size of varices
Conn 1991	Propranolol (n=51)/placebo (n=51) Dose determined prior to randomisation by the response of HVPG to increasing doses of propranolol during hepatic vein catheterisation Placebo details not reported	Cirrhosis, endoscopically documented oesophageal varices (all sizes) and portal hypertension who had not previously bled from oesophageal varices or from an unknown upper gastrointestinal site Cirrhosis diagnosis approximately 50% had histological confirmation	Variceal bleeding (D)	Inclusion: all sizes of varices with subgroup analysis for size of varices
Pagliaro 1989C	Propranolol (n=85)/placebo (n=89) Propranolol twice daily. Dose titrated to achieve a 25% reduction in resting HR. Vitamin K (10 mg) twice daily	Cirrhosis and large oesophageal varices (F3 varices occupying more than a third of the oesophageal lumen); no previous upper gastrointestinal bleeding. Cirrhosis diagnosis biopsy-proven 43%	Survival (TTE) Variceal bleeding (D) gastrointestinal bleeding (D) Bleeding mortality (D)	Medium or large varices stratum
Pascal 1989	Propranolol (n=118)/placebo (n=112) Starting dose 20 mg of conventional formulation twice daily. Titrated up to 160 mg or 320 mg of long-acting once daily to achieve a 20–25% reduction in resting HR Identical placebo once daily	Cirrhosis and Child-Pugh score <14; grade II or II (medium or large) oesophageal varices at endoscopy. Cirrhosis confirmed by liver biopsy or biochemical and clinical data	Survival (TTE) gastrointestinal bleeding (D) Bleeding mortality (D)	Medium or large varices stratum
Sarin 2013	Propranolol (n=77)/placebo (n=73) Starting dose 20 mg twice daily. Incremental dosing used to achieve target HR (dose increased every alternate day to achieve a target HR of 55/minute or to the maximum	Cirrhosis, small oesophageal varices (grade 1 or 2 by Conn's classification or small as per Baveno); no history of variceal bleeding. Cirrhosis diagnosis clinical, radiological or histological	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding (D)	Small varices stratum

Study	Intervention and comparison	Population	Outcomes	Comments
	dose of 360 mg/day). Placebo details not reported			

TTE: outcome reported as a time-to-event outcome
D: outcome reported as a dichotomous outcome

Table 70: Summary of studies included in the review: endoscopic band ligation versus no intervention

Study	Intervention and comparison	Population	Outcomes	Comments
Lay 1997	Band ligation (n=62)/no intervention (n=64) EVL: each varix was ligated with 1–3 rubber bands (adapted endoscopic ligating device, Bard Interventional Products, Billerica, MA). Endoscopic treatment was performed weekly until the oesophageal varices were eradicated.	Cirrhosis, high risk oesophageal varices (all patients had blue varices of F2 or F3 size with red colour signs), no known previous gastrointestinal bleeding.	Survival (TTE) Variceal bleeding (TTE) gastrointestinal bleeding (D)	Medium or large varices stratum
Lo 1999	Band ligation (n=66)/no intervention (n=67) EVL: each varix ligated with 1–2 rubber bands (Bard Interventional Products, Billerica, MA, USA). Performed at intervals of 3 weeks until all varices were obliterated or too small to be ligated. Sucralfate granules 1 g 4 times per day were administered to patients during the course of EVL treatment. After obliteration, patients in the treatment group underwent follow-up endoscopy every 3 months. Repeat EVL was performed in case of variceal recurrence.	Cirrhosis and portal hypertension, endoscopically assessed high risk oesophageal varices (F2 or F3 , associated with a moderate degree of red colour signs)	Survival (TTE) Variceal bleeding (TTE) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D)	Medium or large varices stratum
Sarin 1996	Band ligation (n=35)/no intervention (n=33) EVL: 1 or 2 bands applied at each variceal column at regular 7–10 day intervals until total variceal obliteration achieved (no variceal column visible) or it was not possible to suck in a varix for band ligation (grade 1 varices). Endoscopy performed every	Portal hypertension; high risk varices (included blue varices of F2 or F3 size with at least 1 of the red colour signs) without previous history of upper or lower gastrointestinal bleeding. Cirrhosis not an inclusion criterion. 6/68 had	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D)	Medium or large varices stratum

Study	Intervention and comparison	Population	Outcomes	Comments
	3 months after the eradication of varices.	causes of portal hypertension other than cirrhosis		
Svoboda 1999	Band ligation (n=52)/no intervention (n=50) Ligation performed using an endoscopic ligation device (Suction oesophageal varices ligator, Pauldrach Medical, Germany). Later multiband ligators were also used (Wilson-Cook medical, USA or Microvasive, USA). Three sessions at 2-week intervals, and then every month until the varices were too small to treat. Repeated if recurrence of varices occurred. In each session the largest number possible (up to 6) of elastic bands were positioned in the distal oesophagus. All patients given ACE inhibitor enalapril (later quinapril) 2× 5–10 mg orally to decrease portal pressure.	Cirrhosis, oesophageal varices of grades III and IV; oesophageal varices of grade II with signs of high risk (Paquet's classification).	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D)	Medium or large varices stratum
Triantos 2005	Band ligation (n=25)/no intervention (n=27) At least 1 band on each varix (Multiband ligator 6 shooter, Wilson-Cook, Ireland). Subsequent sessions at 14-day intervals until the varices were too small to ligate (no effect of suction).	Cirrhosis and varices of any size (assessed endoscopically); no prior bleeding from portal hypertensive sources.	gastrointestinal bleeding (D)	Inclusion: all sizes of varices with subgroup analysis for size of varices

TTE: outcome reported as a time-to-event outcome
D: outcome reported as a dichotomous outcome

Table 71: Summary of studies included in the review: endoscopic band ligation versus non-selective beta-blockers

Study	Intervention/comparison	Population	Outcomes	Comments
Abdelfattah 2003 Abstract Egypt follow-up: mean 30	EVL (n=44)/propranolol (n=66) No further intervention details given	Cirrhosis and grade II or III oesophageal varices that had never bled	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding – variceal only (D)	Medium or large varices stratum

Study	Intervention/comparison	Population	Outcomes	Comments
months (range not reported)				Withdrawal in propranolol arm due to side effects n=22
Abdelfattah 2006 Abstract Egypt follow-up: 18 to 24 months (mean not reported)	EVL (n=51)/propranolol (n=52) No further intervention details given	People with cirrhosis with risky varices (primary prophylaxis)	Mortality (D) gastrointestinal bleeding (D)	Medium or large varices stratum
Chen 1998 Abstract Taiwan follow-up: mean 12 months	EVL (n=26)/propranolol (n=30) EVL performed at 2–3 week intervals using Microvasive speedband ligator until complete eradication of varices (recurrent varices treated with repeat EVL) Propranolol given to reduce HR by 25% All patients: patients who bled received EVL	Oesophageal varices (size and risk not defined) and no prior gastrointestinal bleeding	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding – variceal only (D)	Medium or large varices stratum Withdrawal in propranolol arm n=2
De 1999 Full paper India follow-up: mean 18 months (range not reported)	EVL (n=15)/propranolol (n=15) EVL weekly to fortnightly until obliteration Propranolol starting dose 40 mg 3 times daily then titrated to achieve a 25% reduction in pulse rate	Cirrhosis and III to IV oesophageal varices and no history of bleeding (HVPG ≥12 mmHg)	Variceal bleeding (D) gastrointestinal bleeding (D)	Medium or large varices stratum
De La Mora 2000 Abstract Mexico mean not reported	EVL (n=12)/propranolol (n=12) EVL until eradication Propranolol administered in increasing dose until a HR of 60 or a	Cirrhosis and no bleeding history with large, high risk varices	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding – variceal only (D)	Medium or large varices stratum

Study	Intervention/comparison	Population	Outcomes	Comments
(range1 to 19 months)	20% decrease from baseline occurred.			
Drastich 2011 Full paper Czech Republic follow-up: median 10 months.	EVL (n=40)/propranolol (n=33) EVL using multiband ligator device (Six shooter, Wilson-Cook), up to 6 bands placed in each session. Performed at 2 week intervals until oesophageal varices eradicated (complete disappearance or too small to be ligated). EVL continued in patients with recurrence. Propranolol starting dose 20 mg twice daily. Adjusted in 20–40 mg increments at weekly intervals to achieve a HR reduction of 25% (not below 55 bpm or systolic BP <80 mmHg).	Portal hypertension due to liver cirrhosis and large oesophageal varices (>5 mm). Cirrhosis diagnosis on the basis of clinical, ultrasonographic and biochemical examination (if necessary, liver biopsy to confirm). Excluded non-cirrhotic cause of portal hypertension; history of gastrointestinal bleeding, sclerotherapy, EVL or shunt; malignant disease; gastric or duodenal ulcer; congestive heart failure; renal insufficiency; treatment with beta-blockers, nitrates, ACE inhibitors or verapamil; antiviral therapy; AV block, sick-sinus syndrome, bradycardia; decompensated diabetes; pregnancy, lactation.	Mortality (TTE) Variceal bleeding (TTE) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D) Weakness (D)	Medium or large varices stratum Variceal bleeding and survival as TTE data not reported in Cochrane so extracted separately
Gheorghe 2002 Abstract Romania mean 15 months (range not reported)	EVL (n=25)/propranolol (n=28) EVL performed using a six-shutter Saed Ligator at 3-week intervals until variceal obliteration achieved Propranolol – detail not reported	People with cirrhosis on the liver transplantation waiting list. High risk oesophageal varices (varices >5 mm, red signs, Child-Pugh B or C). No history of variceal bleeding	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding – variceal only (D)	Medium or large varices stratum
Jutabha 2005 Full paper USA mean 12 months (range 1 to 61 months)	EVL (n=31)/propranolol (n=31) EVL using a multiband ligating device (Saeed Six-Shooter). Follow-up banding performed at 4–5 weeks. EVL performed until obliteration or reduction to a small size and EVL not possible. Recurrent varices also underwent EVL.	Cirrhosis and large (>5 mm or Paquet grade 3-4) or high-risk (medium size 3–5mm with red signs) non-bleeding varices. Cirrhosis was biopsy-proven or clinically evident. No previous upper gastrointestinal bleeding; no prior sclerotherapy or EVL, TIPS or surgical; no current beat-blocker; life expectancy at least	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D)	Medium or large varices stratum

Study	Intervention/comparison	Population	Outcomes	Comments
	Propranolol – either long acting at starting dose of 80 mg and seen weekly to adjust dose by 80 mg increments (max 400mg) to reduce HR by 25% OR 40 mg twice daily and increased every 2 weeks by 40–80 mg as tolerated. In EVL group, patients prescribed proton pump inhibitors once daily until obliteration of varices.	24 months. Excluded serious recurrent or outgoing comorbid illness and contraindication to beta-blockers. Other exclusion criteria detailed including moderate or large gastric or duodenal ulcers, large-volume or tense ascites or HCC.		
Lay 2006 Full paper China and Taiwan follow-up: mean 35 months (range 1 to 72 months)	EVL (n=50)/propranolol (n=50) EVL with 1–3 rubber bands on each varix until the varices were too small to ligate (max 10 rubber bands per session). Propranolol at a starting dose of 40 mg twice daily. Increased by 10 mg twice daily until either a reduction in the resting HR of 20% or to the maximum dose.	Cirrhosis and oesophageal varices (risk score from Beppu et al., corresponded to all patients having blue varices of F2 or F3 size with at least 1 red colour sign) at high risk and no previous upper gastrointestinal bleeding. Excluded: other disease reducing life expectancy	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding (D) Bleeding mortality (D)	Medium or large varices stratum Survival data (survival after first bleed) not used as starting point for life-table was bleeding. Free from bleeding data not used (all bleeding not variceal bleeding) 2 patients in EVL and 3 patients in propranolol

Study	Intervention/comparison	Population	Outcomes	Comments
				lost to follow-up. 3 patients in each group non-compliant.
Lui 2002 Full paper UK mean 20 months (range 1 to 48 months)	EVL (n=44)/propranolol (n=66) EVL performed every 2 weeks until eradication (complete or grade I only) with single or multiband. Further EVL if grade II or larger varices recurred. Propranolol starting dose of 40mg twice daily and incremental dosing used to achieve the target daily dose of 160mg.	Cirrhosis and grade II or III oesophageal varices that had never bled. Cirrhosis diagnosed based on histology or a combination of radiology, laboratory and clinical parameters. Excluded if <18 or >75 years; advanced systemic illness; non-cirrhotic cause of portal hypertension; on vasoactive agents; contraindications to beta-blockers.	Mortality (TTE) Variceal bleeding (TTE) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D)	Medium or large varices stratum Variceal bleeding and survival as TTE data not reported in Cochrane so extracted separately
Norberto 2007 Fully paper Italy 14 months (range not reported).	EVL (n=31)/propranolol (n=31) EVL performed using a multiband ligator with 6 or 7 bands (Six shooter, Wilson-Cook). Performed every 2 weeks until varices completely eradicated. EVL performed again on recurrent varices. Propranolol started at 10mg twice daily and increased by 20mg/day until a 25% reduction in HR. Maximum dose 160 mg/day. EVL group also received proton pump inhibitors	Cirrhosis and studied for liver transplant. Oesophageal varices F3 or F2 blue with red signs according to Beppu, and no previous bleeding. Cirrhosis diagnosed on the basis of clinical, biochemical or histological analysis. Excluded if <18 or >85 years; gastric varices; previous endoscopic, radiological or surgical treatment of varices; HCC; portal vein thrombosis; heart, respiratory or renal failure; contraindications to beta-blockers; treatment with nitrates, Ca antagonists or anti-arrhythmic drugs; pregnancy; neoplasia.	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D)	Medium or large varices stratum
Perez-Ayuso 2010 Full paper	EVL (n=39)/propranolol (n=36) EVL performed at 3 week intervals	Cirrhosis, high-risk oesophageal varices (large or medium size, 3–5 mm, with red colour signs), no	Mortality (TTE) Variceal bleeding (D)	Medium or large varices

Study	Intervention/comparison	Population	Outcomes	Comments
Mexico follow-up 55 months (range 1 to 119 months).	until eradication (absence of ligable varices). Up to 6 bands placed in each session using a multiband ligator (Six shooter, Wilson-Cook). Religation performed if at least 1 varix >5mm reoccurred. Propranolol starting dose 20 mg twice daily and increased every 3 days to achieve a 25% reduction in heart rate, to a heart rate <55 bpm, to a systolic blood pressure <90 mmHg or a maximum of 320 mg daily.	history of bleeding from varices and no current treatment with beta-blockers. Cirrhosis diagnosed on the basis of clinical, biochemical, histological or ultrasonographic evidence. Excluded younger than 18 and older than 70; big gastric varices, evidence of portal thrombosis, malignancy, contraindication to beta-blockers, previous variceal endoscopic treatment, TIPS, surgical shunt or renal failure.	gastrointestinal bleeding (D) Bleeding mortality (D)	stratum Survival as TTE data not reported in Cochrane so extracted separately
Psilopoulos 2005 Full paper Greece follow-up: 28 months (range 0.5 to 52 months).	EVL (n=30)/propranolol (n=30) EVL performed using multiband ligation device (Speedband or Six shooter). 1 or 2 bands applied to each varix, and up to 6 bands per session. Sessions repeated every 2–3 weeks until variceal eradication or too small to be ligated. Propranolol received 40 mg and dose adjusted to achieve 25% reduction in HR. EVL patients treated with proton pump inhibitors until variceal eradication.	Portal hypertension caused by cirrhosis, grade II or III oesophageal varices (F2 or F3 according to Beppu) with red signs and no history of variceal bleeding. Excluded treatment with nitrates or beta-blockers; <20 or >70 years; gastric or ectopic varices; severe comorbidity; refractory ascites; HCC; marked jaundice; contraindications to beta-blockers; history of EVL, sclerotherapy or TIPS or shunts.	Mortality (TTE) Variceal bleeding (TTE) gastrointestinal bleeding (D) Bleeding mortality (D)	Medium or large varices stratum Survival as TTE data not reported in Cochrane so extracted separately
Sarin 1999 Full paper India mean not reported (range 0.5 to 18	EVL (n=46)/propranolol (n=44) EVL performed using a single rubber band for each varix and as many bands as possible in each session (average 3–9). Performed every week until obliterated or reduced to	Portal hypertension and large grade 3 (3–6 mm) or 4 (>6 mm) varices with no history of bleeding. Cirrhosis was diagnosed on the basis of clinical, biochemical, histologic or ultrasonographic evidence (cirrhosis was not an entry criterion and 6 patients had extrahepatic portal vein	Mortality (D) Variceal bleeding (TTE) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D)	Medium or large varices stratum Variceal bleeding as

Study	Intervention/comparison	Population	Outcomes	Comments
months)	grade I size. EVL repeated if varices recurred and became grade II or larger. Propranolol started with 40 mg and increased the dose by 2–40 mg/day until a 25% reduction in HR achieved.	obstruction and 1 patient had non-cirrhotic portal fibrosis)	Lethargy (D) Hospitalisation (D)	TTE data not reported in Cochrane so extracted separately
Schepke 2004 Full paper Germany follow-up: mean 34 months (0,1 to 73 months).	EVL (n=75)/propranolol (n=77) EVL performed using a multiband ligator (Six shooter, Wilson-Cook), up to 10 bands in each session. Performed weekly until eradication. Religation preformed when at least 1 varix >5 mm recurred. Propranolol started at 40 mg twice daily and increased by 10 mg twice daily until HR reduction of 20% or to the maximum dose.	Cirrhosis and 2 or more oesophageal varices >5 mm, no previous bleeding and a Child-Pugh score below 12. Cirrhosis diagnosis made on histology or unequivocal clinical, sonographic and laboratory findings. Excluded prehepatic portal hypertension, bradycardia, systolic BP <100 mmHg, contraindications to beta-blockers, severe comorbidities, listed for liver transplantation, treatment with beta-blockers or nitrates, TIPS or surgical shunt.	Mortality (TTE) Variceal bleeding (TTE) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D)	Medium or large varices stratum Survival and variceal bleeding as TTE data not reported in Cochrane so extracted separately Withdrawal in propranolol arm due to side effects n=17
Shah 2014 Full paper Pakistan Follow-up: mean 13.2 months	EVL (n=86)/carvedilol (n=82) EVL performed using Saeed Six Shooter Multiband ligator (Wilson-Cook). Repeated every 3 weeks until obliteration of varices achieved (no varices or only small varices which were flattened on air insuffations). Procedure repeated if varices	Cirrhosis without history of variceal bleed; medium or large sized oesophageal varices (grade II-IV). Diagnosis of cirrhosis made on the basis of clinical, radiological, biochemical features and liver histology where available Pregnant or lactating; allergy to carvedilol or reactive airway disease; already on beta-blocker	Mortality (TTE) Variceal bleeding (TTE) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D)	Medium or large varices stratum Not in Cochrane

Study	Intervention/comparison	Population	Outcomes	Comments
	recurred Carvedilol initial dose 6.25mg once a day increased to twice a day after a period of 1 week	treatment; presence of hepatic or other malignancy, which could impair longevity of life or presence of severe systemic illness which could impair the subject's ability to participate in the trial; psychiatric or mentally handicapped people; gastric varices alone.		
Singh 2012 Full paper India Follow-up: 12 months	EVL (n=18)/propranolol (n=20) EVL carried using PentaGun Multiband Ligator - as many bands as possible (3–6 bands). Performed weekly until varices obliterated or reduced to size grade 1. Procedure repeated if varices recurred or became grade 2 or larger. Propranolol started with 40 mg. Dose increased by increments of 20–40 mg/day until a 25% decrease in HR achieved	Patients with portal hypertension and oesophageal varices at high risk of bleeding, who had never had bleeding from varices. Large, grade 3 or 4 varices at high risk (Conn's criteria: grade 3, varices of 3 to 6 mm; grade 4, varices of > 6 mm). Eligibility criteria does not specify cirrhosis but results report all patients had cirrhosis and cirrhosis was diagnosed on the basis of clinical-biochemical, histologic, or ultrasonographic evidence Excluded: receiving antiviral therapy or concomitant hepatoma or tumour, severe cardio-pulmonary or renal disease, bradycardia, bronchial asthma, diabetes mellitus, heart failure, peripheral vascular disease, a psychiatric disorder, glaucoma, or prostatic hypertrophy	Mortality (D) gastrointestinal bleeding (D) Bleeding mortality (D)	Medium or large varices stratum Not in Cochrane
Song 2000 Abstract Korea follow-up: unclear.	EVL (n=31)/propranolol (n=30) No further intervention details given	Cirrhosis and high-risk oesophageal varices (blue coloured, more than enlarged tortuous varix with red colour signs) and no history of bleeding Excluded HCC and history of cardiopulmonary disease	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding (D)	Medium or large varices stratum
Thuluvath 2005 Full paper USA follow-up: mean 27	EVL (n=16)/propranolol (n=15) EVL using a multiband ligator every 2–3 weeks until variceal eradication Propranolol titrated to achieve a HR	Cirrhosis and large oesophageal varices (F2 or F3), no previous bleeding and HVPG ≥12 mmHg. Cirrhosis diagnosis made by clinical or histologic evidence.	Mortality (D) Variceal bleeding (D) gastrointestinal bleeding – variceal only (D)	Medium or large varices stratum

Study	Intervention/comparison	Population	Outcomes	Comments
months.	of <60 bpm or a 25% reduction, or until maximum dose reached	Excluded large gastric varices; previous EVL or sclerotherapy; HCC; contraindications to beta-blockers.	Bleeding mortality (D)	
Tripathi 2009 Full paper UK follow-up: mean 26 months (range not reported)	EVL (n=75)/carvedilol (n=77) EVL performed using a multiband ligator (Speedbander or Six shooter). Performed every 2 weeks until eradication or grade I in size. EVL repeated on recurrence of varices. Carvedilol starting dose of 6.25 mg per day, increased to a target dose of 12.5 mg/day if systolic BP did not fall below 90 mmHg.	Cirrhosis and oesophageal varices grade II or larger in size without previous bleeding. Cirrhosis diagnosis made on the basis of clinical, radiological or laboratory evidence and/or liver biopsy. Excluded <18 or >75 years; pregnant or lactating; childbearing age not on contraception; carvedilol allergy; malignancy affecting survival; systemic illness; psychiatric disease; obstructive airway disease; portal vein thrombosis; mean arterial pressure <55 mmHg or pulse <50 bpm.	Mortality (TTE) Variceal bleeding (TTE) gastrointestinal bleeding – variceal only (D) Bleeding mortality (D)	Medium or large varices stratum Survival and variceal bleeding as TTE data not reported in Cochrane so extracted separately 23 in each group discontinued therapy

D: outcome reported as a dichotomous outcome; TTE: outcome reported as a time-to-event outcome

Table 72: Clinical evidence summary: non-selective beta-blockers versus placebo or no intervention: medium or large varices

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with large varices: non-selective beta-blockers versus placebo (95% CI)
Quality of life	0 (0 studies)	–	–	–	
Survival	398 (2 studies)	LOW[c,e] due to inconsistency, imprecision	HR 1.2 (0.78 to 1.84)	Moderate[a]	
				338 per 1000	52 more per 1000 (from 63 fewer to 194 more)
Free from variceal bleeding	0 (0 studies)	–	–	–	
Variceal bleeding	268 (3 studies) 24 months[b]	VERY LOW[c,d,e] due to inconsistency, indirectness, imprecision	RR 0.28 (0.06 to 1.3)	Moderate	
				364 per 1000	262 fewer per 1000 (from 342 fewer to 109 more)
Upper gastrointestinal bleeding	448 (3 studies) 24 months[b]	MODERATE[e] due to imprecision	RR 0.55 (0.39 to 0.78)	Moderate	
				352 per 1000	158 fewer per 1000 (from 77 fewer to 215 fewer)
Bleeding-related mortality	398 (2 studies) 21 months[b]	MODERATE[e] due to imprecision	RR 0.67 (0.39 to 1.13)	Moderate	
				149 per 1000	49 fewer per 1000 (from 91 fewer to 19 more)

[a] Calculated from the median control group rate at the end of study
[b] Median of the mean follow-up times of the individual studies where reported
[c] Statistical heterogeneity. Cannot investigate predefined subgroups. Random effects model used
[d] Reported as a dichotomous outcome not time-to-event
[e] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 73: Clinical evidence summary: non-selective beta-blockers versus placebo or no intervention: small varices

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with small varices: non-selective beta-blockers versus placebo (95% CI)
Quality of life	0 (0 studies)	–	–	–	
Survival	0 (0)	–	–	–	
Mortality	150 (1 study) 25 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 1.42 (0.24 to 8.27)	Moderate	
				27 per 1000	11 more per 1000 (from 21 fewer to 196 more)
Free from variceal bleeding	0 (0 studies)	–	–	–	
Variceal bleeding	237 (3 studies) 24 months[d]	VERY LOW[a,b,c,e] due to risk of bias, inconsistency, indirectness, imprecision	RR 1.24 (0.31 to 5)	Moderate	
				69 per 1000	17 more per 1000 (from 48 fewer to 276 more)
Upper gastrointestinal bleeding	182 (2 studies) 24.5 months[d]	VERY LOW[a,c,f] due to risk of bias, inconsistency, imprecision	RR 0.9 (0.04 to 20.15)	Moderate	
				95 per 1000	10 fewer per 1000 (from 91 fewer to 1000 more)

[a] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[b] Reported as a dichotomous outcome not time-to-event
[c] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
[d] Median of the mean follow-up times of the individual studies where reported
[e] I-squared value 13%. Heterogeneity by visual inspection of the forest plots (different directions of effect). Cannot perform predefined subgroups. Random effects model used
[f] Statistical heterogeneity. Cannot investigate predefined subgroups. Random effects model used

Table 74: Clinical evidence summary: endoscopic band ligation versus no intervention: medium or large varices

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with large varices: banding ligation versus no intervention (95% CI)
Quality of life	0 (0 studies)	–	–	–	
Mortality (time-to-event)	253 (2 studies)	MODERATE[b] due to risk of bias	HR 0.5 (0.33 to 0.75)	Moderate[a]	
				472 per 1000	199 fewer per 1000 (from 91 fewer to 282 fewer)
Mortality	170 (2 studies) 14–25 months	VERY LOW[b,c,d] due to risk of bias, indirectness, imprecision	RR 0.57 (0.33 to 0.97)	Moderate	
				311 per 1000	134 fewer per 1000 (from 9 fewer to 208 fewer)
Variceal bleeding (time-to-event)	253 (2 studies)	MODERATE[b] due to risk of bias	HR 0.39 (0.25 to 0.63)	Moderate[a]	
				408 per 1000	223 fewer per 1000 (from 127 fewer to 285 fewer)
Variceal bleeding	170 (2 studies) 14–25 months	VERY LOW[b,c,d,e] due to risk of bias, inconsistency, indirectness, imprecision	RR 0.4 (0.17 to 0.93)	Moderate	
				467 per 1000	280 fewer per 1000 (from 33 fewer to 388 fewer)
Upper gastrointestinal bleeding	444 (5 studies) 20.6 months[f]	VERY LOW[b,d,e] due to risk of bias, inconsistency, imprecision	RR 0.49 (0.31 to 0.76)	Moderate	
				394 per 1000	201 fewer per 1000 (from 95 fewer to 272 fewer)
Bleeding-related mortality	297 (3 studies) 25 months[f]	LOW[b] due to risk of bias	RR 0.36 (0.18 to 0.71)	Moderate	
				152 per 1000	97 fewer per 1000 (from 44 fewer to 125 fewer)

[a] Calculated from the median control group rate at the end of study
[b] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[c] Reported as a dichotomous outcome not time-to-event
[d] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with large varices: banding ligation versus no intervention (95% CI)

[e] Statistical heterogeneity. Cannot investigate predefined subgroups. Random effects model used
[f] Median of the mean follow-up times of the individual studies where reported

Table 75: Clinical evidence summary: endoscopic band ligation versus no intervention: small varices

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with small varices: banding ligation versus no intervention (95% CI)
Quality of life	0 (0)	–	–	–	
Survival	0 (0)	–	–	–	
Free from variceal bleeding	0 (0)	–	–	–	
Upper gastrointestinal bleeding	31 (1 study) 20.6 months	VERY LOW[b,c] due to risk of bias, imprecision	See comment	Moderate	
				0 per 1000	70 more per 1000 (from 100 fewer to 240 more)[a]

[a] Manual calculation of absolute risk difference due to zero events in the control arm
[b] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

Table 76: Clinical evidence summary: endoscopic band ligation versus non-selective beta-blockers: medium or large varices

Outcomes	Number of participants (studies)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with large varices: banding ligation versus non-selective beta-blockers (95% CI)

Outcomes	Number of participants (studies)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with large varices: banding ligation versus non-selective beta-blockers (95% CI)
Quality of life	0 (0 studies)	–	–	–	
Survival	790 (7 studies)	MODERATE[e] due to imprecision	HR 1.03 (0.8 to 1.34)	Moderate1	
				333 per 1000	8 more per 1000 (from 56 fewer to 86 more)
Mortality	790 (12 studies) 14.5 months[b]	VERY LOW[c,d,e] due to risk of bias, indirectness, imprecision	RR 0.83 (0.61 to 1.13)	Moderate	
				140 per 1000	24 fewer per 1000 (from 55 fewer to 18 more)
Free from variceal bleeding	805 (7 studies)	VERY LOW[e,f] due to inconsistency, imprecision	HR 0.68 (0.35 to 1.31)	Moderate1	
				273 per 1000	78 fewer per 1000 (from 167 fewer to 68 more)
Variceal bleeding	554 (10 studies) 16.5 months[b]	MODERATE[d] due to indirectness	RR 0.44 (0.27 to 0.71)	Moderate	
				145 per 1000	81 fewer per 1000 (from 42 fewer to 106 fewer)
Upper gastrointestinal bleeding	1610 (20 studies) 19 months[b]	LOW[e,f] due to inconsistency, imprecision	RR 0.71 (0.54 to 0.92)	Moderate	
				159 per 1000	46 fewer per 1000 (from 13 fewer to 73 fewer)
Bleeding-related mortality	1258 (15 studies) 19 months[b]	MODERATE[e] due to imprecision	RR 0.67 (0.42 to 1.08)	Moderate	
				65 per 1000	21 fewer per 1000 (from 38 fewer to 5 more)
Hospitalisation	89 (1 study) 0.5-18 months	LOW[c,e] due to risk of bias, imprecision	RR 0.41 (0.16 to 1.06)	Moderate	
				273 per 1000	161 fewer per 1000 (from 229 fewer to 16 more)
Adverse events - Lethargy	163	MODERATE[c]	OR 0.09	Moderate	

Outcomes	Number of participants (studies)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with large varices: banding ligation versus non-selective beta-blockers (95% CI)
	(2 studies)	due to risk of bias	(0.04 to 0.22)	288 per1000	253 fewer per 1000 (from 206 fewer to 272 fewer)

[a] Calculated from the median control group rate at the end of study
[b] Median of the mean follow-up times of the individual studies where reported
[c] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[d] Reported as a dichotomous outcome not time-to-event
[e] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
[f] I-squared value 13%. Heterogeneity by visual inspection of the forest plots (CIs do not overlap). Predefined subgroup analyses performed but no statistical difference between subgroups. Random effects model used

Table 77: Clinical evidence summary: endoscopic band ligation versus non-selective beta-blockers: small varices

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with small varices: banding ligation versus non-selective beta-blockers (95% CI)
Quality of life	0 (0)	–	–	–	
Survival	0 (0)	–	–	–	
Free from variceal bleeding	0 (0)	–	–	–	

10.4 Economic evidence

10.4.1 Published literature

One economic evaluation was identified that compared non-selective beta-blockers with band ligation for primary prevention of bleeding in patients with varices.[101,106,149] This is summarised in the economic evidence profiles below (Table 78) and the economic evidence table in Appendix I.

No relevant economic evaluations were identified that compared non-selective beta-blockers with no prophylaxis.

No relevant economic evaluations were identified that compared band ligation with no prophylaxis.

See also the economic article selection flow chart in Appendix F.

Table 78: Economic evidence profile: band ligation versus non-selective beta-blockers

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost-effectiveness	Uncertainty
Norberto 2007[149] (Italy)	Partially applicable[(a)]	Potentially serious limitations[(b)]	• RCT (n=62) with candidates for liver transplantation • Medium to large varices • Primary end point was variceal bleeding, (mortality and other outcomes also recorded) • Beta-blocker: propranolol	£1,850	Variceal bleeding: −3% of patients	ICER: £57,812 per bleeding episode averted	No sensitivity analysis was conducted. None of the clinical differences between the 2 arms were statistically important.

Abbreviations: ICER: Incremental cost-effectiveness ratio; QALY: quality-adjusted life year; RCT: randomised controlled trial; TIPS: transjugular intrahepatic portosystemic shunt

(a) Study does not report QALYs; health outcomes and costs are not discounted

(b) Health-related quality of life was not measured; the study had a relatively short time horizon; no sensitivity analysis was performed

10.4.2 Unit costs

See Tables 92 and 93 in Appendix O.

10.5 Evidence statements

10.5.1 Clinical

People with cirrhosis and small oesophageal varices

- For the comparison of non-selective beta-blockers with placebo or no intervention, no evidence was identified for the critical outcomes of quality of life, mortality as a time-to-event outcome or freedom from variceal bleeding as a time-to-event outcome. Evidence of Very Low quality demonstrated no clinically important difference between beta-blockers and placebo or no intervention for the outcomes of mortality, variceal bleeding and upper gastrointestinal bleeding (1 study with 150 patients, 3 studies with 237 patients and 2 studies with 182 patients for each outcome, respectively).
- For the comparison of endoscopic variceal band ligation with no intervention, no evidence was identified for the critical outcomes of quality of life, survival and freedom from variceal bleeding. Evidence of Very Low quality suggested a clinical harm of band ligation on the outcome of upper gastrointestinal bleeding, but evidence was only available from 1 study in a subgroup analysis of 31 patients with small oesophageal varices.
- For the comparison of endoscopic variceal band ligation with non-selective beta-blockers, no evidence was identified for this population stratum.

People with cirrhosis and medium or large oesophageal varices

- For the comparison of non-selective beta-blockers with placebo or no intervention, no evidence was identified for the critical outcomes of quality of life and freedom from variceal bleeding as a time-to-event outcome. Evidence from 2 studies with 398 patients suggested a clinical harm of beta-blockers on survival, but there was some uncertainty and evidence was of Low quality. Evidence of Very Low quality suggested a clinical benefit of beta-blockers on variceal bleeding (3 studies with 268 patients). Evidence of Moderate quality demonstrated a clinically important benefit of beta-blockers on upper gastrointestinal bleeding and bleeding-related mortality (3 studies with 448 patients and 2 studies with 398 patients, respectively).
- For the comparison of endoscopic variceal band ligation with no intervention, no evidence was identified for the critical outcome of quality of life. Two studies with 253 patients reported survival and freedom from variceal bleeding as time-to-event outcomes. These studies provided Moderate quality evidence demonstrating a clinically important benefit of band ligation on survival and variceal bleeding. Evidence of Very low and Low quality demonstrated a clinically important benefit of band ligation on upper gastrointestinal bleeding and bleeding-related mortality (5 studies with 444 patients and 3 studies with 297 patients, respectively).
- For the comparison of endoscopic variceal band ligation with non-selective beta-blockers, no evidence was identified for the critical outcome of quality of life. Evidence of Moderate quality suggested no clinical difference between band ligation and beta-blockers on survival (7 studies with 790 patients). However, a clinically important benefit of band ligation was observed from Very Low quality evidence reporting mortality as a dichotomous outcome (12 studies with 790 patients). Very Low quality evidence demonstrated a clinically important benefit of band ligation on freedom from variceal bleeding (7 studies with 805 patients reported time-to-event data). A similar clinically important benefit of band ligation on variceal bleeding reported as a dichotomous outcome was demonstrated from Moderate quality evidence (10 studies with 554

patients reported as a dichotomous outcome). There was a clinically important benefit of band ligation on upper gastrointestinal bleeding (Low quality, 20 studies with 1610 patients), bleeding-related mortality (Moderate quality, 15 studies with 1258 patients), hospitalisation (Low quality, 1 study with 89 patients) and lethargy due to beta-blockers (Moderate quality, 2 studies with 163 patients).

10.5.2 Economic

People with cirrhosis and small oesophageal varices

- No relevant economic evaluations were identified.

People with cirrhosis and medium or large oesophageal varices

- For the comparison of non-selective beta-blockers with placebo or no intervention, no relevant economic evaluations were identified.
- For the comparison of endoscopic variceal band ligation with placebo or no intervention, no relevant economic evaluations were identified.
- For the comparison of endoscopic variceal band ligation with non-selective beta-blockers, 1 cost-consequences analysis found that band ligation was more costly and more effective compared to beta-blockers for primary prevention of bleeding in patients with varices (£1,850 more per patient, 0.03 fewer deaths per patient, and 0.03 fewer patients with bleeding episodes). This analysis was assessed as partially applicable with potentially serious limitations.

10.6 Recommendations and link to evidence

Recommendation	**21. Offer endoscopic variceal band ligation for the primary prevention of bleeding for people with cirrhosis who have medium to large oesophageal varices.**
Research recommendation	**2. Do non-selective beta-blockers improve survival and prevent first variceal bleeds in people with cirrhosis that is associated with small oesophageal varices?**
Relative values of different outcomes	The GDG discussed the merits of various outcome measures that compared band ligation to non-selective beta-blockers in the prevention of primary bleeding in people with cirrhosis and oesophageal varices. The GDG agreed that health-related quality of life, survival, and freedom from variceal haemorrhage were the critical outcomes for decision-making. Whilst the GDG agreed that survival and freedom from variceal bleeding reported as time-to-event data (hazard ratio) were the preferred measures for decision-making (as the effect of the interventions on the time elapsed before the event occurred was considered an important factor), they wanted to retain as much evidence as possible for these particular outcomes and also considered dichotomous data even though this was quality downgraded and the GDG placed less weighting on dichotomous outcomes for decision-making. The GDG agreed that important outcomes were hospital admission, hospital length of stay, episodes of primary upper gastrointestinal bleeding (regardless of source), bleeding-related mortality and the occurrence of the adverse event of fatigue with beta-blockers. The GDG consensus of opinion was that band ligation, unlike other interventions such as sclerotherapy, is associated with few long-term complications that should be considered as adverse events. The GDG noted that the outcome of upper gastrointestinal bleeding was heterogeneous in the way it was reported by

	individual studies. Some studies only reported variceal bleeding, others reported upper gastrointestinal bleeding (excluding variceal bleeding) and others reported upper gastrointestinal bleeding from varices and other sources.
Trade-off between clinical benefits and harms	The GDG considered 2 pre-specified population strata separately (people with small oesophageal varices and people with medium or large oesophageal varices). It is widely accepted that the intervention would act differently in people with small varices and in people with medium or large varices. For example, band ligation is an invasive procedure and it is unlikely that it would be considered in people with small varices because of the technical difficulties involved. This was reflected in the evidence, with only 1 study reporting band ligation in people with small varices (subgroup data from a study that included people with both small and large varices). This clinical difference between the two population strata is also reflected in the evidence by the rarity of studies including both people with small varices and people with larger varices. Three studies were identified that recruited both populations overall.[8,42,232] Outcomes were only extracted and analysed where the data were provided separately for the 2 strata in a subgroup analysis. Studies that recruited both populations overall and reported outcomes generalised for the whole study population without a subgroup analysis were excluded and listed in the excluded studies list. The GDG did not consider this evidence, as it wished to make separate recommendations for the 2 population strata. **People with small oesophageal varices** *Non-selective beta-blockers versus placebo or no intervention* Only 3 studies[8,42,193] reported on this comparison. The studies did not report outcomes considered as critical by the GDG (quality of life, survival as time-to-event data, or freedom from variceal bleeding as time-to-event data). Dichotomous outcomes were available for mortality, variceal bleeding and upper gastrointestinal bleeding. The GDG considered that there was no clinically important difference between beta-blockers and placebo or no intervention for any of the outcomes reported. *Band ligation versus placebo or no intervention* Only 1 study[232] compared band ligation with placebo in people with small varices as a subgroup analysis (the study recruited people with all sizes of varices). As with the above comparison, this study did not report on the critical outcomes. Only upper gastrointestinal bleeding was reported and this showed clinical harm for band ligation. However, the GDG noted that this was a very small study with only 1 person in the band ligation group having this outcome. *Band ligation versus non-selective beta-blockers* There was no evidence for this comparison in people with small varices. **People with medium or large oesophageal varices** *Non-selective beta-blockers versus placebo or no intervention* Only 4 studies[8,42,153,159] reported this comparison in people with medium or large varices and only 1 critical outcome (survival) was reported which showed a clinical harm for beta-blockers. The GDG noted imprecision in this result and the evidence was of Low quality. Beta-blockers were of clinical benefit over placebo for the outcomes of variceal bleeding, when reported as a dichotomous outcome, and for upper gastrointestinal bleeding and bleeding-related mortality. *Band ligation versus placebo or no intervention*

	Five studies[118,125,191,221,232] reported this comparison in people with medium or large varices. Band ligation was of clinical benefit over placebo or no intervention for the critical time-to-event outcomes of survival and freedom from variceal bleeding, and for the important outcomes of upper gastrointestinal bleeding and bleeding-related mortality. *Band ligation versus non-selective beta-blockers* Twenty studies (25 papers)[1,4,6,38,51,53,54,82,88,106,107,117,127,149,161,166,192,194-196,198,203,206,227,234] reported this comparison for people with medium or large varices. No clinically important difference between beta-blockers and band ligation was observed for survival. However, a clinically important benefit of band ligation was reported for mortality (dichotomous outcome), variceal bleeding (when reported as time-to-event or dichotomous data), upper gastrointestinal bleeding, bleeding-related mortality, hospitalisation and lethargy due to beta-blockers. No evidence was available for any of the comparisons for the quality of life outcome for people with medium or large varices.
Trade-off between net clinical effects and costs	One cost-consequence analysis (Norberto 2007)[149] was identified that directly compared band ligation with beta-blockers in people with medium or large varices. No relevant studies were identified comparing either band ligation or beta-blockers with no intervention, or in people with small varices. Norberto 2007 reported that overall costs were £1,850 greater per person for band ligation compared to beta-blockers with a small (not statistically significant) clinical benefit of 0.03 fewer patients with variceal bleeding and 0.03 fewer deaths per patient. The difference in total costs was mainly due to the higher intervention cost of band ligation, as the follow-up and hospital costs were similar in both arms. However, the GDG noted that this study (also included in the clinical review) had clinical results less favourable for band ligation than the meta-analysed results of the clinical review as a whole. The additional cost of band ligation should therefore be compared against the increased effectiveness shown in the pooled clinical effectiveness figures from the meta-analyses in this chapter, rather than the clinical effectiveness demonstrated in the Norberto 2007 study alone. The GDG also noted that the excess cost of band ligation might be expected to be lower in a study representative of the clinical review. Whilst the higher cost of the initial band ligation procedure would remain, follow-up and hospital costs would be lower for band ligation than for beta-blockers due to the lower rates of variceal bleeding and rehospitalisation in people treated with band ligation. Using the decreased rates of all-cause mortality and variceal bleeding found in the clinical review, with costs of £1,326 for band ligation and £2,653 for treating variceal bleeding (as used in the original economic model for this guideline, see Appendix N) and £56.71 for 1 year of propranolol at 40 mg 3 times per day (NHS Drug Tariff), shows that band ligation would be expected to be cost-effective at a threshold of £20,000 per QALY gained. For band ligation versus no treatment, band ligation would have an incremental cost of £710 (£1,793−£1,083) and incremental effectiveness of 0.597 QALYs, giving an ICER of £1,190 per QALY gained. For band ligation versus beta-blockers, band ligation would have an incremental cost of £1,054 (£1,496−£441) and an incremental effectiveness of 0.072 QALYs, giving an ICER of £14,641 per QALY gained. This assumes that death leads to a loss of 3 QALYs and bleeding leads to a loss of 0.03 QALYs, but does not include the additional financial benefit of decreased subsequent hospitalisations (161 fewer per 1,000 people with band ligation compared to beta-blockers) as the length of these hospitalisations is not known. The GDG noted that these ICERs are estimates, but are consistent with what would be expected if Noberto 2007 was updated with clinical data representative of the clinical review. The GDG concluded that band ligation was likely to be cost-effective

	compared to beta-blockers for people with medium or large varices at a cost-effectiveness threshold of £20,000 per QALY. As the GDG found insufficient clinical evidence to make a recommendation regarding people with small varices, the economics of treating this patient group were not considered.
Quality of evidence	The GDG discussed the included studies and noted the following: • In the Sarin et al. 1996[191] study, 6/68 (9%) of people had a cause for portal hypertension other than cirrhosis and in the Sarin et al. 1999[192] study this proportion was 6%. Overall, the GDG agreed that the small proportion of people with a portal hypertension unrelated to cirrhosis would be unlikely to affect the outcomes significantly, but the evidence quality was downgraded for population indirectness where this study contributed to the majority of the evidence. • Papers by Abdelfattah et al. 2006[4] and Chen et al. 1998[38] did not specify the size of the varices but, as they were included in the Cochrane review (people with high risk varices), they were likely to be people with medium or large varices (and were included within this stratum). • Two studies[82,149] included people with cirrhosis on the transplant waiting list and the GDG agreed that these studies should be included. • For the comparison of beta-blockers versus no intervention, the GDG included studies which had 'no intervention' in addition to placebo controlled study groups. Trial group 'blinding' is difficult in beta-blocker trials as the clinical effects (such as a reduction in pulse rate and lethargy) would alert both the participant and the investigators to the treatment. • Some studies included either proton pump inhibitors or sucrulfate following variceal band ligation and given that this was common following a band ligation procedure, these studies were also included. • There was limited evidence available for hospitalisation rates but it is likely that the incidence of variceal bleeding would reflect hospitalisation as each event would require an inpatient hospital stay. • Seven conference abstracts were included (previously included in the Cochrane review of beta-blockers versus variceal band ligation). Whilst abstracts have not been routinely used as evidence for other review questions it was agreed that, given the Cochrane group had extracted the data and contacted authors for additional information when outcomes or trial methods were not described in the published trial reports, these were included. • There was very limited evidence available for both the comparison of variceal band ligation versus placebo and band ligation versus beta-blockers for people with small varices. This was expected by the GDG, as band ligation is an invasive procedure that would not often be considered in people with small varices. **People with small oesophageal varices** *Non-selective beta-blockers versus placebo or no intervention* There were only 3 studies of small sample size and the reported outcomes were of Very Low quality. *Band ligation versus placebo or no intervention* Subgroup evidence from only 1 study of Very Low quality was available. **People with medium or large oesophageal varices** *Non-selective beta-blockers versus placebo or no intervention*

	Evidence for the critical outcomes was of Low and Very Low quality. Evidence for the important outcomes of upper gastrointestinal bleeding and bleeding-related mortality was of Moderate quality. *Band ligation versus placebo or no intervention* Evidence for the critical outcomes was of Moderate quality and for the important outcomes of upper gastrointestinal bleeding and bleeding-related mortality was of Very low and Low quality. *Band ligation versus non-selective beta-blockers* Evidence was of Moderate quality for the outcomes of survival, variceal bleeding (dichotomous), bleeding-related mortality and adverse events. For the outcomes of variceal bleeding (time-to-event), upper gastrointestinal bleeding and hospitalisation, evidence was of Low or Very Low quality.
Other considerations	**Small varices** Whilst the GDG noted no clinical benefit of beta-blockers in people with small varices there was a paucity of evidence which was of Very Low quality leading to uncertainty over the true effect of beta-blockers in the stratum. Overall the GDG agreed there was insufficient evidence to make a recommendation in people with small varices and instead chose to develop a research recommendation in this area. **Medium or large varices** The GDG: • Chose to recommend band ligation for primary prophylaxis of variceal bleeding in people with cirrhosis who have medium or large varices. While the GDG did not recommend the use of beta-blockers, they acknowledged that beta-blockers may have a role where band ligation is unavailable or contraindicated. • Highlighted that although there was no clinical benefit of band ligation over beta-blockers for the outcome survival, there was a clinical benefit given the reduced occurrence of variceal bleeding and bleeding-related mortality. The GDG agreed that variceal haemorrhage is a severe complication of cirrhosis. It favoured the use of variceal band ligation as this reduced the incidence of this particular outcome compared with beta-blockers. The GDG patient representatives stressed the importance of the implications relating to the psychological aspects of variceal haemorrhage in supporting this recommendation. The GDG also noted the significant survival advantage of band ligation over no intervention, and a reduced number of adverse events in the variceal band ligation group compared to beta-blockers. • Accepted that there are theoretical benefits to the use of beta-blockers other than survival (such as a reduction in bacterial translocation and other complications of portal hypertension). However, these benefits were not seen in the overall survival analysis and the GDG based this recommendation on the available evidence. • Highlighted that the review did not investigate the benefit or harm of band ligation in combination with beta-blockers, and so are unable to make any recommendations relating to combination therapy. **Research recommendation** Bleeding from oesophageal varices is a major complication of cirrhosis. Approximately half of patients with cirrhosis have oesophageal varices and one-third of all patients with varices will experience bleeding at some point. Despite

	improvements in the management of acute haemorrhage in recent decades, the 6-week mortality associated with variceal bleeding remains of the order of 10–20%. Risk of variceal bleeding increases with variceal size. Whether non-selective beta-blockers are of benefit as primary prophylaxis in people with cirrhosis and small oesophageal varices has not been adequately studied.

11 Primary prevention of bacterial infections in cirrhosis and upper gastrointestinal bleeding

11.1 Introduction

People with cirrhosis and upper gastrointestinal bleeding are frequently found to have bacterial infections during or soon after the bleeding episode, although at present, it is uncertain whether infection or bleeding is the primary event. Approximately 20% of patients have an infection at admission and 50% develop infections during the first days of hospitalisation in the absence of antibiotic prophylaxis.[222] The most common infections are spontaneous bacterial peritonitis (SBP) (25%), followed by urinary tract infection (20%), pneumonia (15%), bacteraemia and cellulitis. Infections are culture positive in 50–70% of cases and gram-positive cocci are implicated in 50% of bacterial infections.[65] Those diagnosed with bacterial infection within 48 hours of admission have a higher risk of death and a higher risk of early rebleeding, defined as recurrence of bleeding within 7 days after admission.[19]

It is essential that the prophylactic antibiotic is active against both Enterobacteriaceae and non-enteral streptococci. However, recent studies show an increasing prevalence of infections caused by antibiotic-resistant bacteria, especially in nosocomial episodes.[64] The NICE guideline for the management of acute upper gastrointestinal bleeding recommended that prophylactic antibiotic therapy should be offered to patients with suspected or confirmed variceal bleeding.[140] The purpose of this review was to find the most clinically and cost-effective route of administration of prophylactic antibiotics, therefore placebo controlled trials were excluded from the review.

11.2 Review question: What is the most clinical and cost-effective prophylactic antibiotic for the primary prevention of bacterial infections in people with cirrhosis and upper gastrointestinal bleeding?

For full details see review protocol in Appendix C.

Table 79: PICO characteristics of review question

Population	Adults and young people (16 and over) with confirmed cirrhosis and upper gastrointestinal bleeding (as determined by endoscopy)
Intervention(s)	**Prophylactic antibiotics including:** <u>Intravenous:</u> • beta-lactams: ○ penicillins (amoxicillin, ampicillin, co-amoxiclav [amoxicillin + clavulanic acid; Augmentin] and tazocin [piperacillin and tazobactam]) ○ 3rd generation cephalosporins (including cefotaxime, ceftazidime, ceftriaxone) • aminoglycoside (gentamicin, tobramycin, amikacin) • quinolones (ciprofloxacin, pefloxacin, ofloxacin, floxacin) • carbopenums (meropenum, ertapenem, impenem) • glycopeptide antibiotic vancomycin • glycylcycline antibiotic tigecycline <u>Oral antibiotic therapy (administered with a nasogastric tube or be preceded by the use of IV) including:</u> • quinolones (ciprofloxacin, norfloxacin, pefloxacin, ofloxacin, floxacin, levofloxacin,

	moxifloxacin) • penicillins (amoxicillin, co-amoxiclav [Augmentin], phenoxymethylpenicillin [also considered penicillin V]) • sulfonamides (trimethoprim, trimethoprim/sulphamethoxazole [Septrin] or co-trimoxazole) • third-generation cephalosporin (cefalexin) • other oral antibiotics: clarythromycin, erythromycin, colistin, clindamycin, doxycycline, azithromycin, metronidazole Exclusions: Other routes or modes of administration and dose comparisons (dose 1 versus dose 2)
Comparison(s)	IV versus oral IV versus IV Oral versus oral Any combinations of drugs above (that is IV + oral combination versus monotherapy) Exclusions: Placebo/no treatment
Outcomes	**Critical outcomes:** Occurrence of bacterial infections Health-related quality of life All-cause mortality (time-to-event data) **Important outcomes:** Adverse effect: renal failure Length of hospital stay Readmission rate Antibiotic complications (for example Clostridium difficile, diarrhoea)
Study design	RCTs Systematic reviews of RCTs

11.3 Clinical evidence

One Cochrane review[36,37] including 3 studies[67,182,209] and 1 additional study[110] was included in the review; these are summarised in Table 80 below. Evidence from these studies is summarised in the clinical evidence summary tables below (Table 81, Table 82, Table 83 and Table 84). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

The Cochrane review had a wider remit review considering both placebo controlled and head-to-head trials. Only the results of head-to-head trials of antibiotics were extracted for this report. Additionally, while 'number of days of hospitalisation' was one of the secondary outcomes in the Cochrane review, the review only reported this for studies comparing antibiotics with placebo. From examining the primary papers, 1 paper[182] was found to report this outcome and these data were extracted from that paper. Most of the patients in the included studies had more severely decompensated cirrhosis (Child-Pugh B or C) but 1 paper (which was published as an abstract) did not report on disease severity[209].

Of the outcomes in the review protocol (Table 79 and Appendix C), only occurrence of bacterial infection and length of hospital stay were reported in the literature. Since the preferred time-to-event data on the outcome all-cause mortality were not reported in the studies, dichotomous data were extracted and presented. However, this outcome was considered indirect. No evidence was found on any of the other outcomes. As there was only 1 study identified comparing one pair of interventions, no meta-analyses were conducted.

Table 80: Summary of studies included in the review

Study, country	Population	Intervention and comparison (treatment period)	Outcomes	Comments
Fernandez 2006[67]: (in Cochrane review[37]) Spain	124 adults with advanced cirrhosis and upper gastrointestinal bleeding. Child-Pugh score: 52B/59C	n=61 IV ceftriaxone (1 g) versus n=63 oral norfloxacin (400 mg twice daily) (both for 7 days)	All bacterial infections All-cause mortality	Nasogastric tube was used before treating the haemorrhage (some had balloon tamponade); oral treatment was administered within 12 hours of admission so it is assumed that it was administered through this nasogastric tube.
Kim 2011 [110] South Korea	113 adults with active gastrointestinal haemorrhage in the prior 24 hours, decompensated liver cirrhosis[(a)] Mean age (SD): 53.9 (9.7) years Mean Child-Pugh score: 8.5 (SD 1.7) ciprofloxacin versus 8.7 (SD 1.7) ceftriaxone	n=63 IV ceftriaxone (2 g) versus n=50 oral ciprofloxacin (500 mg twice daily) (both for 7 days)	All bacterial infections	Nasogastric tube was placed before treating the haemorrhage; oral treatment was administered within 12 hours of admission so it is assumed that it was administered through this nasogastric tube.
Sabat 1998[182] (in Cochrane review[37]) Spain	56 adults with cirrhosis and gastrointestinal bleeding Child-Pugh score: 4A/31B/11C	n=28 oral norfloxacin (400 mg twice daily) versus n=28 oral norfloxacin (400 mg twice daily) + IV ceftriaxone (2 g) (both for 7 days)	All bacterial infections All-cause mortality	Nasogastric tube was placed before treating the haemorrhage; study then reports that norfloxacin was administered orally or through nasogastric tube.
Spanish Group 1998[209] (in Cochrane review[37]) Spain	365 adults with cirrhosis and gastrointestinal haemorrhage. Child-Pugh score: unclear	n=183 oral norfloxacin (800 mg) versus n=182 oral ofloxacin (400 mg) (both for 5 days)	All bacterial infections	No details from conference abstract if nasogastric tube was used to administer oral drugs.

(a) Defined as Child-Turcotte-Pugh score of 7 or greater. According to the protocol definitions, this would be considered moderate decompensation, rather than decompensated liver disease (protocol considered 10 or more as decompensated liver disease)

Table 81: Clinical evidence summary: IV ceftriaxone 2 g versus oral ciprofloxacin 500 mg twice daily (both for 7 days)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with oral ciprofloxacin 500 mg twice daily	Risk difference with IV ceftriaxone 2 g (95% CI)
Bacterial infections	129 (1 study) 7 days	Moderate[a] due to risk of bias	RR 0.13 (0.03 to 0.56)	206 per 1000	179 fewer per 1000 (from 91 fewer to 200 fewer)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

Table 82: Clinical evidence summary: IV ceftriaxone 1 g versus oral norfloxacin 400 mg twice daily (both for 7 days)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with oral norfloxacin 400 mg twice daily	Risk difference with IV ceftriaxone 1 g (95% CI)
Bacterial infections	111 (1 study) 10 days	Very low[a,b] due to risk of bias, imprecision	RR 0.42 (0.18 to 1.01)	263 per 1000	153 fewer per 1000 (from 216 fewer to 3 more)
All-cause mortality (dichotomous)	111 (1 study) 10 days	Very low[a,b,c] due to risk of bias, indirectness, imprecision	RR 1.41 (0.52 to 3.79)	105 per 1000	43 more per 1000 (from 50 fewer to 293 more)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(c) Downgraded by 1/2 increments because the majority of the evidence had indirect outcomes

Table 83: Clinical evidence summary: oral norfloxacin 800 mg versus oral ofloxacin 400 mg (both for 5 days)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with oral ofloxacin 400 mg	Risk difference with oral norfloxacin 800 mg (95% CI)
Bacterial infections	365 (1 study) 10 days	Very low[a,b] due to risk of bias, imprecision	RR 0.96 (0.58 to 1.58)	148 per 1000	6 fewer per 1000 (from 62 fewer to 86 more)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 84: Clinical evidence summary: oral norfloxacin 800 mg + IV ceftriaxone (combination) versus oral norfloxacin 800 mg (monotherapy) (both for 7 days)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with oral norfloxacin 800 mg	Risk difference with oral norfloxacin 800 mg + IV ceftriaxone (95% CI)
Bacterial infections	46 (1 study) 3 weeks	Very low[a,b] due to risk of bias, imprecision	RR 0.69 (0.17 to 2.73)	182 per 1000	56 fewer per 1000 (from 151 fewer to 315 more)
All-cause mortality (dichotomous)	46 (1 study) 3 weeks	Very low[a,b,c] due to risk of bias, indirectness, imprecision	RR 0.46 (0.04 to 4.71)	91 per 1000	49 fewer per 1000 (from 87 fewer to 338 more)
Length of hospital stay (days)	46 (1 study) 3 weeks	Very low[a,b] due to risk of bias, imprecision		The mean length of hospital stay in the control groups was 12 days	The mean length of hospital stay in the intervention groups was 0 higher (4.07 lower to 4.07 higher)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(c) Downgraded by 1/2 increments because the majority of the evidence had indirect outcomes

11.4 Economic evidence

11.4.1 Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

11.4.2 Unit costs

See Table 94 in Appendix O.

11.5 Evidence statements

11.5.1 Clinical

- IV ceftriaxone (2 g) showed a clinical benefit over oral ciprofloxacin (500 mg twice daily) for the outcome of bacterial infections (1 study, n=129, Moderate quality).
- IV ceftriaxone (1 g) showed a clinical benefit over oral norfloxacin (400 mg twice daily) for the outcome of bacterial infections, but was considered to cause clinical harm for the outcome of all-cause mortality (1 study, n=111, Very Low quality for both outcomes).
- There was no clinically important difference between oral norfloxacin (800 mg) and oral ofloxacin (400 mg) for the outcome of bacterial infections (1 study, n=365, Very Low quality).
- Combination therapy with oral norfloxacin (800 mg) and IV ceftriaxone (2 g) showed a clinical benefit over monotherapy with norfloxacin only (800 mg) for the outcomes of bacterial infections and all-cause mortality; however, there was no clinically important difference between combination therapy and monotherapy for the outcome of length of hospital stay (1 study, n=46, Very Low quality for all outcomes).

11.5.2 Economic

- No relevant economic evaluations were identified.

11.6 Recommendations and link to evidence

Recommendation	**22. Offer prophylactic intravenous antibiotics for people with cirrhosis who have upper gastrointestinal bleeding.** **23. Review intravenous antibiotics prescriptions in line with the prescribing intravenous antimicrobials section in NICE's antimicrobial stewardship guideline.**
Relative values of different outcomes	The NICE clinical guideline 141 'Acute upper gastrointestinal bleeding: Management'[140] recommends that all patients with suspected or confirmed variceal upper gastrointestinal bleeding should be offered antibiotic prophylaxis. However, the GDG noted there is considerable variation in practice regarding the type of antibiotic and the route of delivery and hence they were particularly interested in the clinical and cost-effectiveness of different prophylactic antibiotics for the prevention of bacterial infections in people with cirrhosis and upper gastrointestinal bleeding. The occurrence of bacterial infections, health-related quality of life and all-cause

	mortality were identified by the GDG as the critical outcomes. Renal failure, length of hospital stay, hospital readmission rate and antibiotic-related complications (for example, antibiotic resistance or *Clostridium difficile* diarrhoea) were agreed to be important outcomes.
Trade-off between clinical benefits and harms	The GDG noted that for the critical outcome of occurrence of bacterial infection there was a clinical benefit favouring the use of intravenous antibiotics over oral preparations. There were insufficient data to identify a particular antibiotic class as being more beneficial. No data were available for any of the comparisons for the critical outcome of quality of life and evidence for the outcome of all-cause mortality was only available for 2 of the 4 comparisons. The GDG agreed that IV administration was more appropriate predominantly because of a reduction in bacterial infections, but also due to the difficulties of oral administration in people with haematemesis and critical illness, many of whom require intubation and ventilation in an intensive care environment. In such cases, placement of a nasogastric tube is relatively contra-indicated in the 48-hour period following treatment of bleeding varices necessitating IV antibiotic administration. The GDG considered the duration of antibiotic treatment but were unable to make a recommendation based on the evidence available. The included studies assess IV treatment for 5 days or 7 days. However, current clinical practice for prophylaxis of bacterial infections is to use IV antibiotics for around 48 hours before stopping or switching to oral administration if there is evidence of bacterial infection. The GDG agreed that a short course of intravenous antibiotics should be considered, with the duration being guided by the clinical status of the patient. The GDG noted the existing NICE guidance on reviewing intravenous antibiotics within a 48–72 hour period within NG15 Antimicrobial stewardship and agreed it would be helpful to make a cross reference to this.
Trade-off between net clinical effects and costs	No relevant economic evaluations were identified. The GDG examined the cost differences between intravenous ceftriaxone (£47.90 and 95.90 for 1 g and 2 g respectively) and oral antibiotics (between £1.05 and £8.57 depending on the antibiotic used) for a 5-day course. Considering these costs compared with the clinical benefits of treatment, the GDG felt that the costs for the intravenous delivery of antibiotics for a 48-hour period (current practice) would be outweighed by their clinical benefits; making intravenous antibiotics a cost-effective option compared to oral antibiotics at a cost-effectiveness threshold of £20,000 per QALY gained.
Quality of evidence	There was significant heterogeneity in the design of the 4 included studies as they all used different comparators. Given this, the GDG considered each study individually. The GDG noted that most studies included patients with cirrhosis and active upper gastrointestinal haemorrhage. All of the evidence examined was of Very Low quality with the exception of the outcome of bacterial infections for the comparison IV ceftriaxone versus oral ciprofloxacin, which showed a clinical benefit of IV ceftriaxone (Moderate quality evidence).
Other considerations	The GDG discussed the clinical issues related to this question. In particular they discussed concerns related to inappropriate antibiotic use in the development of resistance and of *Clostridium difficile* infection. It was felt that the choice of antibiotics should be determined by local microbiological practices. Clinicians should consider the prevalence of organisms in the local population, resistance profiles and the fact that the 'at risk' population consists of immune compromised people with decompensated cirrhosis with likely bacterial translocation of predominantly gram-negative organisms. Whilst the GDG noted that it would be possible for patients to receive both oral and IV antibiotics at home, patients with an upper gastrointestinal bleed are likely to be hospitalised for a minimum of 5 days. The GDG considered making a research recommendation with regards to the class of

	antibiotic and the duration of therapy, however ethical considerations would most likely preclude research in this area.

12 Transjugular intrahepatic portosystemic shunt (TIPS) versus large-volume paracentesis (LVP) for ascites

12.1 Introduction

Sodium and fluid retention are frequent complications of end-stage liver disease and, as a consequence, around 50% of patients with cirrhosis will develop ascites over a 10-year period of follow-up.[85] Ascites significantly impairs the quality of life of patients with cirrhosis and is associated with poor prognosis: 1-year and 5-year survival rates of 85% and 56%, respectively are reported.[165]

The first-line of management of ascites is restriction of dietary sodium by use of a 'no added salt' diet which contains <90 mmol/day (5.2 g of salt/day) and diuretic therapy, using spironolactone with or without added furosemide. These measures are effective in the majority of people with ascites secondary to cirrhosis. Refractory ascites develops in <10% of cases and is due to either inadequate urinary sodium excretion despite diuretics, or development of diuretic-induced complications that preclude the use of an effective dosage.[189] Patients with large volume or refractory ascites may be treated by therapeutic paracentesis[84] accompanied by volume expansion using 20% human albumin solution, providing 6-8 g of albumin per litre of ascites drained,[180] but this procedure generally needs to be repeated at regular intervals to maintain control of the fluid retention. Alternatively, transjugular intrahepatic portosystemic shunt (TIPS) has been advocated as a treatment in patients with refractory ascites needing frequent paracentesis (>3/month) who are not candidates for liver transplantation.[47] However this procedure is complicated by the development of hepatic encephalopathy. TIPS decrease the effective vascular resistance of the liver by the creation of a tract between the higher-pressure portal vein and the lower-pressure hepatic vein, decreasing the portal venous pressure. This in turn lessens the congestive pressure in veins in the intestine reducing production of ascites. The tract is maintained by placement of a special mesh tube known as a stent. Early stents were bare metal but more recently polytetrafluoroethylene (PTFE)-coated grafts have been employed to improve patency.[246]

It is generally perceived that both TIPS and LVP improve quality of life in patients with refractory ascites, but there is disagreement concerning the impact of TIPS on long-term survival. Therefore, the GDG decided to compare the clinical and cost-effectiveness of TIPS compared with large-volume paracentesis (LVP) with albumin in the management of patients with refractory ascites due to cirrhosis.

12.2 Review question: What is the clinical and cost-effectiveness of transjugular intrahepatic portosystemic shunt (TIPS) compared with large-volume paracentesis (LVP) with albumin in the management of diuretic-resistant ascites due to cirrhosis?

For full details see review protocol in Appendix C.

Table 85: PICO characteristics of review question

Population	Adults and young people (16 and over) with confirmed cirrhosis and diuretic-resistant (refractory) ascites
Intervention	Transjugular intrahepatic portosystemic shunt (TIPS)
Comparison	Large-volume paracentesis (LVP) with albumin infusion

Outcomes	**Critical outcomes:** • Re-accumulation of ascites • Health-related quality of life • Transplant-free survival **Important outcomes:** • Spontaneous bacterial peritonitis • Renal failure • Hepatic encephalopathy • Length of hospital stay • Readmission rate
Study design	RCTs Systematic reviews of RCTs

12.3 Clinical evidence

We searched for randomised trials comparing the effectiveness of TIPS with LVP for the management of diuretic-resistant ascites due to cirrhosis.

One Cochrane systematic review[181] and one RCT[139] were included in the review; the study characteristics are summarised below (Table 86). Evidence from these studies is summarised in the clinical evidence summary below Table 87. See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

The Cochrane review was partially included because the review did not include all the outcomes specified in our protocol. Consequently, the papers included in the Cochrane were examined individually to extract the additional outcomes: transplant-free survival, SBP, quality of life and hospital readmissions if available. Additional information on the population, intervention and comparison was also extracted from the primary research papers and summarised in Table 86. One RCT included in the Cochrane review (Lebrec 1996 study[119]) was excluded from the current review. After examination of the original paper, the TIPS intervention was deemed to be different to current UK practice for the treatment of refractory ascites. This method also differed from the method used by the other studies in the review. This method involved the insertion of multiple stents (eight people had 2 stents, two people had 3 stents) irrespective of portal pressure reduction. Current practice is to use a single stent dilated to induce a fall in portal pressure below 12 mmHg, but no further. A further stent is only deployed if the portal pressure does not fall below 12 mmHg following the first stent insertion. This portal pressure reduction, goal approach was used in the other papers in the review. The method used in the Lebrec 1996 study may result in differences in the effectiveness of the intervention and a higher hepatic encephalopathy complication rate.

Two of the included studies specifically reported that all patients in the TIPS group received LVP prior to TIPS.[87,190] The other 3 studies did not report whether patients received LVP prior to TIPS.[139,179,185] However, it is unlikely that the TIPS procedure would have been performed without a prior LVP and it was assumed that the interventions were sufficiently similar in all the studies to pool together. Some studies reported total paracentesis and others reported large-volume paracentesis. Again, these are likely to be sufficiently similar interventions and evidence was pooled together from all studies.

Table 86: Summary of studies included in the review

Study	Population	Intervention and comparison	Outcomes	Comments
Gines 2002[87] RCT Included in Cochrane Saab 2006[181] Spain and USA Follow-up greater than 26 months after inclusion	Consecutive patients with cirrhosis and refractory ascites. Defined by lack of response to low sodium diet and spironolactone (400 mg/day) plus furosemide (160 mg/day) or appearance of diuretic-induced complications. Excluded patients with chronic hepatic encephalopathy.	Intervention: TIPS (n=35). A total paracentesis and albumin performed before TIPS to facilitate procedure. Initially dilated to 8 mm diameter further dilated to 10 mm if portocaval pressure gradient remained above 12 mmHg. Then if still above 12 mmHg a second stent was placed. Comparison: total paracentesis plus intravenous albumin (8 g/litre removed). Patients in the paracentesis group with past history of variceal bleeding and/or moderate or large oesophageal varices were treated with beta-blockers (n=35). All patients: received diuretics during follow-up only if urine sodium under diuretic therapy was >20 mEq/day (doses not provided). Prophylactic antibiotic therapy was given throughout the study period to patients with past history of bacterial peritonitis or for 7 days to patients developing gastrointestinal bleeding.	Survival without liver transplant Recurrence of ascites Spontaneous bacterial peritonitis (SBP) Renal failure Hepatic encephalopathy	SBP and transplant-free survival not reported in Cochrane so extracted separately
Narahara 2011[139] RCT Japan Mean duration of follow-up in intervention group was 828 days and 388 days in the control group	Consecutive Japanese patients with cirrhosis and refractory ascites who have good hepatic and renal function. Excluded patients with episodes of hepatic encephalopathy.	Intervention: TIPS (n=30); expandable stent placed and dilated to obtain a portosystemic pressure gradient of below 12 mmHg. Stent was initially dilated to 6 or 8 mm in diameter. If the pressure remained above 12 mmHg the stent was further dilated to 8 or 10 mm. Comparison: LVP plus albumin (n=30) received sodium restriction and treatment with	Survival overall Recurrent ascites (note: study defined complete response as the elimination of ascites – therefore the number of people that did not have a complete response were	Did not use a 'covered' stent as they were unavailable. When a treatment failure was recorded, the patient was allowed to cross over

Study	Population	Intervention and comparison	Outcomes	Comments
		diuretics. LVP was performed along with intravenous infusion of albumin (6 g/litre ascites) removed. All patients received diuretics before and after randomisation but doses were adjusted according to clinical need.	calculated as having recurrence of ascites) Hepatic encephalopathy	to the other treatment. Reported survival overall but not transplant-free survival (no patients had liver transplantation)
Rossle 2000[179] RCT Included in Cochrane Saab 2006[181] Germany Followed up to 60 months after inclusion	Consecutive patients with cirrhosis and refractory ascites. Patients were considered to be refractory if they did not have a response after 4 weeks of treatment with spironolactone 400 mg/day plus furosemide 120 mg/day or were intolerant to treatment. Excluded patients with hepatic encephalopathy of grade 2 or higher.	Intervention: TIPS (n=29) Balloon expandable Palmaz-Schatz stent (Johnson & Johnson) implanted in 21 patients and a self-expandable nitinol stent (Memotherm) in 8 patients. Mean stent diameter was 9 mm. Anticoagulant therapy with intravenous heparin (with adjustment to achieve partial thromboplastin time of 40–60 seconds) for 1 week and low molecular weight heparin (0.3–0.4 ml) for 4 weeks. After shunt placement, the doses of diuretics were adjusted according to clinical need (assessed in terms of urine production, body weight and the presence or absence of oedema). In the case of shunt insufficiency, reestablishment of the shunt was only performed when severe ascites returned. Comparison: (n=31) large-volume paracentesis (4 or more litres) with albumin infusion only when clinically indicated (8 g/litre removed). During follow-up patients assigned to paracentesis received dietary treatment and treatment with diuretics given at tolerable doses.	Transplant-free survival Hepatic encephalopathy Hospital readmission	Hospital readmission and transplant-free survival not reported in Cochrane so extracted separately

Study	Population	Intervention and comparison	Outcomes	Comments
Salerno 2004[185] RCT Included in Cochrane Saab 2006[181] Italy Patients were followed up on average 18.2 ± 2.3 months after inclusion	Consecutive patients with cirrhosis and refractory or recidivant ascites. Patients were considered to be refractory if there was a lack of response to a low sodium diet and spironolactone 400 mg/day plus furosemide 160 mg/day. Patients were considered recidivant by the recurrence of at least 3 episodes of tense ascites within a 12-month period despite low sodium diet and adequate diuretic doses. Excluded if had history of recurrent episodes of hepatic encephalopathy of grade 2 or higher.	Intervention: TIPS (n=33). Expandable stent (Wallstent or Memotherm) was placed and dilated to obtain a portal pressure gradient below 12 mmHg. Comparison: LVP with albumin infusion (8 g/litre removed) (n=33) All patients: diuretic drugs were not withdrawn, but their doses adjusted according to the clinical need and tolerability of the patient. A low sodium diet (80 mEq/day) was prescribed throughout the study period. Recurrent tense ascites was treated with LVP and albumin (with TIPS angiography and stent patency restored when portocaval pressure gradient >12 mmHg)	Transplant-free survival Re-accumulation Hepatic encephalopathy Hospital readmission	Hospital readmission and transplant-free survival not reported in Cochrane so extracted separately. When a failure of treatment the patient could cross to the other treatment.
Sanyal 2003[190] RCT Included in Cochrane Saab 2006[181] USA, Canada Patients followed up to 12 months after inclusion	Consecutive patients with cirrhosis and refractory ascites, which was defined according to the International Ascites Club criteria. Excluded patients with active encephalopathy (grade 2 or higher).	Intervention: TIPS (n=52). Prior to TIPS, total paracentesis with intravenous infusion of albumin (6-8 mg/litre removed). Details of TIPS procedure not reported (mean decrease in HVPG from 19.8 (4.8) to 8.3 (3.6) mmHg. Remained on restricted sodium diet and treatment for diuretics (as below). Underwent angiography at 6 and 12 months for sonographic or clinical suspicion of recurrent portal hypertension. Comparison: total paracentesis with intravenous infusion of albumin (6-8 g/litre removed) as required, restriction of sodium and treatment with diuretics (as below).	Transplant-free survival Re-accumulation Quality of life SBP Acute renal failure Hepatic encephalopathy Hospital readmission	Hospital readmission, quality of life, SBP and transplant-free survival not reported in Cochrane so extracted separately

Study	Population	Intervention and comparison	Outcomes	Comments
		All patients placed on a restricted sodium diet (50–66 mEq/day) and treatment with diuretics (4-step combination of loop-acting furosemide 40–160mg/day and distal acting spironolactone 100–400mg/day was given with dose escalation by one step at a time permitted for >10lb weight gain). Repeat total paracentesis with infusion of albumin was performed in both groups for tense, symptomatic ascites with weight gain >10lb from immediately previous nadir weight despite maximal diuretic therapy or inability to use an effective dose of diuretics due to diuretic-related side effects.		

Table 87: Clinical evidence summary: TIPS versus LVP

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with TIPS versus LVP (95% CI)
Ascites re-accumulation[6]	305 (4 studies) 12 months	Low[b,c] due to inconsistency, imprecision	RR 0.57 (0.40 to 0.82)	888 per 1000	382 fewer per 1000 (from 160 fewer to 533 fewer)
Quality of life – physical score (scale not reported, better indicated by lower values)	109 (1 study) 12 months	Low[c,d] due to risk of bias, imprecision	–	The mean quality of life – physical score in the control groups was 5.69	The mean quality of life – physical score in the intervention groups was 3.36 lower (7.53 lower to 0.81 higher)
Quality of life – mental score SF-36 score (scale not reported, better indicated by lower values)	109 (1 study) 12 months	Very low[c,d] due to risk of bias, imprecision	–	The mean quality of life – mental score in the control groups was 3.96	The mean quality of life – mental score in the intervention groups was 2.13 lower (5.45 lower to 1.19 higher)
Transplant-free survival[7]	365 (5 studies)	Low[c,e] due to inconsistency, imprecision	HR 0.58 (0.35 to 0.96)	653 per 1000[a]	194 fewer per 1000 (from 15 fewer to 343 fewer)
Spontaneous bacterial peritonitis	179 (2 studies)	Low[c] due to imprecision	RR 1.05 (0.35 to 3.1)	75 per 1000	4 more per 1000 (from 49 fewer to 157 more)
Acute renal failure	179 (2 studies)	Moderate[c] due to imprecision	RR 0.64 (0.35 to 1.18)	260 per 1000	94 fewer per 1000 (from 169 fewer to 47 more)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with TIPS versus LVP (95% CI)
Hepatic encephalopathy[8]	365 (5 studies)	Low[c,i] due to inconsistency, imprecision	RR 1.64 (1.14 to 2.36)	355 per 1000	227 more per 1000 (from 50 more to 483 more)

(a) Calculated from the median control group rate at the end of study
(b) Downgraded by 1 increment because of heterogeneity, I^2=79%, p=0.003, unexplained by subgroup analysis
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(d) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(e) Downgraded by 1 increment because of heterogeneity, I^2=65%, p=0.02, unexplained by subgroup analysis
(f) One study defined complete response as the elimination of ascites – therefore the number of people that did not have a complete response were calculated as having recurrence of ascites
(g) One study reported overall survival but no patients had transplantation
(h) Hepatic encephalopathy was an exclusion criteria in all studies, however some studies reported new cases of hepatic encephalopathy, some worsening cases of hepatic encephalopathy and some relapse of hepatic encephalopathy
(i) Downgraded by 1 increment because of heterogeneity, I^2=58%, p=0.05, unexplained by subgroup analysis

Hospital readmissions were reported in 3 studies[179,185,190] but not in a format that could be combined in a meta-analysis. Salerno 2004[185] reported 28 out of 33 patients in the TIPS group were readmitted to hospital during follow-up compared to 30 out of 33 in the LVP group. The total number of days spent in hospital per patient in the TIPS group was 47 (±3.8) compared to 40 (±4.2) in the LVP group. Sanyal 2003[190] reported 3.2 (±3.1) unscheduled hospitalisation per patient per year in the TIPS group compared to 2.4 (±2.83) in the LVP group. Rossle 2000[179] reported that the TIPS group had a total of 52 (±29) days compared to 72 (±48) days hospital stay during follow-up in the LVP group.

12.4 Economic evidence

12.4.1 Published literature

One economic evaluation was identified with the relevant comparison and has been included in this review.[87] This is summarised in the economic evidence profile below (Table 88) and the economic evidence table in Appendix I.

See also the economic article selection flow chart in Appendix F.

Table 88: Economic evidence profile: TIPS versus LVP

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects[(c)]	Cost-effectiveness	Uncertainty
Gines 2002[87] (USA, Spain)	Partially applicable[(a)]	Potentially serious limitations[(b)]	• Cost-consequences analysis • Based on an RCT at 2 hospitals (US, Spain); n=70[(c)] • Time horizon: 2 years	£2,104[(d) (e) (f)]	Death: RR: 1.11; ARD: 57 more per 1,000 Ascites re-accumulation: RR: 0.59; ARD: 343 fewer patients per 1,000 RR: 0.18; ARD: 8,029 fewer episodes per 1,000 patients Renal failure: RR: 0.53; ARD: 229 fewer per 1,000 Spontaneous bacterial peritonitis: RR: 0.5; ARD: 57 fewer per 1,000 Hepatic encephalopathy: RR: 1.17; ARD: 114 more per 1,000	Death: LVP dominates (is cheaper and more effective than) TIPS Ascites re-accumulation: ICER: £6,137 per patient with re-accumulation averted; £262 per re-accumulation incident averted	No sensitivity analysis was undertaken. Differences in the outcomes of ascites re-accumulation and renal failure were significant (at a level of p=0.05); differences in death, SBP and hepatic encephalopathy were not (though significantly more patients in the TIPS group had severe hepatic encephalopathy).

Abbreviations: ARD: absolute risk difference; LVP: large-volume paracentesis; RCT: randomised controlled trial; TIPS: transjugular intrahepatic portosystemic shunt

(a) Study was partially conducted in USA – differences in healthcare system may make results less applicable to UK; discounting does not appear to have been used; no quality of life data collected

(b) Clinical outcomes and resource usage based on a single RCT; unit costs derived from a single Spanish hospital; costs associated with some complications were not included, unclear whether costs of hospitals stays were included; no sensitivity analysis conducted

(c) See clinical evidence in Section 12.3 for more information on the Gines et al. 2002 RCT

(d) US dollars converted using 2000 purchasing power parities[151]

(e) The cost year was not reported so it was assumed to be 2000 (study ended December 2000)

(f) The study presented Spanish and US costs; the Spanish costs are presented here as more applicable to the UK

12.4.2 Unit costs

See Table 95 in Appendix O.

12.5 Evidence statements

12.5.1 Clinical

- TIPS was considered to have a clinically important benefit over LVP for the following outcomes:
  - Re-accumulation of ascites at 12 months (4 studies, 305 patients, Low quality)
  - Transplant-free survival (5 studies, 365 patients, Low quality)
  - Acute renal failure (2 studies, 179 patients, Moderate quality)
- LVP was considered to have a clinically important benefit over TIPS for the following outcomes:
  - Hepatic encephalopathy (5 studies, 365 patients, Low quality)
- There was no clinically important difference between TIPS and LVP for the following outcomes:
  - Quality of life at 12 months (1 study, 109 patients, Low to Very Low quality evidence for the physical and mental SF-36 scores respectively)
  - Spontaneous bacterial peritonitis (2 studies, 179 patients, Low quality)

12.5.2 Economic

- One cost-consequences analysis found that TIPS was more costly than LVP for the management of diuretic-resistant ascites (£2,104 more per patient) and had 0.06 more deaths per patient, but 8.0 fewer instances of ascites re-accumulation per patient. This analysis was assessed as partially applicable with potentially serious limitations.

12.6 Recommendations and link to evidence

Recommendation	**24. Consider a transjugular intrahepatic portosystemic shunt for people with cirrhosis who have refractory ascites.**
Research recommendation	**3. What is the quality of life in people who have had a transjugular intrahepatic portosystemic shunt (TIPS)?**
Relative values of different outcomes	The GDG discussed the merits of various outcome measures that compared transjugular intrahepatic portosystemic shunt (TIPS) with large-volume paracentesis (LVP) plus albumin in people with cirrhosis and refractory ascites. Health-related quality of life, re-accumulation of ascites and transplant-free survival were critical outcomes of interest. Important outcomes were agreed as occurrence of spontaneous bacterial peritonitis, renal failure, hepatic encephalopathy, length of hospital stay and readmission rates.
Trade-off between clinical benefits and harms	There was clinical benefit in favour of TIPS for the critical outcomes of re-accumulation of ascites and transplant-free survival (albeit Low quality evidence). Only 1 study reported the critical outcome of health-related quality of life using the SF-36 tool. The paper reported that there was an improvement in scores for both groups, but there was no significant difference between the two groups. In the study, the SF-36 score was broken down into physical and mental components but the scale for each was not reported. Therefore, the GDG found it difficult to assess whether the difference between groups was of clinical importance. Although there was a lack of evidence for quality of life, the GDG discussed the

	potential benefits of a successful TIPS procedure over repeated LVP. For LVP, patients are required to attend hospital regularly every few weeks. This procedure is not a cure for the refractory ascites. The GDG discussed that avoiding repeated LVP is likely to improve quality of life in the majority of patients who are able to undergo the TIPS procedure. Given this, the GDG felt able to make a 'consider' recommendation. Evidence was available from 2 studies for the outcome of renal failure (Moderate quality) and SBP (Low quality), with contradictory results in small numbers of patients. The GDG, despite the Moderate quality, placed less weight on these results given the small number of reported outcomes. When considering the evidence the GDG noted that the data available were largely older studies and of Low quality. The GDG felt that if an RCT had been conducted now, patient selection (particularly screening for hepatic encephalopathy and cardiac dysfunction) would have been more stringent and may have shown more benefit for TIPS. The GDG also noted that technical failures affected the outcomes in the TIPS groups and modern radiological techniques and stent design may reduce the effect of these technical failures. For example, even the most recent study (2011) used uncovered stents which are more likely to stenose. In addition, older studies may have included patients with more severe liver disease, who would not be considered for TIPS now. The GDG discussed the benefits of TIPS in improving the nutritional status of patients with cirrhosis. Whilst the GDG did not look for evidence about nutritional status following both procedures, anecdotally it was felt that, when TIPS was successful, it indirectly improved the person's nutritional status. The GDG noted the suggestion of some benefit of TIPS for health-related quality of life, although this was not clinically significant. The GDG noted that in patients with hepatic encephalopathy there was a clinical benefit of LVP over TIPS. Hepatic encephalopathy can become disabling and require reversal of the TIPS procedure by blockage of the stent. The GDG noted that the TIPS procedure is unsuitable for people with a past or current history of hepatic encephalopathy.
Trade-off between net clinical effects and costs	One cost–consequence analysis (Gines 2002)[87] was identified comparing LVP with TIPS in patients with refractory ascites (not responding to low sodium diet). This study reported that TIPS was more costly than LVP for the management of diuretic-resistant ascites (£2104 more per patient) and had 0.06 more deaths per patient, but 8.0 fewer instances of ascites re-accumulation per patient. The difference in total costs was mainly due to the high procedural TIPS costs outweighing lower follow-up costs in comparison to LVP. The GDG agreed that the largest component to the cost of LVP would be albumin infusions and the main follow-up costs related to TIPS would be the cost of admission related to episodes of hepatic encephalopathy. The GDG also noted that this study was the only study included in the clinical review not favouring TIPS in terms of transplant-free survival, and that the TIPS procedural methodology has improved since the publication of the study meaning that the reported clinical effectiveness data are not representative of the clinical review. The GDG considered that with clinical outcomes more favourable to TIPS, the follow-up costs of TIPS would be lower, although the initial costs of TIPS would always be higher than for LVP. The relative costs of TIPS and LVP will depend on the number of times LVP needs to be repeated – this can vary very largely between individuals, with some people needing frequent repetition and others re-accumulating fluid very slowly. For example, with TIPS costing £2900 and LVP costing £670 (costs from Parker 2013[157]), a TIPS procedure that prevented more than 4 LVP procedures would be cost saving in addition to decreasing mortality. The GDG concluded that the cost-effectiveness of TIPS would vary from person to person depending on their individual circumstances, particularly their suitability for

	TIPS and the frequency of their need for LVP. In many patients TIPS would be likely to be cost-effective at a threshold of £20,000 per QALY gained as the benefit in transplant-free survival, and the benefit to quality of life from reduced re-accumulation of ascites are likely to justify the increased costs of TIPS. In patients who prior to TIPS had need for repeated LVP, TIPS would even be cost saving. As a result, clinicians should use their judgement regarding which people may be most suitable for TIPS, but for those for whom TIPS is suitable the GDG believes that TIPS should be considered by the clinician and patient.
Quality of evidence	Five RCTs were identified (4 had been included in a previous Cochrane review and 1 was a more recent RCT). The Cochrane review papers were individually assessed to identify data for the critical and important outcomes not reported by Cochrane authors. The evidence quality ranged from Very Low to Moderate, and for all critical outcomes the evidence quality was Low. For many outcomes, heterogeneity was observed (re-accumulation of ascites, transplant-free survival and hepatic encephalopathy). It was not possible to explore this heterogeneity using the predefined subgroup analyses due to the small number of studies in each subgroup. The population identified in the protocol was people with cirrhosis and diuretic refractory ascites. Three papers reported that people were receiving LVP prior to the placement of TIPS; however 2 studies did not report this in their inclusion criteria. It was assumed that this was the case in all studies. The GDG agreed that the phrases 'total paracentesis' and 'large-volume paracentesis' were, in practical terms, the same or a very similar procedure. A further study (RCT) by Lebrec 1996[119] was subsequently excluded by the GDG. This report was included in the Cochrane review, but the GDG noted that the procedure used for TIPS insertion was significantly different from current TIPS protocols. Individuals in the intervention arm received 2–3 uncovered stents in parallel with the aim of creating a larger shunt than in many other studies. The paper was excluded from the analysis. The GDG noted that the absolute number of patients in this paper was small. In the paper by Gines 2002[87] the LVP group continued to receive beta-blockers and the TIPS group did not. The GDG noted that beta-blockers can pre-dispose to hepatic encephalopathy and are associated with mortality in advanced cirrhosis. For the critical outcome of re-accumulation of ascites either 'complete response – a complete resolution of ascites', or 'incomplete response – leaving some residual ascites' were considered a positive effect. However, studies used different definitions for the levels of response and this contributed to heterogeneity in the data (represented by the high I^2 value). Trial inclusion criteria selected patients who were 'fit for TIPS' which may have improved the outcomes in both groups to above that which could be expected in clinical practice. There was variation in the operational definitions used within studies, for example, hepatic encephalopathy was defined by some studies as 'Grade 1' or 'overt'; others reported new hepatic encephalopathy whilst others cited 'on-going' hepatic encephalopathy. This may contribute to the heterogeneic results for the outcome of hepatic encephalopathy. The GDG also noted that in the Gines 2002[87] study (in which LVP patients were given beta-blockers), for hepatic encephalopathy, less benefit was reported for LVP than in other studies. In addition, variation between studies in the use of covered or uncovered stents and in the use of 'low salt' or 'no added salt' diets was highlighted by the GDG.
Other considerations	The GDG emphasised that all patients with cirrhosis and refractory ascites should be reviewed by a hepatologist and considered for transplantation. Those who are suitable for transplantation may undergo TIPS as a 'holding procedure' while on the transplant waiting list. Those who are not suitable for transplantation would undergo TIPS as a definitive procedure. The GDG was in agreement that there is currently wide variation in UK practice and were concerned that there are patients who may

	benefit from TIPS but who are not being offered this service. **Research recommendation** Prior to TIPS, people may have had several problems resulting from portal hypertension, including variceal bleeding from veins in the stomach, oesophagus, or intestines, ascites or hydrothorax – all of which will have had a detrimental effect on their quality of life. TIPS should alleviate these problems, but little is known about the consequential effect on quality of life and any effects that potential problems following TIPS (for example, hepatic encephalopathy, shunt blockages, infection and cardiac problems) have on each person. It is therefore important to assess what benefits TIPS has to the quality of life of people with advanced liver disease.

13 Primary prevention of spontaneous bacterial peritonitis (SBP) in people with cirrhosis and ascites

13.1 Introduction

Bacterial infections are common in people with cirrhosis and are more frequent in people with decompensated disease. Once infection develops, renal dysfunction, hepatic encephalopathy and a deterioration of liver function may follow, which adversely affect survival. In people with cirrhosis, a bacterial infection is present either at the time of hospital admission or during hospitalisation in 25–35% of people[63] and bacterial infection increases the probability of death, reaching 30% at 1 month and 63% by 1 year.[12]

The most common infections in people with cirrhosis are spontaneous bacterial peritonitis (SBP) (25%), followed by urinary tract infection (20%), pneumonia (15%), bacteraemia (either spontaneous or following a therapeutic procedure), and cellulitis.[104] Clinical factors associated with an increased risk of infection are severity of liver disease, low ascitic fluid protein levels, prior SBP and hospitalisation.[63] Enterobacteriaceae and non-enterococcal streptococci cause the majority of spontaneous infections in cirrhosis[12,63] and translocation of enteric bacteria from the gut appears to be the main cause of Gram-negative infections in cirrhosis.[104]

Since most episodes of SBP in cirrhosis result from the translocation of enteric bacteria, the prophylactic administration of an oral antibiotic that effectively reduces the concentration of these bacteria in the gut whilst preserving the protective anaerobic flora could reduce the incidence of SBP. However, the use of prophylactic antibiotics is associated with a risk of developing multi-resistant organisms and infections caused by antibiotic-resistant bacteria have a higher hospital mortality than those caused by susceptible bacteria.[63] Since these factors lead to uncertainty around the use of antibiotics for the primary prevention of SBP in people with cirrhosis and ascites, the GDG wanted to examine the clinical and cost-effectiveness of oral antibiotics compared to placebo in patients with cirrhosis at risk of SBP associated with the presence of ascites.

13.2 Review question: What is the clinical and cost-effectiveness of antibiotics compared with placebo for the primary prevention of spontaneous bacterial peritonitis (SBP) in people with cirrhosis and ascites?

For full details see the review protocol in Appendix C.

Table 89: PICO characteristics of review question

Population	Adults and young people (16 and over) with confirmed cirrhosis and ascites
Intervention	**Prophylactic antibiotics including:** Oral antibiotic therapy until resolution of the ascites: • Quinolones (ciprofloxacin, norfloxacin, pefloxacin, ofloxacin, floxacin) • Penicillins (amoxicillin, co-amoxiclav (Augmentin)) • Sulfonamides (co-trimoxazole [trimethoprim/sulphamethoxazole, Septrin]) • Third-generation cephalosporins (cefalexin)
Comparisons	Placebo or no treatment

Outcomes	**Critical outcomes:** • Occurrence of SBP (determined from abdominal paracentesis with ascitic fluid neutrophil [polymorph] counts ≥250 cells/mm^3 [0.25×10^9 per litre] on microscopy) • Health-related quality of life • All-cause mortality (time-to-event) **Important outcomes:** • Adverse event: incidence of resistant organisms • Adverse effect: renal failure • Adverse effect: liver failure • Length of hospital stay • Readmission rate
Study design	RCTs Systematic reviews of RCTs

13.3 Clinical evidence

One Cochrane review[41] (comprised of adults with cirrhosis and ascites regardless of aetiology of cirrhosis or severity of the disease) included 9 studies overall, however only 5 relevant studies were included in this review;[66,92,175,208,224]. One additional RCT published since the Cochrane review was identified and included in this review.[223] The included studies are summarised in Table 90 below. Evidence from the included studies is summarised in the clinical evidence summary below (Table 91). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, GRADE tables in Appendix J and excluded studies list in Appendix L.

Two studies included in the Cochrane review were excluded from this review because they looked at a head-to-head comparison of different antibiotics, not a placebo controlled comparison. The other 2 studies included in the Cochrane review[86,204] were excluded from this systematic review because they included a large proportion (>15%) of patients who previously had SBP. Two papers,[175,208] with small proportions of people with prior SBP, were included in this review. One study, Rolachon et al. (1995),[175] reported how many patients developed SBP who also had SBP previously, so it was possible to amend the analysis to include primary prophylaxis only. The numerator and denominator were amended in the analysis for this outcome. For other outcomes it was not possible to determine how many cases in the numerator had previously had SBP and hence this study was considered as having an indirect population. For Soriano et al. (1991),[208] it was unclear whether or not those who developed SBP had primary or secondary antibiotic prophylaxis, but as there were only 7 patients who developed SBP in the study (in the placebo group) and only 1 patient in this group had SBP previously, it was unlikely to alter the overall result substantially. Sensitivity analysis was performed by removing this participant and there was minimal effect.

All the studies used a combination of clinical, laboratory, and ultrasonographic data or histology to confirm cirrhosis. The participants were typically Child-Pugh B or C (7 or greater) but some studies included small numbers of participants who were Child-Pugh A (1 study[92] did not report severity of cirrhosis). Five of the studies included people at high risk of SBP, defined by having an ascitic protein concentration of <15 g/litre. One study[223] included people at low risk of SBP, defined by having an ascitic protein concentration of ≥15g/litre.

All studies compared a quinolone (either norfloxacin or ciprofloxacin) to placebo, usually administered for between 6 and 12 months, but 1 study[208] treated patients for the hospitalised period only (<4 weeks) and another study only treated people for 1 month.[223]

Some outcomes prioritised for this review were not reported in the Cochrane review so these were extracted from the individual papers. These include time-to-event data for all-cause mortality,

incidence of antibiotic-resistant organisms (the Cochrane review reported this for 2 relevant studies[175,208]), liver failure, renal failure, and length of hospital stay. There was no evidence found on health-related quality of life or on readmission rates.

For the 3 studies that reported all-cause mortality as time-to-event data it was possible to calculate the hazard ratio and these were combined in a meta-analysis. Since 3 of the studies did not report time-to-event all-cause mortality data, dichotomous data were presented in a separate forest plot. As each study reported at varying follow-up times, the results from these studies were not combined into a meta-analysis but were instead presented separately as subgroups by follow-up duration.

Four studies looked at antibiotic resistance but did not report in a format that could be combined in a meta-analysis. One study[66] reported no cases of quinolone-resistant SBP in the 12-month study. However, it was unclear what investigations were performed or if investigations were performed for all patients. Two studies[92,175] took stool samples from a proportion of patients to detect resistant microorganisms, and it was unclear why not all patients were tested. One of these studies [175] tested 10 of the 28 patients who were treated with antibiotics and reported that none had acquired resistance to ciprofloxacin after 6 months. The other study[92] was a multi-centre study that tested 46 patients at 5 centres (24/53 antibiotic, 22/54 placebo) and found that 42% (10/24) of those tested who were treated with antibiotics had norfloxacin-resistant organisms at some point during the 6-month treatment period compared with 14% (3/22) of those in the placebo group. The final study[223] reported that *E.Coli* resistant to ciprofloxacin was found in 6 out of the 7 people who developed a urinary tract infection during the trial.

All studies that reported liver failure only did so where this outcome lead to death, so this was extracted. While the I^2 statistic was at an acceptable level for the forest plot on liver failure (see forest plot K.5 in Appendix K), there appeared to be heterogeneity in point estimates in opposite directions (that is, favouring opposing treatment). It was not possible to explore heterogeneity in the pre-specified subgroups (see the review protocol in Appendix C) since all treatments were in the same antibiotic class and trial participants had a similar level of risk of ascites (defined as <15 g/litre ascitic protein). Any subgrouping by severity would result in only 1 study in each subgroup, so this subgroup analysis was not performed. As it was not possible to explain the heterogeneity, a random effects meta-analysis was used and the outcome was downgraded for inconsistency.

Heterogeneity was also found for the meta-analysis on length of hospital stay, where each study reporting the outcome had point estimates favouring opposing treatments. It was not possible to explore heterogeneity for this outcome as there were only 2 studies, and any possible subgroup analysis would leave only 1 study in each subgroup. As it was not possible to explain the heterogeneity, a random effects meta-analysis was used and the outcome was downgraded for inconsistency.

Table 90: Summary of studies included in the review (all also included in Cochrane review[41])

Study and country	Population	Level of protein in ascites (g/litre)	Intervention and comparison (treatment period, follow-up)	Outcomes	Comments
Fernandez 2007[66] Spain	74 adults with cirrhosis and ascites Child-Pugh score: mean 9.9 (SD 1.5) versus 10.4 (SD 1.5) (included ≥9 only)	<15g/litre Mean (SD): 9 (4) versus 9 (3)	n=38 oral norfloxacin (400 mg/day) versus n=36 placebo (both for 12 months)	SBP All-cause mortality (time-to-event) Incidence of bacterial microorganisms Renal failure Liver failure	
Grange 1998[92] France	107 adults with cirrhosis and ascites Child-Pugh score: unclear[(a)]	<15g/litre Mean (SD): 9.3 (2.9) versus 10.4 (2.8)	n=53 oral norfloxacin (400 mg/day) versus n=54 placebo (both for 6 months[(b)])	SBP All-cause mortality (dichotomous) Incidence of bacterial microorganisms Liver failure	Unclear about severity of cirrhosis in patients
Rolachon 1995[175] France	60 adults with cirrhosis and ascites Child-Pugh score: 1A/35B/24C	<15g/litre Mean (SD): 9.4 (2.9) versus 10.3 (2.8)	n=28 oral ciprofloxacin (750 mg/day) versus n=32 placebo (both for 6 months)	SBP All-cause mortality (dichotomous) Incidence of bacterial microorganisms Liver failure Length of hospital stay	6% of patients had SBP previously[(c)]
Soriano 1991[208] Spain	63 adults with cirrhosis and ascites Child-Pugh score: 3A/27B/33C	<15g/litre Mean (SD): 7.1 (3.4) versus 6.7 (2.9)	n=32 oral norfloxacin (400 mg/day) versus n=31 placebo (for hospitalisation period: mean 26.75 [SD 14.87] versus 23.68 [SD 13.48] days)	SBP All-cause mortality (dichotomous) Length of hospital stay	11% of patients had SBP previously[(c)]
Tellez-Avila 2013[223] Mexico	95 adults with cirrhosis and ascites Child-Pugh score: 14A/62B/19C	≥15g/litre	n=49 oral ciprofloxacin (500 mg/day) Versus n=46 placebo	SBP All-cause mortality (time-to-event)	

Study and country	Population	Level of protein in ascites (g/litre)	Intervention and comparison (treatment period, follow-up)	Outcomes	Comments
			(both for 1 month with 6 month follow-up)		
Terg 2008[224] Argentina	100 adults with cirrhosis and ascites Child-Pugh score: mean 8.5 (SD 1.5) versus 8.3 (SD 1.3)	<15g/litre Mean (SD): 8.4 (3.1) versus 8.5 (3.6)	n=50 ciprofloxacin (500 mg/day) versus n=50 placebo (both for 12 months)	SBP All-cause mortality (time-to-event) Renal failure Liver failure	

(a) Most had advanced cirrhosis with a history of complications which was likely to mean that most have a minimum Child-Pugh score of B
(b) Study reported this to be 4.2 versus 4.4 months follow-up; this may be the average follow-up, taking into account those that stopped treatment for reasons such as death
(c) These were included because it was either possible to extract data on primary prophylaxis only or because the numbers of patients who had secondary prophylaxis was so small that it was unlikely to alter the results substantially (see text)

Table 91: Clinical evidence summary: prophylactic oral antibiotics versus placebo

Outcomes	Number of participants (studies), Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with antibiotic prophylaxis versus placebo or no treatment (95% CI)
Spontaneous bacterial peritonitis	482 (6 studies) 6 months	MODERATE[a] due to risk of bias	RR 0.22 (0.11 to 0.46)	187 per 1000	146 fewer per 1000 (from 146 fewer to 166 fewer)
All-cause mortality (time-to-event)	263 (3 studies) 6-12 months	HIGH	HR 0.40 (0.22 to 0.73)	280 per 1000	157 fewer per 1000 (from 67 fewer to 210 fewer)
All-cause mortality (dichotomous) - Mortality at ~1-month follow-up	59 (1 study) 25.5 days	VERY LOW[a,c,d,e] due to risk of bias, indirectness, imprecision	RR 0.39 (0.08 to 1.85)	167 per 1000	102 fewer per 1000 (from 154 fewer to 142 more)

Outcomes	Number of participants (studies), Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with antibiotic prophylaxis versus placebo or no treatment (95% CI)
All-cause mortality (dichotomous) - Mortality at ~4-month follow-up	107 (1 study) 132 days	VERY LOW[a,c,e] due to risk of bias, indirectness, imprecision	RR 0.82 (0.35 to 1.91)	185 per 1000	33 fewer per 1000 (from 120 fewer to 168 more)
All-cause mortality (dichotomous) – Mortality at 6-month follow-up	53 (1 study) 6 months	LOW[a,c,d,e] due to risk of bias, indirectness, imprecision	RR 0.76 (0.24 to 2.43)	222 per 1000	53 fewer per 1000 (from 169 fewer to 317 more)
Adverse event: renal failure	168 (2 studies) 12 months	LOW[a,e] due to risk of bias, imprecision	RR 0.54 (0.31 to 0.96)	332 per 1000	153 fewer per 1000 (from 13 fewer to 229 fewer)
Adverse event: liver failure	328 (4 studies) 8.5 months	VERY LOW[a,e] due to risk of bias, imprecision	RR 1.43 (0.54 to 3.79)	35 per 1000	15 more per 1000 (from 16 fewer to 98 more)
Length of hospital stay	123 (2 studies) 3.4 months	VERY LOW[a,e,f] due to risk of bias, inconsistency, imprecision	–	The mean length of hospital stay in the control groups was 20.8 days	The mean length of hospital stay in the intervention groups was 3.12 lower (14.15 lower to 7.92 higher)

(a) *Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias*
(b) *Calculated from median control group rate at 12 months*
(c) *The majority of the evidence had indirect outcomes*
(d) *The majority of the evidence had an indirect population*
(e) *Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs*
(f) *Downgraded by 1/2 increments because the confidence intervals across studies show minimal or no overlap and heterogeneity, $I^2 \geq 50\%$, $p \leq 0.05$, unexplained by subgroup analysis*

13.4 Economic evidence

13.4.1 Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

13.4.2 Unit costs

See Tables 96 and 97 in Appendix O.

13.5 Evidence statements

13.5.1 Clinical

- Antibiotics were considered more clinically effective than placebo for the following outcomes:
  - Incidence of spontaneous bacterial peritonitis (6 studies, n=482, Moderate quality)
  - All-cause mortality at:
    - all time points (time-to-event data from 3 studies, n=263, High quality)
    - ~1 month (n=59), ~4 months (n=107), 6 months (n=53), all one study, dichotomous data and Low to Very Low quality.
  - Renal failure (2 studies, n=168, Low quality).
- It was unclear whether or not antibiotics had a clinical benefit on liver failure leading to death or length of hospital stay compared to placebo (liver failure: 4 studies, n=328, Very Low quality; length of hospital stay: 2 studies, n=123, Very Low quality).
- There was no evidence on health-related quality of life.

13.5.2 Economic

- No relevant economic evaluations were identified.

13.6 Recommendations and link to evidence

Recommendation	**25. Offer prophylactic oral ciprofloxacin or norfloxacin[b] for people with cirrhosis and ascites with an ascitic protein of 15 g/litre or less, until the ascites has resolved.**
Research recommendation	**4. How frequently does antibiotic resistance occur, and how significant are antibiotic treatment-related complications when antibiotics are used for the primary prevention of spontaneous bacterial peritonitis in people at high risk of having, or developing, cirrhosis?**
Relative values of different outcomes	The main purpose of the use of antibiotics in people with cirrhosis and ascites is to prevent the occurrence of spontaneous bacterial peritonitis (SBP). The GDG was particularly interested in the impact of antibiotics on these patients in terms of their

[b] At the time of publication (July 2016), neither ciprofloxacin nor norfloxacin had a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

	quality of life and whether or not antibiotics have an impact on overall mortality in these very sick patients. The GDG agreed critical outcomes of interest as the incidence of SBP, quality of life and all-cause mortality. Length of hospital stay, readmission rates, incidence of resistant organisms, renal failure, and liver failure were agreed to be important additional outcome measures which may be influenced by antibiotic treatment. The GDG noted that people with cirrhosis and ascites are generally managed in the outpatient setting but the development of SBP necessitates hospital admission. The GDG was interested not only in hospital admission rates but also whether length of hospital stay, associated with the occurrence of SBP or adverse events, was reduced for patients receiving antibiotics. As there is a risk of the development of antibiotic resistance, particularly with the long-term use of antibiotics, the GDG felt that it was important to consider this outcome. People with cirrhosis and ascites are at risk of renal or liver failure, so the GDG wished to consider whether the use of antibiotics influenced the occurrence of these complications. Length of hospital stay and admission rates were agreed to be important outcomes as they can be used as surrogates for quality of life and for the cost of treating SBP and any other complications or adverse effects.
Trade-off between clinical benefits and harms	The evidence presented identified a clinical benefit of prophylactic antibiotics on the critical outcomes of SBP and all-cause mortality for people with cirrhosis and ascites who were judged to be at a high risk of SBP (an ascitic fluid protein level of less than 15 g/litre). No evidence was available for the critical outcome of health-related quality of life. A clinical benefit of antibiotic prophylaxis was also seen in this group when assessing the important outcome of renal failure. The clinical effectiveness of antibiotics on the outcomes of liver failure and length of hospital stay was unclear. Overall, the GDG agreed that the clinical benefit observed for mortality and occurrence of SBP warranted the use of prophylactic antibiotics in people with cirrhosis, ascites and a high risk of SBP. The GDG acknowledged that this recommendation represents a change in UK practice, as only secondary prophylaxis is currently recommended. The mean length of treatment in most of the included studies varied between 6 and 12 months. However, the GDG recommended that treatment be continued for the total duration of time an individual has ascites. Antimicrobials should only be withdrawn once ascites has been successfully resolved, either by a re-compensation of the liver disease, sodium restriction and diuretic therapy, TIPS insertion or liver transplantation. The GDG noted 4 trials that reported data on the development of antibiotic resistance. The data were not sufficient to influence the GDG's recommendation as this outcome was only reported from a proportion of people in the trials; however, the GDG did not underestimate the importance of this outcome for healthcare professionals in deciding whether to initiate primary prophylaxis. It was noted that the Grange 1998 paper[92] reported a 14% incidence in the development of norfloxacin-resistant organisms during the 6-month follow-up in the placebo group and a higher percentage in the treatment group (42%) among the patients whose stool samples they tested. However, stool samples were tested from fewer than half of patients. The GDG commented that all-cause mortality outcome data should include any instances of fatal antibiotic-resistant infections occurring during the treatment period. Evidence from 1 study was identified in people with ascites and a lower risk of SBP (ascitic fluid protein level of 15 g/litre or more). This study showed a similar benefit of antibiotics on the all-cause mortality to that in studies in people at high risk. However, unlike studies in people at high risk, this study appears to show no benefit of antibiotics on the outcome of SBP, albeit with very large confidence intervals. This apparent difference in the effectiveness of antibiotics may reflect the fact that people were at lower risk, or may be due to other factors such as the shorter treatment period of 1 month in this study. As this study carried a low weighting in the meta-analysis, there was no statistical heterogeneity in the outcome of SBP

	overall. Consequently, as this study carried little weighting and the majority of the evidence was in people at high risk of SBP, the GDG made their recommendation in people at high risk. The GDG chose not to make a recommendation for people with low risk of SBP. No evidence was identified for other antibiotics such as penicillin or other quinolones. The GDG acknowledged that they could not be certain about the clinical effectiveness of other antibiotics and therefore chose to limit the recommendation to the use of either oral ciprofloxacin or oral norfloxacin.
Trade-off between net clinical effects and costs	No relevant economic evaluations were identified. The GDG noted the low cost of therapy with ciprofloxacin (£40 for 500 mg per day for a year) and moderate cost of norfloxacin (£313 for 400 mg per day for a year), and that both costs and benefits of treatment could continue indefinitely for as long as the patient had ascites, which would vary considerably between individuals. However, the GDG also noted the high cost of treating an episode of SBP. Using a 5-day course of piperacillin/tazobactam, together with a 7-day hospital stay, paracentesis and an ultrasound would cost £1,925 to treat a single episode of SBP (see unit costs in Appendix O). This would be cost saving if prophylactic treatment with ciprofloxacin prevented just 1 case of SBP per 71 person-years of prophylactic treatment, and for norfloxacin 1 case of SBP per 6 person-years of prophylactic treatment. The clinical evidence reviewed in this chapter found prophylactic treatment to reduce incidence of SBP by an average of 142 cases per 1000 people (1 in every 7 people) within 6 months, and thus it would be cost saving with the average reduction of SBP in published studies. The benefits to all-cause mortality shown above make this intervention even more favourable. The GDG was therefore confident that the reduction in episodes of SBP and the costs associated with treating SBP would make primary prophylaxis using ciprofloxacin or norfloxacin cost saving or at worst highly cost-effective compared to no prophylaxis at a cost-effectiveness threshold of £20,000 per QALY gained.
Quality of evidence	Six studies were identified from the evidence search. The GDG noted that the studies spanned a 22-year period and that 3 studies were over 15 years old. The most recent study was from 2013. All but 1 study was included in a Cochrane review. The quality of available evidence for the critical outcome of the incidence of SBP was Moderate. The quality of the evidence for all-cause mortality (time-to-event) was of High quality. The quality of evidence for all other outcome measures was low or very low. For dichotomous outcomes, most studies were at risk of bias due to a comparable or a higher rate of drop-outs. While it was unsurprising that many patients dropped out given that these patients were very sick, there is still a risk that the effect estimate does not reflect the true effect. These patient drop-outs also had a larger effect on the outcome because of the small size of the trials, which are difficult to recruit to given the relatively rare occurrence of SBP. Furthermore, the GDG noted a smaller number of occurrences of liver failure than expected in both groups. They felt that this may be partly due to the small size of the studies which is likely to have influenced the imprecision for this outcome. It was noted that in 2 of the studies participants were included even if they had been diagnosed with SBP previously (secondary prophylaxis). The GDG agreed that these studies could be included, but only if 15% or fewer of the participants had previously had SBP. Two papers which were included in the Cochrane review were excluded from this review on this basis. In Gines 1990,[86] all participants had prior SBP. In Singh 1995,[204] 14% in the treatment group and 17% in the no prophylaxis group had previously had SBP. The GDG noted that Singh 1995 was the only paper that compared trimethoprim-sulfamethoxazole with a placebo, but that it had been excluded, because of the pre-specified conditions in the protocol. The GDG also commented on the fact that a small number of participants were classified as Child-Pugh Grade A in the available evidence. The GDG felt that this may

	have been a mistake, as people with ascites would normally be classified as a minimum of Child-Pugh B. The GDG believed that the small number of participants who were classified as Child-Pugh A did not allow any recommendation to be made for low risk people with cirrhosis. The GDG noted the lack of evidence to assess the efficacy of prophylaxis in low risk groups (that is, with an ascitic fluid albumin level of greater than 15 g/litre). The GDG highlighted that the Soriano 1991 paper[208] described a short period of prophylaxis (mean 26 days), whilst the participants were inpatients for another decompensating event. One study[223] also reported a very short intervention period of 4 weeks which may explain the lack of effect of the antibiotic to prevent SBP. All other studies reported people receiving treatment for 6 to 12 months. The GDG agreed that Soriano 1991 should remain in the analyses, but considered only as indirect evidence as it may represent a slightly different population. The GDG felt that this was likely to be the reason for the heterogeneity in the outcome length of hospital stay. The Soriano 1991 study[208] reported that, on average, those in the placebo group had a shorter hospital stay, while the other study which reported this outcome reported the opposite. The indirectness of this study did not affect the overall quality of the critical outcomes on which the GDG based their decision. The GDG commented on the lack of availability of good quality evidence on the development of antibacterial resistance. It was also noted that microbial resistance patterns vary between countries and may have changed significantly since these trials were conducted.
Other considerations	The GDG highlighted that the prime management goal is the treatment of ascites. Antimicrobial prophylaxis is not an alternative to the effective treatment of ascites but is seen as an additional measure. As SBP is an infection of the ascitic fluid, the GDG recommended that antibiotics should be stopped once ascites has resolved. The GDG noted that existing American Liver Association guidelines suggest that antibiotics for primary prophylaxis of SBP should be considered for people at high risk of developing this complication, which was defined as an ascitic fluid protein of less than 1.5 g/dl (15 g/litre) along with impaired renal function or liver failure. The GDG discussed the limitations of the certain antibiotics. Quinolones have interactions which may need to be taken into consideration for some people with cirrhosis and ascites (for example, with concurrent sulphonylurea treatment for diabetes). The GDG were uncertain of the wide availability of norfloxacin in the UK. The GDG also noted that other antibiotics are widely used for secondary prophylaxis, including co-trimoxazole, co-amoxiclav and colistin. The popularity of these antibiotics among healthcare professionals is due to a perceived lower incidence of antibiotic resistance and development of *Clostridium difficile* diarrhoea. There was no available evidence on these drugs to make recommendations for primary prophylaxis. **Research recommendation** Spontaneous bacterial peritonitis is the most common serious infection in patients with cirrhosis, occurring in 25% of people who develop ascites. It is associated with significant morbidity and mortality rates of 20–40%. It occurs most commonly in people with advancing liver disease; approximately 70% of cases occur in people with Child-Pugh class C cirrhosis. Several oral antibiotics that have been investigated for the prophylaxis of SBP have shown benefits and a significant reduction in the incidence of SBP in people at high risk of having, or developing, cirrhosis. They are, however, associated with antibiotic resistance, adverse reactions and drug interactions. There is a lack of good quality, recent evidence regarding the prevalence and consequences of antibacterial resistance that may occur during long-term oral antibiotic therapy when used for the prevention of spontaneous bacterial peritonitis.

14 Volume replacers in hepatorenal syndrome

14.1 Introduction

Hepatorenal syndrome is a common complication in hospitalised people with decompensated cirrhosis, with an annual incidence in those with ascites of approximately 8%. It is characterised by impaired renal function.[184] The diagnostic criteria of hepatorenal syndrome include: i) cirrhosis with ascites; ii) serum creatinine >133 µmol/L; iii) no improvement of serum creatinine after volume expansion with albumin; iv) absence of shock; v) no current or recent treatment with nephrotoxic drugs; and vi) absence of parenchymal kidney disease as indicated by proteinuria >500 mg/day, microhematuria (>50 red blood cells per high power field), or abnormal renal ultrasonography. Two types of hepatorenal syndrome have been described: type 1 is characterised by rapidly progressive reduction in renal function as defined by a doubling of the initial serum creatinine to a level greater than 220 µmol/litre or a 50% reduction of the initial 24-hour creatinine clearance to a level lower that 20 ml per minute in less than 2 weeks; type 2 has a slower time course.[184] People with type 2 hepatorenal syndrome are especially predisposed to develop type 1 hepatorenal syndrome after infections, such as SBP. The prognosis of people with type 1 hepatorenal syndrome is bleak, with a mortality rate exceeding 50% after 1 month.[11] Although people with type 2 hepatorenal syndrome have a median survival of 6 months, their prognosis is markedly worse than those with cirrhosis and ascites without renal impairment.[11]

Terlipressin, a vasoconstrictor most active in the splanchnic circulation, has been used to treat hepatorenal syndrome. It is given with a plasma volume expander, which serves to maintain the blood volume and increase the blood oncotic pressure, reducing the movement of free fluid into the peritoneum.

Liver transplantation is the definitive treatment for hepatorenal syndrome but vasopressor agents are used to treat hepatorenal syndrome in conjunction with a plasma volume expander as a bridge to transplantation in those participants considered suitable for this procedure. Human albumin solution is widely used as the plasma volume expander in people with hepatorenal syndrome due to cirrhosis but it is expensive. Therefore, the GDG decided to investigate which is the most clinical and cost-effective volume replacer for people with hepatorenal syndrome who are also receiving vasoactive drugs.

14.2 Review question: Which is the most clinical and cost-effective volume replacer for patients with hepatorenal syndrome due to cirrhosis who are also receiving vasoactive drugs?

For full details see review protocol in Appendix C.

Table 92: PICO characteristics of review question

Population	• Adults and young people (16 and over) with confirmed cirrhosis and hepatorenal syndrome. Hepatorenal syndrome is defined as reversible renal dysfunction occurring in patients with cirrhosis (with a serum creatinine >133 micromol/litre and an absence of other identifiable causes of renal failure). • People who are also receiving vasoconstrictors (vasopressin, ornipressin, terlipressin, octreotide, midodrine, noradrenaline, norepinephrine, dopamine)
Interventions	• IV albumin • IV crystalloids (Ringer's lactate solution, 0.9% sodium chloride (saline), Hartmann's solution, dextrose) • IV polygel, plasma or colloid expanders (all polygel, plasma or colloid expanders

	grouped together, for example, haemocel, gelofusion/gelofusine, dextran, manitol, voluven)
Comparisons	• IV albumin versus IV crystalloids • IV albumin versus polygel, plasma or colloid expanders • IV crystalloids versus polygel, plasma or colloid expanders
Outcomes	**Critical outcomes** • Survival (time-to-event) or mortality at 3 months • Health-related quality of life (continuous) • Reversal of hepatorenal syndrome or improved renal function (dichotomous – as defined by the study) at 3 months (reduction of serum creatinine below 133 micromol/litre, creatinine clearance, renal function returning to functioning kidneys without the requirement for drugs) **Important outcomes** • Time to discharge from hospital (time-to-event) • Readmission to hospital (dichotomous) • Adverse events of volume replacement (infection) • Adverse events of volume replacement (heart failure)
Study design	Randomised controlled intervention trials

14.3 Clinical evidence

No relevant clinical studies comparing volume replacers in people with cirrhosis and hepatorenal syndrome and receiving vasoactive drugs were identified.

No studies were included in the review. See the excluded studies list in Appendix L.

14.4 Economic evidence

14.4.1 Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

14.4.2 Unit costs

See Table 98 in Appendix O.

14.5 Evidence statements

14.5.1 Clinical

- No relevant clinical studies were identified.

14.5.2 Economic

- No relevant economic evaluations were identified.

14.6 Recommendations and link to evidence

Recommendations	No recommendation was made.
Research recommendation	5. **What is the most clinically and cost-effective volume replacer for patients with hepatorenal syndrome due to cirrhosis who are also receiving vasoactive drugs?**
Relative values of different outcomes	The GDG compared albumin with other volume replacers for the management of hepatorenal syndrome in people with cirrhosis. The population considered was people receiving vasoactive drugs, as the GDG agreed that people with hepatorenal syndrome would always be given a vasopressor. The GDG also agreed that vasopressors would never be given in isolation without volume expansion, and they highlighted that there is evidence in the literature that vasopressors used without volume expansion are not efficacious.[152] Current European and American guidelines support the use of intravenous albumin in combination with vasopressors for patients with hepatorenal syndrome. The GDG looked specifically for any evidence that other volume replacers may be more clinically and cost-effective than albumin. The GDG prioritised the critical outcomes of survival, health-related quality of life and reversal of hepatorenal syndrome when considering its recommendation for this question. The GDG also considered the important outcomes of hospital length of stay, hospital readmission rates and any reported adverse events related to the use of volume expanders.
Trade-off between clinical benefits and harms	No relevant clinical studies were identified. In the absence of clinical evidence, the GDG discussed the clinical benefits and harms associated with albumin and other volume replacers. The GDG agreed that a volume expander should be given due to the low intravascular volume state of patients with cirrhosis and hepatorenal syndrome. The GDG was in agreement that there would be potential harm in not offering volume expansion in combination with vasopressors in this situation. It also noted that whilst these patients are intravascularly deplete, the pathophysiology of decompensated cirrhosis is such that they are also fluid overloaded, but that the majority of fluid is outside the vascular compartment. People with decompensated cirrhosis are therefore more prone to complications of fluid overload, such as pulmonary oedema, if given intravenous fluids. The GDG felt that the ideal volume expander in this situation should be able to provide its effect with a minimum of infused fluid (that is, have a high oncotic pressure). The GDG favoured albumin as the volume expander in treating patients with hepatorenal syndrome as it provides volume expansion with a minimum of infused fluid (especially if the 20% solution is used); is not associated with any increased risk of renal injury; and is currently the most-studied fluid in patients with decompensated cirrhosis and infections, as discussed below. Potential harms of albumin were discussed, such as evidence that giving albumin can decrease the production of endogenous albumin. The GDG noted that there were some excluded studies, which provided data on the efficacy of intravenous fluids in patients with cirrhosis outside the specific situation of hepatorenal syndrome (for example, in people with SBP or people undergoing large-volume paracentesis). The GDG agreed that studies in these populations should be excluded, as there are different underlying physiological mechanisms which result in the need to give volume replacement and differences in effectiveness of volume replacers would be expected. Therefore, it was concluded that evidence from these populations could not be used as an indirect evidence base. These studies were not systematically reviewed, but some of these studies are referred to in 'Other considerations' below regarding the use of albumin and other volume replacers in people with cirrhosis. During large-volume paracentesis, intravenous volume

	replacement is recommended and current evidence supports the use of intravenous albumin in this situation.[20] In SBP, intravenous albumin is recommended in combination with antibiotics when the serum creatinine is >1 mg/dl, blood urea nitrogen >30 mg/dl, or total bilirubin >4 mg/dl but is not necessary in patients who do not meet these criteria.[201] Albumin has been shown to be superior to hydroxyethyl starch in spontaneous bacterial peritonitis (SBP).[68] A separate consideration was the data (predominantly from critical care literature) which suggest that synthetic colloids (starch and gelatine) are associated with an increased risk of renal injury or the need for renal replacement therapy in patients with sepsis. The GDG considered that patients with decompensated cirrhosis are already at an increased risk of renal injury and so these synthetic colloids may not be appropriate.[137,163,171] Although the GDG favoured the use of albumin over other volume expanders, given the lack of clinical evidence the GDG did not feel it could make a recommendation based on these potential benefits and harms without proof that patient outcomes would be improved. Instead the GDG chose to make a recommendation for future research.
Trade-off between net clinical effects and costs	No relevant published economic evidence was identified. The GDG noted the relative expense of albumin in comparison to crystalloids, however it also noted the points raised above regarding the greater safety of albumin and the greater quantity of evidence for the use of albumin in other situations. On balance, given the lack of evidence regarding comparative clinical effectiveness or cost-effectiveness, the GDG was not able to recommend any specific volume replacer.
Quality of evidence	No relevant clinical studies were identified.
Other considerations	The GDG noted the original question of the most clinically and cost-effective volume expander in the management of hepatorenal syndrome. There was no clinical evidence identified that specifically answered this question, however the GDG did feel that this was a clinically important question to have asked, as it is an issue that is often debated in the management of such patients. The GDG discussed the current dichotomy between intensive therapy unit (ITU) and hepatology wards in the volume replacers used for this population. For this reason, the GDG felt it was important to make a recommendation for future research to investigate the most clinically and cost-effective volume expander to be used in patients with hepatorenal syndrome receiving vasoactive drugs. **Research recommendation** Hepatorenal syndrome develops in people with cirrhosis with ascites and is characterised by impaired renal function.[184] Terlipressin, a vasoconstrictor most active in the splanchnic circulation, is used to treat hepatorenal syndrome but it is given with a plasma volume expander, which serves to maintain the blood volume and increase the blood oncotic pressure, reducing the movement of free fluid into the peritoneum. Human albumin solution is the recommended intravenous volume replacement during large-volume paracentesis[20] and in patients with SBP, in combination with antibiotics, when the serum creatinine is greater than 1 mg/dl, blood urea nitrogen greater than 0 mg/dl, or total bilirubin greater than 4 mg/dl.[201] However, in hepatorenal syndrome there are no clinical studies examining the benefits and harms associated with albumin compared with other volume replacers. People with hepatorenal syndrome have a low intravascular volume state and there is general agreement that they require volume expansion in combination with vasopressors. Whilst these people have intravascular depletion, the pathophysiology of decompensated cirrhosis is such that they are also fluid overloaded, but that the

	majority of fluid is outside the vascular compartment. People with decompensated cirrhosis are therefore more prone to complications of fluid overload, such as pulmonary oedema, if given intravenous fluids. The ideal volume expander to be used in hepatorenal syndrome should be able to provide its effect with a minimum of infused fluid (that is, have a high oncotic pressure).

15 Management of an episode of acute hepatic encephalopathy

15.1 Introduction

Hepatic encephalopathy is a brain dysfunction caused by liver impairment or portosystemic shunting; it manifests as a wide spectrum of neurological or psychiatric abnormalities ranging from subtle changes in cognition to clinically obvious changes in intellect, behaviour, motor function and consciousness.[60] It is a debilitating complication of cirrhosis, severely affecting the lives of patients and their carers. It is the commonest complication of cirrhosis and the risk for the first episode of hepatic encephalopathy is 5–25% within 5 years after the diagnosis of cirrhosis and is greatest in those with decompensated liver disease.[18]

The diagnosis requires the detection of clinical signs and symptoms suggestive of hepatic encephalopathy in a person with cirrhosis or portosystemic shunting, who does not have an alternative cause of brain dysfunction. Clinically apparent or *overt* hepatic encephalopathy may arise over a period of hours or days in someone who has previously been stable. Such an episode of acute hepatic encephalopathy may be precipitated by a number of diverse events, such as infection, gastrointestinal haemorrhage, electrolyte disturbance, alcohol misuse or constipation, although in 50% of instances no obvious cause is identified.[216]

It is considered that hepatic encephalopathy is reversible with treatment. At present, therapy for an episode of acute hepatic encephalopathy is directed at the precipitating cause, as well as reducing the production and absorption of gut-derived neurotoxins, particularly ammonia, mainly through bowel cleansing, and the use of non-absorbable disaccharides, such as lactulose. However, there are no universally accepted standards for the treatment of an episode of acute hepatic encephalopathy and consequently, clinical practice is often dictated by local guidelines and personal preferences.

Since there is an unmet need for national recommendations, the GDG decided to compare the clinical and cost- effectiveness of treatments for an episode of acute hepatic encephalopathy in people with cirrhosis. The management of hepatic encephalopathy resulting from acute liver failure or portosystemic shunting, the treatment of minimal hepatic encephalopathy, the treatment of persistent hepatic encephalopathy, and prophylaxis to prevent recurrent hepatic encephalopathy are not included in this review.

For guidance on the prevention of recurrence of episodes of overt hepatic encephalopathy see Rifaximin for preventing episodes of overt hepatic encephalopathy (NICE TA337).

15.2 Review question: What is the most clinically and cost-effective intervention for the first-line treatment of an episode of acute hepatic encephalopathy in people with cirrhosis?

For full details see review protocol in Appendix C.

Table 93: PICO characteristics of review question

Population	Adults and young people (16 and over) with confirmed cirrhosis, presenting at their GP or emergency care with an episode of acute hepatic encephalopathy.
Interventions	• Non-absorbable disaccharides (combined within drug class): ○ lactulose

	○ lactitol • Oral non-absorbable antibiotics (individual drug level, not combined within class): ○ aminoglycosides (neomycin) ○ rifaximin ○ vancomycin • Other oral antibiotics (metronidazole) • Phosphate enemas (combined within drug class) • Polyethylene gycol electrolyte solution, PEG 3350 • Amino acids (IV or oral): ○ l-ornithine-l-aspartate (LOLA) ○ branch chain amino acids (combined within drug class) • IV flumazenil • Oral probiotics (combined within drug class) • Sodium benzoate • Oral zinc • Molecular Adsorbent Recirculating System (MARS) • Combination therapy (any combinations of the above) • Placebo/no treatment
Comparisons	Any head-to-head comparison (combination or mono therapy) Any intervention versus placebo/no treatment
Outcomes	**Critical outcomes:** • Survival • No improvement in hepatic encephalopathy • Health-related quality of life **Important outcomes:** • Time to discharge from hospital • Adverse events (diarrhoea, flatulence, abdominal pain, nausea, gastrointestinal bleeding, renal failure)
Study design	RCTs or systematic reviews of RCTs

15.3 Clinical evidence

We searched for randomised controlled trials comparing the effectiveness of current therapies or combinations of therapies for an episode of acute hepatic encephalopathy, either against each other or placebo. Twenty-one randomised controlled trials (23 papers) were included in the review;[5,7,33,34,71,95,97,115,126,133,156,169,176-178,199,217,218,220,236,237,240,241] these are summarised in Table 94.

The review population for this question was people with cirrhosis with an episode of acute hepatic encephalopathy. Studies including people with minimal or chronic hepatic encephalopathy were excluded from the review. The review only included treatment of the acute event and studies looking at primary or secondary prophylaxis of hepatic encephalopathy were excluded. A variety of mono and/or combined therapies were used; the evidence is presented separately for each comparison. A network meta-analysis (NMA) was not performed due to an incomplete network with no closed loops, and the variation in how the critical outcome of 'no improvement in hepatic encephalopathy' was reported across studies. Evidence from these studies is summarised in the clinical evidence summary tables below (Table 95, Table 96, Table 97, Table 98, Table 99, Table 100, Table 101, Table 102, Table 103, Table 104, Table 105, Table 106, Table 107, Table 108, Table 109, Table 110, Table 111).

See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

Table 94: Summary of studies included in the review

Study	Population	Intervention and comparison	Outcomes
Abid 2011[5] Country: Pakistan	n=120 • Cirrhosis diagnosed on the basis of clinical findings, ultrasound and/or histology. • Aged >18 years • Hepatic encephalopathy grades 1 to 4 • Patients were grouped as minimal hepatic encephalopathy if NCT-A completion took > 30 seconds and no other sign of encephalopathy. • Hyperammonaemia • With/without a single reversible precipitating factor of hepatic encephalopathy (for example constipation, hypokalaemia, urinary tract infection, respiratory tract infection, spontaneous bacterial peritonitis, dehydration)	LOLA + Lactulose + Metronidazole + any necessary concomitant medications for the treatment of precipitating factor(s) versus Lactulose + Metronidazole + any necessary concomitant medications for the treatment of precipitating factor(s) Duration of treatment: daily over 4 hours for 3 consecutive days	• Mortality • Complete improvement of hepatic encephalopathy (by 2 grades from baseline) • Partial improvement of hepatic encephalopathy (by 1 grade from baseline) • No improvement/deterioration of hepatic encephalopathy • Duration of hospitalisation • Adverse events
Ahmad 2008[7] Country: Pakistan	n=80 • Cirrhosis diagnosed on the basis of clinical, laboratory and ultrasonographic features. • Adults with a diagnosis of cirrhosis • Clinically overt encephalopathy (west Haven 1–4) developed spontaneously without any precipitating factor • Hyperammonaemia	LOLA + Lactulose + Metronidazole versus Placebo + Lactulose + Metronidazole Duration of treatment: daily over 4 hours for 5 consecutive days	• In-hospital mortality • Number of patients who achieved hepatic encephalopathy grade 0 at 5 days • Adverse reactions to medicine (nausea, vomiting) at 5 days
Cerra 1983[34]	n=22 • Cirrhosis proven by clinical evaluation of	Branch chain amino acids (BCAA)	• Mortality at follow-up • Improvement to Grade 0 hepatic

Study	Population	Intervention and comparison	Outcomes
Country: USA	biopsy studies. Patients were screened by means of history, physical examination, mental status examination, EEG and metabolic and laboratory data. • Adults aged 18–85 with chronic hepatic disease and at least acute grade 2 encephalopathy who were judged to require parenteral nutritional support.	versus Neomycin Duration of treatment: 4 to 14 days with a follow-up period of at least 7 days after the study or until death or discharge	encephalopathy at follow-up • Improvement to Grade 0–1 at follow-up
Cerra 1985[33] Country: USA	n=75 • 'For most patients that diagnosis was cirrhosis'. 65–75% of the patients in each group had this diagnosis made by biopsy, the rest by clinical criteria. • Adults aged 18 to 85 years with chronic hepatic disease and at least acute grade 2 encephalopathy. • The patients were screened by history and physical examination, electroencephalogram and by metabolic laboratory data. Encephalopathy was graded by a trained independent observer on a scale of 0–4.	BCAA versus Neomycin Duration of treatment: up to 14 days	• Mortality
Fiaccadori 1984[71] Country: Italy	n=48 • Diagnosis of cirrhosis based on clinical and laboratory data and confirmed in all cases but one by liver biopsy. • (1) Presence of cirrhosis • (2) Presence of hepatic encephalopathy • (3) No evidence of hepatorenal syndrome.	Lactulose enema versus IV BCAA versus IV BCAA + Lactulose	• Lactulose enema versus IV BCAA • Complete mental recovery (defined as consciousness regained and returned to grade 0 hepatic encephalopathy by day 7) • IV BCAA versus IV BCAA + Lactulose • Coming out of coma by day 7

Study	Population	Intervention and comparison	Outcomes
		Duration of treatment: 7 days	
Gyr 1996[95] Country: multiple	n=49 • Hospitalised patients having chronic liver failure with mild to moderate degree of portal systemic encephalopathy (PSE, stage I–III or clinical PSE score 3–14). • PSE episodes resulting from common precipitating situations such as severe bleeding and infection were excluded, resulting in a selection of patients with apparently more spontaneous and stable PSE in chronic liver disease. • (No specific information on cirrhosis diagnosis provided)	Flumazenil versus Placebo Duration of treatment: 3 hour treatment period + 5 hour post-treatment observation period	• Mortality during study period • Mortality at 4 weeks follow-up • Improvement of at least 2 points in PSE score from baseline • Adverse events at 3 hour treatment period + 5 hour post-treatment observation period
Hassanein 2007[97] Country: multiple	n=70 • Cirrhosis was determined by medical history, and confirmed clinically, biochemically and radiologically. • Patients 18 years of age or older, presenting with manifestations of cirrhosis and hepatic encephalopathy grade 3 or 4.	MARS + Standard medical therapy (SMT) versus SMT Duration of treatment: 6 hours daily for 5 days or until a 2 grade improvement in hepatic encephalopathy; follow-up until 180 days post-treatment	• Mortality at 5 days • People with an improvement of hepatic encephalopathy by 2 grades at any time during the 5 day study period • Serious adverse events at 5 days
Lacetti 2000[115] Country: Italy	n=54 • The diagnosis of cirrhosis was made by pertinent clinical, laboratory and morphological procedures performed during previous hospitalisation.	Flumazenil versus Placebo	• Mortality at 24 hours • Improvement in neurological status (increase in Glasgow coma score by 3 points) at 24 hours

Study	Population	Intervention and comparison	Outcomes
	• People with a diagnosis of liver cirrhosis who presented with hepatic encephalopathy in the emergency department or developed hepatic encephalopathy during their hospital stay: of those, only individuals with chronic liver failure and more severe stages of hepatic encephalopathy (stages III–IV) were included.	Conventional treatment similar for both groups (the following additional treatments were permitted: saline, glucose, lactulose enemas, BCAAs) Duration of study: 10 days treatment and a further 10 days follow-up	• Side effects at 24 hours
Loguercio 1987[126] Country: Italy	n=40 • Diagnosis of cirrhosis: Conn and Lieberthal method • 'Cirrhotic patients'	Oral probiotics versus Oral lactulose Duration of treatment: 10 days treatment and a further 10 days follow-up	• Improvement in hepatic encephalopathy symptoms at day 10 • Meteoism, abdominal pain, diarrhoea, hyperammonaemia, worsening of hepatic encephalopathy, constipation at 20 days
Mas 2003[133] Country: Spain	n=103 • After hospital admission, patients underwent detailed physical, neurological and psychometric assessment. • Consecutive cirrhotic patients with an acute hepatic encephalopathy episode, diagnosed in specified 13 hospitals in Spain from November 1995 to December 1997: with clinical, psychometric and electroencephalographic evidence of grade I–III hepatic encephalopathy of <2 days duration and PSE index >0	Rifaximin versus Lactitol Duration of treatment: every 8 hours for maximum of 10 days	• Mortality considered unrelated to the study medication within 28 days of the last dose • Unchanged/failure (hepatic encephalopathy clinical syndrome not improved and blood ammonia levels not decreased/increase in blood ammonia, increase in PSE index and/or shift to a higher stage of hepatic encephalopathy) • Adverse events post-treatment
Paik 2005[156]	n=54 • Diagnosis of cirrhosis based on clinical and	Rifaximin	• Improvement in hepatic encephalopathy grade at 7 days

Study	Population	Intervention and comparison	Outcomes
Country: South Korea	laboratory findings. • Hospital inpatients with episodic hepatic encephalopathy affected by decompensated cirrhosis. • The participants showed signs of grade I to III hepatic encephalopathy, according to Conn's modification of Parsons-Smith classification, and had serum ammonia levels > 75 µmol/litre. Of the 64 participants, 26 (40.6%) had 'acute hepatic encephalopathy' and 38 (59.4%) had "'recurrent hepatic encephalopathy'.	versus Oral lactulose Duration of treatment: antibiotics 3x daily for 7 days; lactulose 1x daily for 7 days	• Improvement in hepatic encephalopathy index (taking into account hepatic encephalopathy grade, number connection test (NCT), blood ammonia and severity of flapping tremor) at 7 days • Adverse effects at 7 days
Rahimi 2014[169] Country: USA	n=50 • Cirrhosis was defined by clinical features, including a history consistent with chronic liver disease (CLD) as well as documented complication of CLD and/or imaging results consistent with cirrhosis and/or liver histology consistent with cirrhosis. • (1) Age 18 to 80 years • (2) Diagnosis of cirrhosis from any cause • (3) Presence of any grade of hepatic encephalopathy • (4) Availability of a legally authorised representative (LAR) for interview and consent.	Polyethylene glycol electrolyte solution, PEG 3350 (after 24 hours participants were allowed to receive lactulose per the standard care) versus Oral lactulose Duration of treatment: 4 hours; until discharge from hospital or death	• Mortality • Improvement of 1 or more grade in hepatic encephalopathy at 24 hours • Improvement to grade 0, or 2 days at grade 1 after initial improvement of at least 1 grade • No improvement of hepatic encephalopathy grade at 24 hours • Overall length of stay • Number of adverse events at 24 hours
Rossi-Fanelli 1982[178]; 1984[177]; 1986[176] Country: Italy	n=34 • (1) Presence of cirrhosis, diagnosed on clinical, biochemical and histological findings • (2) Presence of hepatic coma (grade 3–4	BCAA – BS692 versus Oral lactulose	• Mortality up to 10 days after mental state recovery • Mean time of arousal • Number of patients achieving complete mental recovery

Study	Population	Intervention and comparison	Outcomes
	hepatic encephalopathy) assessed by 2 independent observers according to the classification of Adams & Foley as reported by Fischer et al. • (3) Absence of signs of hepatorenal syndrome assessed according to the criteria established at the symposium held in Sassari.	Duration of treatment: until 48 hours after mental state recovery; following this patients who did not recover underwent a combination treatment in both treatment arms; until 10 days after the start of therapy.	(consciousness regained and returned to grade 0 hepatic encephalopathy) • Unresponsive: number of patients not achieving complete mental recovery
Sharma 2013[199] Country: India	n=120 • Diagnosis of cirrhosis was based on laboratory tests, endoscopic evidence, sonographic findings, and liver histology if available. • Patients at a tertiary care centre aged 18 to 80 years with cirrhosis and overt hepatic encephalopathy.	Rifaximin + lactulose versus Oral lactulose Duration of treatment: 3x daily until complete recovery of hepatic encephalopathy or a maximum of 10 days if no recovery, discharge from hospital or death	• Mortality • Number of participants achieving complete reversal of hepatic encephalopathy (according to West Haven criteria) within 10 days • Length of hospital stay • Side effects related to study medications
Strauss 1986[217] Country: Brazil	n=29 • Diagnosis made 'mainly on a histological basis' • Diagnosed cirrhosis. Hepatic encephalopathy characterised as a disturbance of consciousness assessed semi-quantitatively as grades I to IV. • Patients were treated equally for exogenous precipitating factors of the encephalopathy. Diuretics were always withdrawn and gastrointestinal bleeding due to oesophageal varices was treated with Sengstaken-Blakemore balloon and	IV BCAA versus Neomycin Duration of treatment: not directly specified (until complete recovery of consciousness)	• Mortality during treatment • Time to recovery during treatment

Study	Population	Intervention and comparison	Outcomes
	blood transfusion. Potassium was supplemented if necessary and laxatives were used only in constipated patients. Infections were treated with antibiotics, mainly ampicillin (1–4 g orally) or according to specific antibiograms.		
Strauss 1992[218] Country: Brazil	n=39 • Histopathological and/or clinical-biochemical diagnosis of cirrhosis • 8 of the 39 patients randomised had previous episodes of hepatic encephalopathy (but people with chronic hepatic encephalopathy or on specific treatment for hepatic encephalopathy at the time of randomisation or in the week before it were excluded)	Neomycin versus Placebo In both cases: patients in grades III and IV also received enriched solution of BCAAs (Portamin) with IV hypertonic glucose Duration of treatment: unclear. Patients were followed up and analysed for mortality for 1 year after discharge	• Therapeutic failure and death at 5th day of treatment • Time until regression to grade 0 hepatic encephalopathy
Sushma 1992[220] Country: India	n=74 • Diagnosis of cirrhosis was made by liver biopsy or clinical criteria when liver biopsy was not possible. • People with diagnosis of cirrhosis or had had a surgical portal-systemic anastomosis; hepatic encephalopathy of <7 days	Sodium benzonate versus Oral lactulose Duration of treatment: until recovery or death	• Mortality during treatment • Mean duration of therapy before complete recovery • Number of patients with complete response (recovery to normal mental status with no evidence of asterixis) • Number of participants who continued in grade 1+ mental status despite therapy for 21 days • Number of complications during treatment

Study	Population	Intervention and comparison	Outcomes
Uribe 1981[236] Country: Mexico	n=18 • Biopsy-proven cirrhosis • Development within 24 hours of an acute episode of hepatic encephalopathy (at least grade 2+ severity) plus 2 of the following abnormalities: arterial ammonia levels above 120 ug% (normal <90 ug%); abnormal slow waves in the EEG as blindly judged by a neurologist; time taken to perform a NCT at least double the normal range (>60 seconds, normal is >30 seconds) or patient unable to perform the test due to mental confusion or coma.	Lactose Enema versus Neomycin Duration of treatment: treatment continued until 48 hours after recovery then study was concluded	• Mortality within 1 month from end of study • Clinical-biochemical improvement (improvement of 1 grade in mental state – Conn's grading 0–4), a reduction of 30 seconds in time taken to perform the number connection test (NCT) and ammonia reduction of 50 ug%) • Treatment side effects
Uribe 1987[237]	• n=15 (trial also continued with a lactitol versus lactose comparison, n=45) • Cirrhosis (unclear how cirrhosis diagnosed) • Development within 24 hours of an acute episode of hepatic encephalopathy (at least grade 2+ severity according to Conn's classification) plus 2 of the following abnormalities: arterial ammonia levels above 120 ug% (normal <90 ug%); abnormal slow waves in the EEG and protracted performance on a number connection test of at least double the normal time (normal <30 seconds) or inability to perform the test due to mental confusion or coma.	20% lactitol enema versus placebo tap water enema Duration of treatment: variable and response dependent	• Death • Therapeutic response (defined as (i) sustained improvement of 1 grade in mental state during ≤48 hours or (ii) improvement of more than 2 grades in mental state).
Vilstrup 1990[240] Country: Denmark	n=77 • Cirrhosis and hepatic encephalopathy Grade II/III/IV, according to the Fogarty classification.	IV BCAA versus	• Mortality at 16 days • Number of participants who woke up (to hepatic encephalopathy grade 0 or I by Fogarty classification at 16 days)

Study	Population	Intervention and comparison	Outcomes
	• Insufficient information provided.	Placebo/glucose Duration of treatment: until recovery or death (maximum of 16 days of treatment)	• Number of participants who had treatment failures other than death (hepatic encephalopathy deeper than grade I [Fogarty classification]) after 16 days
Wahren 1983[241] Country: Sweden	n=50 • EEG and neurological examinations • Clinical and laboratory evidence of cirrhosis verified histologically by liver biopsy, autopsy, angiography, laparoscopy, laparotomy	IV BCAA versus Placebo/glucose Duration of treatment: given until 1 day after hepatic encephalopathy had improved to grade 0 or 1, for a maximum of 5 days. Last blood collected the morning after the end of the intervention.	• Mortality at 5 days • Positive response to treatment at 5 days • No response to treatment at 5 days • Negative response to treatment at 5 days

15.3.1 Non-absorbable disaccharides versus single therapy

Table 95: Clinical evidence summary: non-absorbable disaccharides versus neomycin

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with neomycin	Risk difference with non-absorb disaccharides (95% CI)
Mortality	18 (1 study)	VERY LOW[a, b] due to risk of bias, imprecision	RR 1.25 (0.09 to 17.02)	Moderate	
				100 per 1000	25 more per 1000 (from 91 fewer to 1000 more)
Clinical-biochemical improvement (improvement of 1 grade in mental	18	LOW[a, b]	RR 1.25	Moderate	

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with neomycin	Risk difference with non-absorb disaccharides (95% CI)
state [Conn's grading 0–4], a reduction of 30 seconds in time taken to perform the NCT and ammonia reduction of 50 ug%)	(1 study)	due to risk of bias, imprecision	(0.77 to 2.03)	700 per 1000	175 more per 1000 (from 161 fewer to 721 more)
Side effects	18 (1 study)	MODERATE[a] due to risk of bias	Not estimable	0 per 1000	No events in either the intervention or control arm

[a] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[b] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 96: Clinical evidence summary: non-absorbable disaccharides versus Rifaximin

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Rifaximin	Risk difference with non-absorbable disaccharides (95% CI)
Mortality (considered unrelated to medication; at 28 days)	103 (1 study)	LOW[a] due to imprecision	RR 1.89 (0.18 to 20.17)	Moderate	
				20 per 1000	18 more per 1000 (from 16 fewer to 383 more)
Unchanged/failure (hepatic encephalopathy clinical syndrome not improved and blood ammonia levels not decreased/increase in blood ammonia, increase in PSE index and/or a shift to a higher stage of hepatic encephalopathy)	103 (1 study)	LOW[a] due to imprecision	RR 1.05 (0.46 to 2.36)	Moderate	
				180 per 1000	9 more per 1000 (from 97 fewer to 245 more)
Improvement in hepatic encephalopathy grade (at 7 days)	54 (1 study)	LOW[a, b] due to risk of bias,	RR 0.9 (0.66 to 1.21)	Moderate	
				813 per 1000	81 fewer per 1000

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Rifaximin	Risk difference with non-absorbable disaccharides (95% CI)
		imprecision			(from 276 fewer to 171 more)
Improvement in hepatic encephalopathy index (taking into account hepatic encephalopathy grade, NCT, blood ammonia and severity of flapping tremor; at 7 days)	54 (1 study)	LOW[a, b] due to risk of bias, imprecision	RR 1.13 (0.95 to 1.35)	Moderate	
				844 per 1000	110 more per 1000 (from 42 fewer to 295 more)
Adverse events	157 (2 studies)	VERY LOW[a, c] due to inconsistency, imprecision	RR 0.8 (0.19 to 3.39)	Moderate	
				46 per 1000	9 fewer per 1000 (from 37 fewer to 110 more)

[a] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
[b] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[c] The point estimate varies widely across studies, unexplained by subgroup analysis

Table 97: Clinical evidence summary: non-absorbable disaccharides versus BCAA

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with BCAA	Risk difference with non-absorbable disaccharides (95% CI)
Mortality (up to 10 days after mental recovery)	34 (1 study)	LOW[a] due to imprecision	RR 1.25 (0.4 to 3.87)	Moderate	
				235 per 1000	59 more per 1000 (from 141 fewer to 674 more)
Time of arousal (hours)	34 (1 study)	MODERATE[a] due to imprecision		27.6	The mean time of arousal in the intervention groups was 3.9 hours more

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with BCAA	Risk difference with non-absorbable disaccharides (95% CI)
					(11.43 lower to 19.23 higher)
Complete mental recovery (study 1 defines as consciousness regained and returned to grade 0 hepatic encephalopathy; study 2 defines as come out of coma by day 7)	66 (2 studies)	MODERATE[a] due to imprecision	RR 0.67 (0.47 to 0.94)	Moderate	
				822 per 1000	271 fewer per 1000 (from 49 fewer to 436 fewer)

[a] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 98: Clinical evidence summary: non-absorbable disaccharides versus PEG 3350

Outcomes	Number of Participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with PEG 3350	Risk difference with non-absorbable disaccharides (95% CI)
Mortality (at 24 hours)	50 (1 study)	LOW[a] due to imprecision	RR 2 (0.19 to 20.67)	Moderate	
				40 per 1000	40 more per 1000 (from 32 fewer to 787 more)
Hepatic encephalopathy resolution (defined as an improvement to grade 0, or 2 days at grade 1 after an initial improvement of at least 1 grade)	48 (1 study)	LOW[a, b] due to risk of bias, imprecision	HR 0.57 (0.31 to 1.05)	Moderate	
				_[c]	_
Improvement of 1 or more in hepatic encephalopathy grade (hepatic encephalopathy spectral analysis (SA) score; at 24 hours)	48 (1 study)	LOW[a, b] due to risk of bias, imprecision	RR 0.57 (0.38 to 0.85)	Moderate	
				913 per 1000	393 fewer per 1000 (from 137 fewer to 566 fewer)
Length of hospital stay (days)	50 (1 study)	LOW[a] due to		4 days	The mean length of hospital stay (days) in the intervention groups

Outcomes	Number of Participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with PEG 3350	Risk difference with non-absorbable disaccharides (95% CI)
		imprecision			was 4 higher (0.85 lower to 8.85 higher)
Adverse events (at 24 hours)	50 (1 study)	LOW[a] due to imprecision	RR 1.67 (0.45 to 6.24)	Moderate	
				120 per 1000	80 more per 1000 (from 66 fewer to 629 more)

[a] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
[b] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[c] Not possible to calculate control risk

Table 99: Clinical evidence summary: non-absorbable disaccharides versus probiotics

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with probiotics	Risk difference with non-absorbable disaccharides (95% CI)
Improvement in hepatic encephalopathy symptoms (at day 10)	38 (1 study)	VERY LOW[a, b] due to risk of bias, imprecision	RR 0.93 (0.65 to 1.33)	Moderate	
				790 per 1000	55 fewer per 1000 (from 277 fewer to 261 more)
Adverse events (at 20 days)	31 (1 study)	VERY LOW[a, b] due to risk of bias, imprecision	RR 8.53 (1.21 to 60.33)	Moderate	
				63 per 1000	474 more per 1000 (from 13 more to 1000 more)

[a] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with probiotics	Risk difference with non-absorbable disaccharides (95% CI)

[b] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 100: Clinical evidence summary: non-absorbable disaccharides versus sodium benzoate

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with sodium benzoate	Risk difference with non-absorbable disaccharides (95% CI)
Mortality	74 (1 study)	LOW[a] due to imprecision	RR 0.92 (0.37 to 2.29)	Moderate	
				211 per 1000	17 fewer per 1000 (from 133 fewer to 272 more)
Complete response (recovery to normal mental status with no evidence of asterixis)	74 (1 study)	MODERATE[a] due to imprecision	RR 1.02 (0.81 to 1.28)	Moderate	
				790 per 1000	16 more per 1000 (from 150 fewer to 221 more)
Continued in grade 1+ mental status despite therapy for 21 days	74 (1 study)	LOW[a] due to imprecision	RR 0.35 (0.04 to 3.23)	Moderate	
				79 per 1000	51 fewer per 1000 (from 76 fewer to 176 more)
Complications during treatment	74 (1 study)	HIGH	RR 0.9 (0.76 to 1.08)	Moderate	
				921 per 1000	92 fewer per 1000 (from 221 fewer to 74 more)

[a] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

15.3.2 Combination therapy (1 intervention + non-absorbable disaccharides) versus non-absorbable disaccharides

Table 101: Clinical evidence summary: Rifaximin + non-absorbable disaccharides versus non-absorbable disaccharides

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-absorbable disaccharides	Risk difference with Rifaximin+non-absorbable disaccharides (95% CI)
Mortality	120 (1 study)	MODERATE[a] due to imprecision	RR 0.48 (0.29 to 0.81)	Moderate	
				491 per 1000	255 fewer per 1000 (from 93 fewer to 349 fewer)
Complete reversal of hepatic encephalopathy (according to West Haven criteria; at 10 days)	120 (1 study)	MODERATE[a] due to imprecision	RR 1.5 (1.12 to 2)	Moderate	
				509 per 1000	255 more per 1000 (from 61 more to 509 more)
Length of hospital stay (days)	120 (1 study)	MODERATE[a] due to imprecision		8.2	The mean length of hospital stay in the intervention groups was 2.4 days shorter (3.86 to 0.94 lower)
Side effects related to study medications	120 (1 study)	LOW[a] due to imprecision	RR 1.09 (0.51 to 2.32)	Moderate	
				175 per 1000	16 more per 1000 (from 86 fewer to 231 more)

[a] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 102: Clinical evidence summary: BCAA + non-absorbable disaccharides versus non-absorbable disaccharides

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-absorbable disaccharides	Risk difference with BCAA+non-absorbable disaccharides (95% CI)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-absorbable disaccharides	Risk difference with BCAA+non-absorbable disaccharides (95% CI)
Mortality (at 16 days)	65 (1 study)	LOW[a] due to imprecision	RR 1.13 (0.56 to 2.3)	Moderate	
				303 per 1000	39 more per 1000 (from 133 fewer to 394 more)
Wake up (study 1 defines as woke up to hepatic encephalopathy grade 0 or I by Fogarty classification at 16 days; study 2 defines as came out of coma by day 7)	97 (2 studies)	VERY LOW[a, b, c] due to risk of bias, inconsistency, imprecision	RR 1.24 (0.91 to 1.69)	Moderate	
				570 per 1000	137 more per 1000 (from 51 fewer to 393 more)
Treatment failures other than death (hepatic encephalopathy deeper than grade I [Fogarty classification] despite other improvements; at 16 days)	65 (1 study)	LOW[a] due to imprecision	RR 0.69 (0.21 to 2.21)	Moderate	
				182 per 1000	56 fewer per 1000 (from 144 fewer to 220 more)

[a] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
[b] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[c] Downgraded by 1 or 2 increments because the point estimate varies widely across studies, unexplained by subgroup analysis

Table 103: Clinical evidence summary: Flumazenil + non-absorbable disaccharides versus non-absorbable disaccharides

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo (concurrent lactulose)	Risk difference with Flumazenil (95% CI)
Mortality (during the observation period, 3 hour treatment + 5 hour observation)	49 (1 study)	VERY LOW[a, b] due to risk of bias, imprecision	OR 0.1 (0 to 5.09)	Moderate	
				48 per 1000	43 fewer per 1000 (from 48 fewer to 156 more)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo (concurrent lactulose)	Risk difference with Flumazenil (95% CI)
Clinically relevant response (improvement of at least 2 points in PSE score, PSE score on a 0–16 scale, at 8 hours)	49 (1 study)	LOW[a] due to risk of bias	Peto OR 7.39 (1.49 to 36.61)	Moderate	
				0 per 1000	250 more per 1000 (from 80 more to 420 more)
Adverse events (at 8 hours – flushing, nausea and vomiting, irritability)	49 (1 study)	VERY LOW[a, b] due to risk of bias, imprecision	Peto OR 6.47 (0.84 to 49.99)	Moderate	
				0 per 1000	140 more per 1000 (from 0 more to 290 more)

[a] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[b] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

15.3.3 Combination therapy (2 interventions + non-absorbable disaccharides) versus combination therapy (1 intervention + non-absorbable disaccharides)

Table 104: Clinical evidence summary: Flumazenil + BCAA + non-absorbable disaccharides versus BCAA + non-absorbable disaccharides

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo (concurrent lactulose and BCAA)	Risk difference with Flumazenil (95% CI)
Mortality (at 24 hours)	54 (1 study)	VERY LOW[a, b] due to risk of bias, imprecision	RR 1.11 (0.39 to 3.22)	Moderate	
				192 per 1000	21 more per 1000 (from 117 fewer to 426 more)
Improvement in neurological status	54	VERY LOW[a, b]	RR 1.46	Moderate	

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo (concurrent lactulose and BCAA)	Risk difference with Flumazenil (95% CI)
(increase in Glasgow coma score by 3 points); at 24 hours	(1 study)	due to risk of bias, imprecision	(0.97 to 2.19)	539 per 1000	248 more per 1000 (from 16 fewer to 641 more)
Side effects (at 24 hours)	54 (1 study)	LOW[a] due to risk of bias	Not estimable	0 per 1000	No events in either the intervention or control arm

[a] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[b] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 105: Clinical evidence summary: LOLA + metronidazole + non-absorbable disaccharides versus metronidazole + non-absorbable disaccharides

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo (lactulose+metronidazole)	Risk difference with LOLA (lactulose+metronidazole) (95% CI)
Mortality (inpatient stay)	200 (2 studies)	VERY LOW[a, b] due to indirectness, imprecision	RR 0.55 (0.21 to 1.42)	Moderate	
				108 per 1000	49 fewer per 1000 (from 85 fewer to 45 more)
Complete improvement defined as improvement of 2 grades from baseline (day 3)	108 (1 study)	HIGH	RR 1.8 (1.32 to 2.46)	Moderate	
				463 per 1000	370 more per 1000 (from 148 more to 676 more)
Achieved hepatic encephalopathy grade 0 (at 5 days)	80 (1 study)	VERY LOW[b, c] due to risk of bias, imprecision	RR 1.19 (0.99 to 1.44)	Moderate	
				775 per 1000	147 more per 1000 (from 8 fewer to 341 more)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo (lactulose+metronidazole)	Risk difference with LOLA (lactulose+metronidazole) (95% CI)
Adverse events	200 (2 studies)	VERY LOW[b, c] due to risk of bias, imprecision	Peto OR 7.39 (0.15 to 372.38)	Moderate	
				0 per 1000	10 more per 1000 (from 20 fewer to 40 more)

[a] The majority of the evidence had indirect outcomes
[b] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
[c] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

Narrative findings

LOLA + lactulose + metronidazole versus placebo + lactulose + metronidazole

Abid 2011[5] report that the median duration of hospitalisation following treatment with LOLA was 96 hours (range 48–574) compared to placebo which was 96 hours ([range 90–240] p=0.025).

15.3.4 Single therapy versus placebo

Table 106: Clinical evidence summary: non-absorbable disaccharides versus placebo

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo	Risk difference with non-absorbable disaccharides (95% CI)
Mortality	15 (1 study)	LOW[a] due to risk of bias	OR 0.03 (0 to 0.4)	Moderate	
				600 per 1000	557 fewer per 1000

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo	Risk difference with non-absorbable disaccharides (95% CI)
					(from 225 fewer to 600 fewer)
Therapeutic response defined as (i) sustained improvement of 1 grade in mental state during ≤48 hours or (ii) improvement of more than 2 grades in mental state	15 (1 study)	VERY LOW[a,b] due to risk of bias, imprecision	RR 3.82 (0.95 to 15.36)	Moderate	
				200 per 1000	564 more per 1000 (from 10 fewer to 1000 more)

[a] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[b] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 107: Clinical evidence summary: BCAA versus placebo

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo	Risk difference with BCAA (95% CI)
Mortality (at 5 days)	50 (1 study)	LOW[a, b] due to risk of bias, imprecision	RR 2 (0.8 to 5.02)	Moderate	
				200 per 1000	200 more per 1000 (from 40 fewer to 804 more)
Positive response to treatment (at 5 days)	42 (1 study)	VERY LOW[a, b] due to risk of bias, imprecision	RR 1 (0.55 to 1.83)	Moderate	
				500 per 1000	0 fewer per 1000 (from 225 fewer to 415 more)

[a] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[b] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 108: Clinical evidence summary: Neomycin (+BCAA in grades III and IV) versus placebo (+BCAA in grades III and IV)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with placebo (concurrent BCAA in grade III/IV)	Risk difference with neomycin (95% CI)
Mortality (at day 5)	39 (1 study)	VERY LOW[a, b] due to risk of bias, imprecision	RR 0.95 (0.15 to 6.08)	Moderate	
				105 per 1000	5 fewer per 1000 (from 89 fewer to 533 more)
Time until regression to grade 0 hepatic encephalopathy (hours)	39 (1 study)	LOW[a, b] due to risk of bias, imprecision		49.47	The mean time until regression to grade 0 hepatic encephalopathy in the intervention groups was 13.36 hours shorter (27.47 lower to 0.75 higher)

[a] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[b] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

15.3.5 Single therapy versus single therapy

Table 109: Clinical evidence summary: BCAA versus neomycin

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with neomycin	Risk difference with BCAA (95% CI)
Mortality	129 (3 studies)	VERY LOW[a, b, c] due to risk of bias, inconsistency, imprecision	RR 0.57 (0.36 to 0.89)	Moderate	
				400 per 1000	172 fewer per 1000 (from 44 fewer to 256 fewer)

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with neomycin	Risk difference with BCAA (95% CI)
Full improvement to grade 0 hepatic encephalopathy	17 (1 study)	VERY LOW[a, c] due to risk of bias, imprecision	RR 2.22 (0.58 to 8.44)	Moderate	
				250 per 1000	305 more per 1000 (from 105 fewer to 1000 more)
Improvement to grade 0 or 1 hepatic encephalopathy	17 (1 study)	LOW[a, c] due to risk of bias, imprecision	RR 1.19 (0.75 to 1.88)	Moderate	
				750 per 1000	143 more per 1000 (from 188 fewer to 660 more)
Time to recovery (hours)	28 (1 study)	MODERATE[a] due to risk of bias		70.8	The mean time to recovery (hours) in the intervention groups was 37.4 lower (56.1 to 18.7 lower)

[a] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[b] Downgraded by 1 or 2 increments because the point estimate varies widely across studies, unexplained by subgroup analysis
[c] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

15.3.6 Combination therapy (one intervention + non-absorbable disaccharides) versus single therapy

Table 110: Clinical evidence summary: BCAA + non-absorbable disaccharides versus BCAA

Outcomes	Number of participants (studies)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with BCAA	Risk difference with BCAA+non-absorbable disaccharides (95% CI)
Came out of coma (at 7 days)	32 (1 study)	VERY LOW[a, b] due to risk of bias, imprecision	RR 1.06 (0.9 to 1.26)	Moderate	
				938 per 1000	56 more per 1000 (from 94 fewer to 244 more)

Outcomes	Number of participants (studies)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with BCAA	Risk difference with BCAA+non-absorbable disaccharides (95% CI)

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

15.3.7 MARS versus standard medical therapy

Table 111: Clinical evidence summary: MARS versus standard medical therapy

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with standard medical therapy	Risk difference with MARS (95% CI)
Mortality (at 5 days)	70 (1 study)	LOW[a] due to imprecision	RR 0.79 (0.25 to 2.5)	Moderate	
				161 per 1000	34 fewer per 1000 (from 121 fewer to 241 more)
Responder (improvement of hepatic encephalopathy by 2 grades at any time; at 5 days)	69 (1 study)	LOW[a, b] due to risk of bias, imprecision	RR 1.54 (0.93 to 2.55)	Moderate	
				400 per 1000	216 more per 1000 (from 28 fewer to 620 more)
Serious adverse events (at 5 days)	70 (1 study)	MODERATE[a] due to imprecision	RR 1.99 (1.02 to 3.89)	Moderate	
				258 per 1000	255 more per 1000 (from 5 more to 746 more)

[a] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
[b] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

15.4 Economic evidence

15.4.1 Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

15.4.2 Unit costs

See Table 99 in Appendix O.

15.5 Evidence statements

15.5.1 Clinical

Single therapy versus placebo

- A clinical benefit was found for non-absorbable disaccharides over placebo for mortality (1 study, 15 patients, Low quality) and therapeutic response (1 study, 15 patients, Very Low quality).
- A clinical harm was found for BCAA over placebo for mortality (1 study, 50 patients, Low quality). No difference was found for positive response to treatment (1 study, 42 patients, Very Low quality).
- Neomycin (+ BCAA in grades III and IV) showed clinical benefit over placebo (+ BCAA in grades III and IV) for the outcome time until regression to grade 0 hepatic encephalopathy (1 study, 39 patients, Low quality), but no clinical difference for mortality (1 study, 39 patients, Very Low quality).

Single therapy versus single therapy

- BCAA showed clinical benefit over neomycin for the outcomes mortality (3 studies, 129 patients, Very Low quality), full improvement to grade 0 hepatic encephalopathy (1 study, 17 patients, Very Low quality), improvement to grade 0 or 1 hepatic encephalopathy (1 study, 17 patients, Low quality) and time to recovery (1 study, 28 patients, Moderate quality).

Non-absorbable disaccharides versus single therapy

- Non-absorbable disaccharides showed clinical harm over neomycin for the outcomes mortality and clinical-biochemical improvement (1 study, 18 patients, Low to Very Low quality). No difference was found for side effects (1 study, 18 patients, Moderate quality).
- Non-absorbable disaccharides showed clinical harm over rifaximin for mortality (1 study, 103 patients, Low quality). No difference was found for the outcomes unchanged/failure (1 study, 103 patients, Low quality), improvement in hepatic encephalopathy grade (1 study, 54 patients, Low quality), improvement in hepatic encephalopathy index (1 study, 54 patients, Low quality) and adverse events (2 studies, 157 patients, Very Low quality).
- Non-absorbable disaccharides showed clinical harm over BCAA for mortality (1 study, 34 patients, Low quality), and complete mental recovery (2 studies, 66 patients, Moderate quality). There was no clinically important difference for the outcome time of arousal (1 study, 34 patients, Moderate quality).
- Non-absorbable disaccharides showed clinical harm over PEG 3350 for the outcomes mortality (1 study, 50 patients, Low quality), hepatic encephalopathy resolution (1 study, 48 patients, Low quality), improvement of 1 or more in hepatic encephalopathy grade (1 study, 48 patients, Low

quality), length of hospital stay (1 study, 50 patients, Low quality), and adverse events (1 study, 50 patients, Low quality).

- Non-absorbable disaccharides showed clinical harm over probiotics for adverse events (1 study, 31 patients, Very Low quality). No difference was found for improvement in hepatic encephalopathy symptoms (1 study, 38 patients, Very Low quality).
- No clinical difference was found for non-absorbable disaccharides compared to sodium benzoate for the outcomes mortality (1 study, 74 patients, Low quality), complete response (1 study, 74 patients, Moderate quality), and complications during treatment (1 study, 74 patients, High quality).

Combination therapy (one intervention + non-absorbable disaccharides) versus non-absorbable disaccharides

- A clinical benefit was found for rifaximin and non-absorbable disaccharides over non-absorbable disaccharides alone for the outcomes mortality, complete reversal of hepatic encephalopathy and length of hospital stay (1 study, 120 patients, Moderate quality). No difference was found for the outcome side effects related to study medication (1 study, 120 patients, Low quality).
- A clinical benefit was found for BCAA and non-absorbable disaccharides over non-absorbable disaccharides alone for the outcomes 'wake up' (2 studies, 97 patients, Very Low quality) and treatment failures other than death (1 study, 65 patients, Low quality). For the outcome mortality no difference was found (1 study, 65 patients, Low quality).
- A clinical benefit was found for flumazenil and non-absorbable disaccharides over non-absorbable disaccharides alone for the outcomes mortality (1 study, 49 patients, Very Low quality) and clinically relevant response (1 study, 49 patients, Low quality). Clinical harm for the combination with flumazenil was found for adverse events (1 study, 49 patients, Very Low quality).

Combination therapy (2 interventions + non-absorbable disaccharides) versus combination therapy (one intervention + non-absorbable disaccharides)

- A clinical benefit was found for flumazenil combined with BCAA and non-absorbable disaccharides over BCAA and non-absorbable disaccharides for the outcome improvement in neurological status (1 study, 54 patients, Very Low quality). No difference was found for the outcomes mortality (1 study, 54 patients, Very Low quality) and side effects (1 study, 54 patients, Low quality).
- A clinical benefit was found for LOLA combined with metronidazole and non-absorbable disaccharides over metronidazole and non-absorbable disaccharides for the outcome mortality (2 studies, 200 patients, Very Low quality), improvement of 2 grades from baseline (1 study, 108 patients, High quality) and achievement of hepatic encephalopathy grade 0 (1 study, 80 patients, Very Low quality). A clinical harm was found for the treatment combination involving LOLA for the outcome adverse events (2 studies, 200 patients, Very Low quality).

Combination therapy (1 intervention + non-absorbable disaccharides) versus single therapy

- A clinical benefit of BCAA combined with non-absorbable disaccharides over BCAA alone was found for the outcome 'came out of a coma' (1 study, 32 patients, Very Low quality).

MARS versus standard medical therapy

- A clinical benefit was found for MARS over standard medical therapy for mortality (1 study, 70 patients, Low quality) and responder (1 study, 69 patients, Low quality). A clinical harm was found for MARS for the outcome serious adverse events (1 study, 70 patients, Moderate quality).

15.5.2 Economic

- No relevant economic evaluations were identified.

15.6 Recommendations and link to evidence

Recommendations	**No recommendation.**
Research recommendation	**6. In people with cirrhosis and an acute episode of hepatic encephalopathy secondary to a clearly identified, potentially reversible precipitating factor, does management of the precipitating event alone improve the hepatic encephalopathy without specific treatment?**
Relative values of different outcomes	The GDG selected the outcomes of mortality, improvement in hepatic encephalopathy and health-related quality of life as critical outcomes, and time to discharge from hospital and treatment-related adverse events as important outcomes for this question. The GDG felt that the length of hospital stay was a useful surrogate marker of improvement in hepatic encephalopathy and was important in assessing the effectiveness of the interventions.
Trade-off between clinical benefits and harms	The GDG noted that the following interventions appeared to show some clinical benefit over non-absorbable disaccharides for at least one outcome: neomycin, rifaximin (for the outcome of mortality only), BCAA, PEG3350 and probiotics (for the outcome of adverse events). However, another study showed no clinical benefit of BCAA over placebo for the outcomes of mortality and positive response of hepatic encephalopathy to treatment. Sodium benzoate showed no clinical benefit over non-absorbable disaccharides. Despite the evidence of clinical effectiveness for some of the interventions compared with non-absorbable disaccharides, the GDG could not discount the Low or Very Low quality of the majority of this evidence. Therefore, the GDG could not be confident that these effect estimates represented the true effect of the interventions. The GDG noted that the evidence for rifaximin versus non-absorbable disaccharides was only available from one study with a high overall mortality rate, conducted in a population which differs considerably from most UK patients with cirrhosis (that is, age and severity). It was agreed that this study was not applicable to a UK population. Rifaximin and BCAAs showed a clinical benefit in combination with non-absorbable disaccharides versus non-absorbable disaccharides alone. The evidence for rifaximin plus non-absorbable disaccharides was of Moderate quality. However, again, evidence was only available from 1 study. The GDG noted the lack of placebo controlled comparisons to assess the effectiveness of individual or combinations of interventions versus treatment of the precipitating cause of the episode of acute hepatic encephalopathy event alone (in the placebo group). Only one study assessed non-absorbable disaccharides (lactitol enemas) versus placebo (tap water enemas) which suggested a clinical benefit of non-absorbable disaccharides. However, the water enema arm of the trial was stopped by the regulator as the mortality rate after the enrolment of the first 5 patients was unacceptably high. Therefore the trial only consisted of 5 people in the placebo arm and 10 people in the lactitol arm. Overall the evidence was Very low or Low quality. In addition, no details were given about the treatment of the precipitating cause of hepatic encephalopathy in either arm. The GDG also noted the improvements in general and intensive care that would have occurred and that there

were no trials where the effects of treatment of the precipitating factor alone on the outcome of hepatic encephalopathy had been examined.

The GDG considered that the current standard treatment of an episode of acute hepatic encephalopathy is directed at reducing the production and absorption of gut-derived neurotoxins, particularly ammonia, mainly through bowel cleansing, and the use of non-absorbable disaccharides, such as lactulose. This treatment is inexpensive and has a low adverse risk profile but there was no evidence in the review to indicate its superiority over other treatment approaches.

Neomycin, although classified as a non-absorbable antibiotic, is absorbed in small quantities and can cause both nephrotoxicity and non-reversible ototoxicity. Although it is inexpensive and has been used in the short-term to treat hepatic encephalopathy when the response to non-absorbable disaccharides is incomplete, the GDG agreed that they would not recommend neomycin based on the limited evidence available.

The GDG discussed the possibility that routine use of rifaximin may be associated with antibiotic resistance but felt that insufficient data were available on this subject. Rifaxamin is currently recommended in the UK by NICE TA337 for the treatment of recurrent episodic hepatic encephalopathy in conjunction with a non-absorbable disaccharide. The GDG did not think there was enough evidence of clinical effectiveness in an episode of acute hepatic encephalopathy trials to warrant an off-licence recommendation.

The GDG were concerned about the practicality of PEG3350 administration; as a preparation made to a total volume of 4 litres it may not be practical to administer to people with grade 3 or 4 hepatic encephalopathy, nor wise in individuals who may also have fluid retention. Its patient acceptability is likely to be low. It was noted that PEG3350 is not licensed for use in an episode of acute hepatic encephalopathy.

The GDG felt that due to the half-life of intravenous flumazenil (40–80 minutes), and the risk of precipitating seizures and arrhythmias, the risks of this treatment would outweigh any potential benefit. This intervention is rarely given in this situation and would require IV administration every few hours as the intervention cannot be given as an infusion. The GDG would not consider recommending flumazenil due to the lack of evidence and these concerns.

The GDG discussed that sodium benzoate is not used to treat episodes of acute hepatic encephalopathy in the UK except rarely in specific instances in tertiary care. The GDG would not consider recommending this intervention as the current evidence did not indicate any clinical benefit over non-absorbable disaccharides.

LOLA is used occasionally in the UK to treat an episode of acute hepatic encephalopathy however its availability outside tertiary care and specialist centres is limited. The evidence suggested a clinical benefit of LOLA when combined with metronidazole and non-absorbable disaccharides, but the GDG agreed that further research is needed to assess the benefits and harms.

Metronidazole is an antibiotic that has a number of adverse effects including dizziness, eye problems and peripheral neuritis. The GDG did not feel that it should be used for this indication.

The GDG felt that MARS is an intensive care intervention and should not be used to treat an episode of acute hepatic encephalopathy.

The GDG discussed that there is currently variation in the treatment of an acute hepatic encephalopathy episode on a national level. The GDG agreed this is an important area and a recommendation is needed to standardise practice. However, the evidence for each comparison was sparse and the evidence was of Low or Very Low quality. The lack of High quality evidence was not necessarily a lack of effect of any of the interventions, but the GDG agreed there was insufficient evidence to make a recommendation. The GDG thought that further research is needed and have made a future research recommendation. They discussed that the most commonly used intervention currently is lactulose, and that this current practice should

	continue until further research is carried out. The GDG noted the lack of High quality evidence from placebo controlled trials in this area, perhaps because it would be unethical not to give an intervention to someone in this clinical situation. However, they all agreed that the identification and treatment of the precipitant of the episode of acute hepatic encephalopathy may account for a proportion of the clinical benefit observed in practice. The GDG also noted that many of the studies were old and treatments of the precipitating cause of an episode of acute hepatic encephalopathy event will have changed considerably since they were published. The GDG agreed that evidence is needed to show effectiveness of current interventions before they could consider head-to-head comparisons and comparisons of different combinations. Thus, it considered that in people with cirrhosis experiencing an episode of acute hepatic encephalopathy with a clear, well-defined precipitating factor it would be reasonable to investigate whether treatment of the precipitating factor alone would be effective in ameliorating the hepatic encephalopathy. This would be done by randomising participants to treatment of the precipitant factor alone or with the addition of lactulose. If it were shown that treatment with lactulose provided additional benefit then the comparative efficacy and safety of other agents such as rifaximin, LOLA or BCAA could be explored.
Trade-off between net clinical effects and costs	No relevant published economic evidence was identified. The GDG considered the standard UK costs of the alternative interventions, but as the GDG decided that there was insufficient clinical evidence to recommend an intervention, that decision was not taken on economic grounds.
Quality of evidence	The GDG noted that the studies included in the evidence base spanned a 33-year period and that many studies were over 15 years old. The most recent study was from 2014. The majority of the evidence found was of Low to Very Low quality. The GDG noted that this may be due to a lack of evidence (with evidence only available from small RCTs for the majority of comparisons), rather than a lack of effect. The GDG also noted that the lack of High quality evidence of any treatment versus placebo made the head-to-head comparisons of single or combination therapy more difficult to interpret. A number of results did have Moderate quality evidence and the GDG noted these when considering a recommendation. When considering the results of individual studies, the GDG noted that many of the studies were conducted more than 15 years ago and some much longer than this. The GDG felt that the standard of care for the management of acute hepatic encephalopathy (particularly intensive care support) would be significantly better now than in the 1980s and 1990s.
Other considerations	A number of studies of rifaximin were excluded because they were looking at the treatment of chronic or minimal hepatic encephalopathy, or the prophylaxis of hepatic encephalopathy. **Research recommendation** Hepatic encephalopathy is a major complication of cirrhosis. Approximately 50% of people with cirrhosis will develop clinically apparent hepatic encephalopathy at some stage after diagnosis – the risk being around 5–25% within 5 years. Hospital admissions are common and inpatient stays often prolonged. The presence of hepatic encephalopathy is associated with a significant increase in mortality; survival after the first episode is 42% at 1 year and 23% at 3 years. At present, treatment of the hepatic encephalopathy is directed primarily at reducing the production and absorption of gut-derived neurotoxins, particularly ammonia, mainly through bowel cleansing, and the use of non-absorbable disaccharides, such as lactulose, although several other agents such as non-

	absorbable antibiotics are also used. However, in approximately 50% of people admitted with episodic hepatic encephalopathy there is a clearly defined precipitating factor (for example, infections, gastrointestinal bleeding or overuse of diuretics). Treatment is often challenging and some people may need to be cared for in an intensive care setting, at least initially. The identification and correction of any precipitating events is important as there is evidence that this alone may improve hepatic encephalopathy without recourse to specific therapies. However, this has not been rigorously tested in a randomised clinical trial.

16 Reference list

1 202 Endoscopic variceal band ligation in comparison with propranolol in prophylaxis of first variceal bleeding in patients with liver cirrhosis. Journal of Hepatology. 2005; 42(Suppl.1):79

2 Review Manager (RevMan) [Computer program]. Version 5. Copenhagen. The Nordic Cochrane Centre, The Cochrane Collaboration, 2015. Available from: http://tech.cochrane.org/Revman

3 WinBUGS [Computer programme] version 1.4. Cambridge. MRC Biostatistics Unit, University of Cambridge, 2015. Available from: http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/

4 Abdelfattah MH, Rashed MA, Elfakhry AA, Soliman MA, Shiha GE. 201 Endoscopic variceal ligation versus pharmacologic treatment for primary prophylaxis of variceal bleeding: a randomised study. Journal of Hepatology. 2006; 44(Suppl.1):S83

5 Abid S, Jafri W, Mumtaz K, Islam M, Abbas Z, Shah HA et al. Efficacy of L-ornithine-L-aspartate as an adjuvant therapy in cirrhotic patients with hepatic encephalopathy. Journal of the College of Physicians and Surgeons--Pakistan. 2011; 21(11):666-671

6 Abulfutuh AR, Morsy M, Solyman AEG, Hendawy SE, Desouky ME, Hadad SE et al. Study of variceal band ligation, propranolol and isosorbide mononitrate in the prevention of the first variceal bleeding. Gastroenterology. 2003; 124(4):A780

7 Ahmad I, Khan AA, Alam A, Dilshad A, Butt AK, Shafqat F et al. L-ornithine-L-aspartate infusion efficacy in hepatic encephalopathy. Journal of the College of Physicians and Surgeons--Pakistan. 2008; 18(11):684-687

8 Andreani T, Poupon R, Balkau BJ, Trinchet JC, Grange JD, Peigney N et al. Preventive therapy of first gastrointestinal bleeding in patients with cirrhosis: results of a controlled trial comparing propranolol, endoscopic sclerotherapy and placebo. Hepatology. 1990; 12(6):1413-1419

9 Aravinthan A, Pietrosi G, Hoare M, Jupp J, Marshall A, Verrill C et al. Hepatocyte expression of the senescence marker p21 is linked to fibrosis and an adverse liver-related outcome in alcohol-related liver disease. PloS One. 2013; 8(9):e72904

10 Arena U, Vizzutti F, Abraldes JG, Corti G, Stasi C, Moscarella S et al. Reliability of transient elastography for the diagnosis of advanced fibrosis in chronic hepatitis C. Gut. 2008; 57(9):1288-1293

11 Arroyo V, Guevara M, Gines P. Hepatorenal syndrome in cirrhosis: pathogenesis and treatment. Gastroenterology. 2002; 122(6):1658-1676

12 Arvaniti V, D'Amico G, Fede G, Manousou P, Tsochatzis E, Pleguezuelo M et al. Infections in patients with cirrhosis increase mortality four-fold and should be used in determining prognosis. Gastroenterology. 2010; 139(4):1246-1256

13 Askgaard G, Gronbaek M, Kjaer MS, Tjonneland A, Tolstrup JS. Alcohol drinking pattern and risk of alcoholic liver cirrhosis: a prospective cohort study. Journal of Hepatology. 2015; 62(5):1061-1067

14 Aykut UE, Akyuz U, Yesil A, Eren F, Gerin F, Ergelen R et al. A comparison of FibroMeter NAFLD Score, NAFLD fibrosis score, and transient elastography as noninvasive diagnostic tools for

hepatic fibrosis in patients with biopsy-proven non-alcoholic fatty liver disease. Scandinavian Journal of Gastroenterology. 2014; 49(11):1343-1348

15 Becker U, Gronbaek M, Johansen D, Sorensen TIA. Lower risk for alcohol-induced cirrhosis in wine drinkers. Hepatology. 2002; 35(4):868-875

16 Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. Hepatology. 2003; 38(6):1449-1457

17 Bendtsen F, Skovgaard LT, Sorensen TI, Matzen P. Agreement among multiple observers on endoscopic diagnosis of esophageal varices before bleeding. Hepatology. 1990; 11(3):341-347

18 Benvegnu L, Gios M, Boccato S, Alberti A. Natural history of compensated viral cirrhosis: a prospective study on the incidence and hierarchy of major complications. Gut. 2004; 53(5):744-749

19 Bernard B, Cadranel JF, Valla D, Escolano S, Jarlier V, Opolon P. Prognostic significance of bacterial infection in bleeding cirrhotic patients: a prospective study. Gastroenterology. 1995; 108(6):1828-1834

20 Bernardi M, Caraceni P, Navickis RJ, Wilkes MM. Albumin infusion in patients undergoing large-volume paracentesis: a meta-analysis of randomized trials. Hepatology. 2012; 55(4):1172-1181

21 Blackwelder WC, Yano K, Rhoads GG, Kagan A, Gordon T, Palesch Y. Alcohol and mortality: the Honolulu Heart Study. American Journal of Medicine. 1980; 68(2):164-169

22 Borroni G, Ceriani R, Cazzaniga M, Tommasini M, Roncalli M, Maltempo C et al. Comparison of simple tests for the non-invasive diagnosis of clinically silent cirrhosis in chronic hepatitis C. Alimentary Pharmacology and Therapeutics. 2006; 24(5):797-804

23 Bota S, Sirli R, Sporea I, Focsa M, Popescu A, Danila M et al. A new scoring system for prediction of fibrosis in chronic hepatitis C. Hepatitis Monthly. 2011; 11(7):548-555

24 Bota S, Sporea I, Sirli R, Popescu A, Gradinaru-Tascau O. How useful are ARFI elastography cut-off values proposed by meta-analysis for predicting the significant fibrosis and compensated liver cirrhosis? Medical Ultrasonography. 2015; 17(2):200-205

25 Cadranel JF, Rufat P, Degos F. Practices of liver biopsy in France: results of a prospective nationwide survey. For the Group of Epidemiology of the French Association for the Study of the Liver (AFEF). Hepatology. 2000; 32(3):477-481

26 Cales P, Zabotto B, Meskens C, Caucanas JP, Vinel JP, Desmorat H et al. Gastroesophageal endoscopic features in cirrhosis. Observer variability, interassociations, and relationship to hepatic dysfunction. Gastroenterology. 1990; 98(1):156-162

27 Canavan C, Eisenburg J, Meng L, Corey K, Hur C. Ultrasound elastography for fibrosis surveillance is cost effective in patients with chronic hepatitis C virus in the UK. Digestive Diseases and Sciences. 2013; 58(9):2691-2704

28 Carbonell N, Pauwels A, Serfaty L, Fourdan O, Levy VG, Poupon R. Improved survival after variceal bleeding in patients with cirrhosis over the past two decades. Hepatology. 2004; 40(3):652-659

29 Cardoso AC, Carvalho-Filho RJ, Stern C, Dipumpo A, Giuily N, Ripault MP et al. Direct comparison of diagnostic performance of transient elastography in patients with chronic hepatitis B and chronic hepatitis C. Liver International. 2012; 32(4):612-621

30 Castera L, Sebastiani G, Le Bail B, de Ledinghen V, Couzigou P, Alberti A. Prospective comparison of two algorithms combining non-invasive methods for staging liver fibrosis in chronic hepatitis C. Journal of Hepatology. 2010; 52(2):191-198

31 Catanzaro R, Milazzo M, Arona S, Sapienza C, Vasta D, Arcoria D et al. Diagnostic accuracy of enhanced liver fibrosis test to assess liver fibrosis in patients with chronic hepatitis C. Hepatobiliary and Pancreatic Diseases International. 2013; 12(5):500-507

32 Caviglia GP, Ciancio A, Rosso C, Abate ML, Olivero A, Pellicano R et al. Non-invasive methods for the assessment of hepatic fibrosis: transient elastography, hyaluronic acid, 13C-aminopyrine breath test and cytokeratin 18 fragment. Annals of Hepatology. 2013; 13(1):91-97

33 Cerra FB, Cheung NK, Fischer JE, Kaplowitz N, Schiff ER, Dienstag JL et al. Disease-specific amino acid infusion (F080) in hepatic encephalopathy: a prospective, randomized, double-blind, controlled trial. Journal of Parenteral and Enteral Nutrition. 1985; 9(3):288-295

34 Cerra FB, McMillen M, Angelico R, Cline B, Lyons J, Faulkenbach L et al. Cirrhosis, encephalopathy, and improved results with metabolic support. Surgery. 1983; 94(4):612-619

35 Chasalani N, Imperiale TF, Ismail A, Carey M, Wilcox CM, Madichetty H et al. Predictors of large esophageal varices in patients with cirrhosis. American Journal of Gastroenterology. 1999; 94(11):3285-3291

36 Chavez-Tapia NC, Barrientos-Gutierrez T, Tellez-Avila F, Soares-Weiser K, Mendez-Sanchez N, Gluud C et al. Meta-analysis: Antibiotic prophylaxis for cirrhotic patients with upper gastrointestinal bleeding - An updated Cochrane review. Alimentary Pharmacology and Therapeutics. 2011; 34(5):509-518

37 Chavez-Tapia NC, Barrientos-Gutierrez T, Tellez-Avila F, I, Soares-Weiser K, Uribe M. Antibiotic prophylaxis for cirrhotic patients with upper gastrointestinal bleeding. Cochrane Database of Systematic Reviews. 2010;(9):CD002907. DOI:10.1002/14651858.CD002907.pub2

38 Chen CY, Sheu MZ, Su SY. Prophylactic endoscopic variceal ligation (EVL) with multiple band ligator for esophageal varices. Gastroenterology. 1998; 114(Suppl.1):A1224

39 Chen SH, Li YF, Lai HC, Kao JT, Peng CY, Chuang PH et al. Effects of patient factors on noninvasive liver stiffness measurement using acoustic radiation force impulse elastography in patients with chronic hepatitis C. BMC Gastroenterology. 2012; 12:105

40 Chrysanthos NV, Papatheodoridis GV, Savvas S, Kafiri G, Petraki K, Manesis EK et al. Aspartate aminotransferase to platelet ratio index for fibrosis evaluation in chronic viral hepatitis. European Journal of Gastroenterology and Hepatology. 2006; 18(4):389-396

41 Cohen MJ, Sahar T, Benenson S, Elinav E, Brezis M, Soares-Weiser K. Antibiotic prophylaxis for spontaneous bacterial peritonitis in cirrhotic patients with ascites, without gastro-intestinal bleeding. Cochrane Database of Systematic Reviews. 2009;(2):CD004791. DOI:10.1002/14651858.CD004791.pub2.

42 Conn HO, Grace N, Bosch J, Groszmann R, Rodes J, Wright SC et al. Propranolol in the prevention of the first hemorrhage from esophagogastric varices: A multicenter, randomized clinical trial. Hepatology. 1991; 13(5):902-912

43 Crossan C, Tsochatzis EA, Longworth L, Gurusamy K, Davidson B, Rodriguez-Peralvarez M et al. Cost-effectiveness of non-invasive methods for assessment and monitoring of liver fibrosis and cirrhosis in patients with chronic liver disease: systematic review and economic evaluation. Health Technology Assessment. 2015; 19(9):1-410

44 Cucchetti A, Trevisani F, Cescon M, Ercolani G, Farinati F, Poggio PD et al. Cost-effectiveness of semi-annual surveillance for hepatocellular carcinoma in cirrhotic patients of the Italian Liver Cancer population. Journal of Hepatology. 2012; 56(5):1089-1096

45 Cucchetti A, Trevisani F, Pecorelli A, Erroi V, Farinati F, Ciccarese F et al. Estimation of lead-time bias and its impact on the outcome of surveillance for the early diagnosis of hepatocellular carcinoma. Journal of Hepatology. 2014; 61(2):333-341

46 D'Amico G, Pagliaro L, Bosch J. Pharmacological treatment of portal hypertension: an evidence-based approach. Seminars in Liver Disease. 1999; 19(4):475-505

47 D'Amico G, Luca A, Morabito A, Miraglia R, D'Amico M. Uncovered transjugular intrahepatic portosystemic shunt for refractory ascites: a meta-analysis. Gastroenterology. 2005; 129(4):1282-1293

48 D'Antiga L, Betalli P, de Angelis P, Davenport M, Di Giorgio A, McKiernan PJ et al. Interobserver Agreement on Endoscopic Classification of Oesophageal Varices in Children. Journal of Pediatric Gastroenterology and Nutrition. 2015; 61(2):176-181

49 Davies SC. Annual report of the Chief Medical Officer: On the State of the Public's Health. London. Department of Health, 2011

50 de Franchis R. Evolving consensus in portal hypertension. Report of the Baveno IV consensus workshop on methodology of diagnosis and therapy in portal hypertension. Journal of Hepatology. 2005; 43(1):167-176

51 de la Mora JG, Farca-Belsaguy AA, Uribe M, de Hoyos-Garza A. Ligation VS propranolol for primary prophilaxis of variceal bleeding using a multiple band liga-tor and objective measurements of treatment adequacy: Preliminary results. Gastroenterology. 2000; 118(4):A1434-A1435

52 de Ledinghen V, Douvin C, Kettaneh A, Ziol M, Roulot D, Marcellin P et al. Diagnosis of hepatic fibrosis and cirrhosis by transient elastography in HIV/hepatitis C virus-coinfected patients. Journal of Acquired Immune Deficiency Syndromes. 2006; 41(2):175-179

53 De BK, Ghoshal UC, Das T, Santra A, Biswas PK. Endoscopic variceal ligation for primary prophylaxis of oesophageal variceal bleed: preliminary report of a randomized controlled trial. Journal of Gastroenterology and Hepatology. 1999; 14(3):220-224

54 Drastich P, Lata J, Petrtyl J, Bruha R, Prochazka V, Vanasek T et al. Endoscopic variceal band ligation compared with propranolol for prophylaxis of first variceal bleeding. Annals of Hepatology. 2011; 10(2):142-149

55 Ebinuma H, Saito H, Komuta M, Ojiro K, Wakabayashi K, Usui S et al. Evaluation of liver fibrosis by transient elastography using acoustic radiation force impulse: comparison with Fibroscan(®). Journal of Gastroenterology. 2011; 46(10):1238-1248

56 El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology. 2007; 132(7):2557-2576

57 Esmat G, Elsharkawy A, El Akel W, Fouad A, Helal K, Mohamed MK et al. Fibroscan of chronic HCV patients coinfected with schistosomiasis. Arab Journal of Gastroenterology. 2013; 14(3):109-112

58 Fahmy MI, Badran HM. Comparison of transient elastography to Doppler indices in prediction of hepatitis C induced liver fibrosis and cirrhosis. Egyptian Journal of Radiology and Nuclear Medicine. 2011; 42(2):111-117

59 Fasani P, Sangiovanni A, De Fazio C, Borzio M, Bruno S, Ronchi G et al. High prevalence of multinodular hepatocellular carcinoma in patients with cirrhosis attributable to multiple risk factors. Hepatology. 1999; 29(6):1704-1707

60 Ferenci P, Lockwood A, Mullen K, Tarter R, Weissenborn K, Blei A. Hepatic encephalopathy--definition, nomenclature, diagnosis, and quantification: final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. Hepatology. 2002; 35(3):716-721

61 Ferlitsch M, Reiberger T, Hoke M, Salzl P, Schwengerer B, Ulbrich G et al. Von Willebrand factor as new noninvasive predictor of portal hypertension, decompensation and mortality in patients with liver cirrhosis. Hepatology. 2012; 56(4):1439-1447

62 Fernandes FF, Ferraz ML, Andrade LE, Dellavance A, Terra C, Pereira G et al. Enhanced liver fibrosis panel as a predictor of liver fibrosis in chronic hepatitis C patients. Journal of Clinical Gastroenterology. 2015; 49(3):235-241

63 Fernandez J, Acevedo J, Castro M, Garcia O, de Lope CR, Roca D et al. Prevalence and risk factors of infections by multiresistant bacteria in cirrhosis: a prospective study. Hepatology. 2012; 55(5):1551-1561

64 Fernandez J, Gustot T. Management of bacterial infections in cirrhosis. Journal of Hepatology. 2012; 56(Suppl.1):S1-12

65 Fernandez J, Navasa M, Gomez J, Colmenero J, Vila J, Arroyo V et al. Bacterial infections in cirrhosis: epidemiological changes with invasive procedures and norfloxacin prophylaxis. Hepatology. 2002; 35(1):140-148

66 Fernandez J, Navasa M, Planas R, Montoliu S, Monfort D, Soriano G et al. Primary prophylaxis of spontaneous bacterial peritonitis delays hepatorenal syndrome and improves survival in cirrhosis. Gastroenterology. 2007; 133(3):818-824

67 Fernandez J, Ruiz del Arbol L, Gomez C, Durandez R, Serradilla R, Guarner C. Norfloxacin vs ceftriaxone in the prophylaxis of infections in patients with advanced cirrhosis and hemorrhage. Gastroenterology. 2006; 131(4):1049-1056

68 Fernandez J, Monteagudo J, Bargallo X, Jimenez W, Bosch J, Arroyo V et al. A randomized unblinded pilot study comparing albumin versus hydroxyethyl starch in spontaneous bacterial peritonitis. Hepatology. 2005; 42(3):627-634

69 Ferraioli G, Tinelli C, Dal Bello B, Zicchetti M, Filice G, Filice C et al. Accuracy of real-time shear wave elastography for assessing liver fibrosis in chronic hepatitis C: a pilot study. Hepatology. 2012; 56(6):2125-2133

70 Ferraioli G, Tinelli C, Lissandrin R, Zicchetti M, Dal Bello B, Filice G et al. Point shear wave elastography method for assessing liver stiffness. World Journal of Gastroenterology. 2014; 20(16):4787-4796

71 Fiaccadori F, Ghinelli F, Pedretti G, Pelosi G, Sacchini D, Zeneroli ML. Branched chain amino acid enriched solutions in the treatment of hepatic encephalopathy: a controlled trial. In: Capocaccia L, Fischer JE, Rossi-Fanelli F (eds), Hepatic encephalopathy in chronic liver failure, New York: Plenum Press, 1984: 323-333

72 Fierbinteanu Braticevici C, Sporea I, Panaitescu E, Tribus L. Value of acoustic radiation force impulse imaging elastography for non-invasive evaluation of patients with nonalcoholic fatty liver disease. Ultrasound in Medicine and Biology. 2013; 39(11):1942-1950

73 Fierbinteanu-Braticevici C, Andronescu D, Usvat R, Cretoiu D, Baicus C, Marinoschi G. Acoustic radiation force imaging sonoelastography for noninvasive staging of liver fibrosis. World Journal of Gastroenterology. 2009; 15(44):5525-5532

74 Finkenstedt A, Dorn L, Edlinger M, Prokop W, Risch L, Griesmacher A et al. Cystatin C is a strong predictor of survival in patients with cirrhosis: Is a cystatin C-based MELD better? Liver International. 2012; 32(8):1211-1216

75 Floreani A, Cazzagon N, Martines D, Cavalletto L, Baldo V, Chemello L. Performance and utility of transient elastography and noninvasive markers of liver fibrosis in primary biliary cirrhosis. Digestive and Liver Disease. 2011; 43(11):887-892

76 Fontana RJ, Lok AS. Noninvasive monitoring of patients with chronic hepatitis C. Hepatology. 2002; 36(5 Suppl 1):S57-S64

77 Friedrich-Rust M, Hadji-Hosseini H, Kriener S, Herrmann E, Sircar I, Kau A et al. Transient elastography with a new probe for obese patients for non-invasive staging of non-alcoholic steatohepatitis. European Radiology. 2010; 20(10):2390-2396

78 Friedrich-Rust M, Rosenberg W, Parkes J, Herrmann E, Zeuzem S, Sarrazin C. Comparison of ELF, FibroTest and FibroScan for the non-invasive assessment of liver fibrosis. BMC Gastroenterology. 2010; 10:103

79 Fuchs CS, Stampfer MJ, Colditz GA, Giovannucci EL, Manson JE, Kawachi I et al. Alcohol consumption and mortality among women. New England Journal of Medicine. 1995; 332(19):1245-1250

80 Fujii H, Enomoto M, Fukushima W, Ohfuji S, Mori M, Kobayashi S et al. Noninvasive laboratory tests proposed for predicting cirrhosis in patients with chronic hepatitis C are also useful in patients with non-alcoholic steatohepatitis. Journal of Gastroenterology. 2009; 44(6):608-614

81 Gaia S, Carenzi S, Barilli AL, Bugianesi E, Smedile A, Brunello F et al. Reliability of transient elastography for the detection of fibrosis in Non-Alcoholic Fatty Liver Disease and chronic viral hepatitis. Journal of Hepatology. 2011; 54(1):64-71

82 Gheorghe C, Gheorghe L, Vadan R, Hrehoret D, Popescu I. Prophylactic banding ligation of high risk esophageal varices inpatients on the waiting list for liver transplantation: an interim report. Journal of Hepatology. 2002; 36(Suppl.1):38

83 Giannini E, Arzani L, Borro P, Botta F, Fasoli A, Risso D et al. Does surveillance for hepatocellular carcinoma in HCV cirrhotic patients improve treatment outcome mainly due to better clinical status at diagnosis? Hepato-Gastroenterology. 2000; 47(35):1395-1398

84 Gines P, Arroyo V, Quintero E, Planas R, Bory F, Cabrera J et al. Comparison of paracentesis and diuretics in the treatment of cirrhotics with tense ascites. Results of a randomized study. Gastroenterology. 1987; 93(2):234-241

85 Gines P, Quintero E, Arroyo V, Teres J, Bruguera M, Rimola A et al. Compensated cirrhosis: natural history and prognostic factors. Hepatology. 1987; 7(1):122-128

86 Gines P, Rimola A, Planas R, Vargas V, Marco F, Almela M et al. Norfloxacin prevents spontaneous bacterial peritonitis recurrence in cirrhosis: results of a double-blind, placebo-controlled trial. Hepatology. 1990; 12(4 Pt.1):716-724

87 Gines P, Uriz J, Calahorra B, Garcia-Tsao G, Kamath PS, Del Arbol LR et al. Transjugular intrahepatic portosystemic shunting versus paracentesis plus albumin for refractory ascites in cirrhosis. Gastroenterology. 2002; 123(6):1839-1847

88 Gluud LL, Krag A. Banding ligation versus beta-blockers for primary prevention in oesophageal varices in adults. Cochrane Database of Systematic Reviews. 2012;(8):CD004544. DOI:10.1002/14651858.CD004544.pub2

89 Grace N. Diagnosis and treatment of gastrointestinal bleeding secondary to portal hypertension. American College of Gastroenterology Practice Parameters Committee. American Journal of Gastroenterology. 1997; 92(7):1081-1091

90 Grace N, Groszmann R, Garcia-Tsao G, Burroughs A, Pagliaro L, Makuch R et al. Portal hypertension and variceal bleeding: an AASLD single topic symposium. Hepatology. 1998; 28(3):868-880

91 GRADE Working Group. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group website. 2011. Available from: http://www.gradeworkinggroup.org/ [Last accessed: 8 October 2015]

92 Grange JD, Roulot D, Pelletier G, Pariente EA, Denis J, Ink O et al. Norfloxacin primary prophylaxis of bacterial infections in cirrhotic patients with ascites: a double-blind randomized trial. Journal of Hepatology. 1998; 29(3):430-436

93 Groszmann R, Bosch J, Grace N, Conn HO, Garcia-Tsao G, Navasa M et al. Hemodynamic events in a prospective randomized trial of propranolol versus placebo in the prevention of a first variceal hemorrhage. Gastroenterology. 1990; 99(5):1401-1407

94 Guechot J, Trocme C, Renversez JC, Sturm N, Zarski JP, ANRS HC EP 23 Fibrostar Study Group. Independent validation of the Enhanced Liver Fibrosis (ELF) score in the ANRS HC EP 23 Fibrostar cohort of patients with chronic hepatitis C. Clinical Chemistry and Laboratory Medicine. 2012; 50(4):693-699

95 Gyr K, Meier R, Haussler J, Bouletreau P, Fleig WE, Gatta A et al. Evaluation of the efficacy and safety of flumazenil in the treatment of portal systemic encephalopathy: a double blind, randomised, placebo controlled multicentre study. Gut. 1996; 39(2):319-324

96 Halfon P, Bacq Y, de Muret A, Penaranda G, Bourliere M, Ouzan D et al. Comparison of test performance profile for blood tests of liver fibrosis in chronic hepatitis C. Journal of Hepatology. 2007; 46(3):395-402

97 Hassanein T, Tofteng F, Brown RS, Jr., McGuire B, Lynch P, Mehta R et al. Randomized controlled study of extracorporeal albumin dialysis for hepatic encephalopathy in advanced cirrhosis. Hepatology. 2007; 46(6):1853-1862

98 Health & Social Care Information Centre. Provisional Monthly Hospital Episode Statistics for Admitted Patient Care, Outpatients and Accident and Emergency Data - April 2012 to December 2012. 2013. Available from: http://www.hscic.gov.uk/catalogue/PUB10706 [Last accessed: 20 October 2015]

99 Health & Social Care Information Centre. Count of finished admission episodes (FAEs) with a primary diagnosis of diseases of liver from 2005/06 to 2014/154, at England level. Hospital Episode Statistics. 2015. Available from: http://www.hscic.gov.uk/media/18177/Liver-Diseases-2005-06-to-2014-15/xls/Liver_Diseases_2005-06_to_2014-15.xlsx [Last accessed: 20 October 2015]

100 Imperiale TF, Chalasani N. A meta-analysis of endoscopic variceal ligation for primary prophylaxis of esophageal variceal bleeding. Hepatology. 2001; 33(4):802-807

101 Imperiale TF, Klein RW, Chalasani N. Cost-effectiveness analysis of variceal ligation vs beta-blockers for primary prevention of variceal bleeding. Hepatology. 2007; 45(4):870-878

102 Ioannou GN, Perkins JD, Carithers RL. Liver transplantation for hepatocellular carcinoma: impact of the MELD allocation system and predictors of survival. Gastroenterology. 2008; 134(5):1342-1351

103 Ioannou GN, Weiss NS, Kowdley K, Dominitz JA. Is obesity a risk factor for cirrhosis-related death or hospitalization? A population-based cohort study. Gastroenterology. 2003; 125(4):1053-1059

104 Jalan R, Fernandez J, Wiest R, Schnabl B, Moreau R, Angeli P et al. Bacterial infections in cirrhosis: a position statement based on the EASL Special Conference 2013. Journal of Hepatology. 2014; 60(6):1310-1324

105 Janssens F, de Suray N, Piessevaux H, Horsmans Y, de Timary P, Starkel P. Can transient elastography replace liver histology for determination of advanced fibrosis in alcoholic patients: a real-life study. Journal of Clinical Gastroenterology. 2010; 44(8):575-582

106 Jutabha R, Jensen DM, Martin P, Savides T, Han S, Gornbein J. Randomized study comparing banding and propranolol to prevent initial variceal hemorrhage in cirrhotics with high-risk esophageal varices. Gastroenterology. 2005; 128(4):870-881

107 Jutabha R, Jensen DM, Martin P, Savides TJ, Lam F, Jensen ME et al. Initial report of a randomized, prospective study of prophylactic propranolol compared to rubber band ligation for prevention of first variceal hemorrhage in cirrhotics with large esophageal varices. Gastroenterology. 2000; 118(4):A212-A213

108 Kayadibi H, Yasar B, Ozkara S, Serdar MA, Kurdas OO, Gonen C. The diagnostic accuracy of the Forns index, platelet count and AST to Platelet Ratio Index derived fibrosis index for the prediction of Hepatitis C virus-related significant liver fibrosis and cirrhosis. Scandinavian Journal of Clinical and Laboratory Investigation. 2014; 74(3):240-247

109 Kettaneh A, Marcellin P, Douvin C, Poupon R, Ziol M, Beaugrand M et al. Features associated with success rate and performance of FibroScan measurements for the diagnosis of cirrhosis in HCV patients: a prospective study of 935 patients. Journal of Hepatology. 2007; 46(4):628-634

110 Kim BI, Kim HJ, Park JH, Park DI, Cho YK, Sohn CI et al. Increased intestinal permeability as a predictor of bacterial infections in patients with decompensated liver cirrhosis and hemorrhage. Journal of Gastroenterology and Hepatology. 2011; 26(3):550-557

111 Kim BK, Park YN, Kim DY, Park JY, Chon CY, Han KH et al. Risk assessment of development of hepatic decompensation in histologically proven hepatitis B viral cirrhosis using liver stiffness measurement. Digestion. 2012; 85(3):219-227

112 Kim MN, Kim SU, Park JY, Kim DY, Han KH, Chon CY et al. Risk assessment of liver-related events using transient elastography in patients with chronic hepatitis B receiving entecavir. Journal of Clinical Gastroenterology. 2014; 48(3):272-278

113 Klatsky AL, Armstrong MA. Alcohol, smoking, coffee, and cirrhosis. American Journal of Epidemiology. 1992; 136(10):1248-1257

114 Klibansky DA, Mehta SH, Curry M, Nasser I, Challies T, Afdhal N. Transient elastography for predicting clinical outcomes in patients with chronic liver disease. Journal of Viral Hepatitis. 2012; 19(2):e184-e193

115 Laccetti M, Manes G, Uomo G, Lioniello M, Rabitti PG, Balzano A. Flumazenil in the treatment of acute hepatic encephalopathy in cirrhotic patients: a double blind randomized placebo controlled study. Digestive and Liver Disease. 2000; 32(4):335-338

116 Lackner C, Struber G, Liegl B, Leibl S, Ofner P, Bankuti C et al. Comparison and validation of simple noninvasive tests for prediction of fibrosis in chronic hepatitis C. Hepatology. 2005; 41(6):1376-1382

117 Lay CS, Tsai YT, Lee FY, Lai YL, Yu CJ, Chen CB et al. Endoscopic variceal ligation versus propranolol in prophylaxis of first variceal bleeding in patients with cirrhosis. Journal of Gastroenterology and Hepatology. 2006; 21(2):413-419

118 Lay CS, Tsai YT, Teg CY, Shyu WS, Guo WS, Wu KL et al. Endoscopic variceal ligation in prophylaxis of first variceal bleeding in cirrhotic patients with high-risk esophageal varices. Hepatology. 1997; 25(6):1346-1350

119 Lebrec D, Giuily N, Hadengue A, Vilgrain V, Moreau R, Poynard T et al. Transjugular intrahepatic portosystemic shunts: comparison with paracentesis in patients with cirrhosis and refractory ascites: a randomized trial. French Group of Clinicians and a Group of Biologists. Journal of Hepatology. 1996; 25(2):135-144

120 Leroy V, Sturm N, Faure P, Trocme C, Marlu A, Hilleret MN et al. Prospective evaluation of FibroTest, FibroMeter, and HepaScore for staging liver fibrosis in chronic hepatitis B: comparison with hepatitis C. Journal of Hepatology. 2014; 61(1):28-34

121 Liu B, Balkwill A, Reeves G, Beral V, Million Women Study Collaborators. Body mass index and risk of liver cirrhosis in middle aged UK women: prospective study. BMJ. 2010; 340:c912

122 Liu J, Jing Z-F, Yan J-H, Luo M, Hu R. Therapeutic effects of octreotide combined with alprostadil on cirrhosis with hepatorenal syndrome. World Chinese Journal of Digestology. 2014; 22(14):2034-2038

123 Llovet JM, Bustamante J, Castells A, Vilana R, Ayuso MC, Sala M et al. Natural history of untreated nonsurgical hepatocellular carcinoma: rationale for the design and evaluation of therapeutic trials. Hepatology. 1999; 29(1):62-67

124 Lo GH, Chen WC, Chen MH, Lin CP, Lo CC, Hsu PI et al. Endoscopic ligation vs. nadolol in the prevention of first variceal bleeding in patients with cirrhosis. Gastrointestinal Endoscopy. 2004; 59(3):333-338

125 Lo GH, Lai KH, Cheng JS, Lin CK, Hsu PI, Chiang HT. Prophylactic banding ligation of high-risk esophageal varices in patients with cirrhosis: a prospective, randomized trial. Journal of Hepatology. 1999; 31(3):451-456

126 Loguercio C, Del Vecchio BC, Coltorti M. Enterococcus lactic acid bacteria strain SF68 and lactulose in hepatic encephalopathy: a controlled study. Journal of International Medical Research. 1987; 15(6):335-343

127 Lui HF, Stanley AJ, Forrest E, Jalan R, Hislop WS, Mills PR et al. Primary prophylaxis of variceal hemorrhage: a randomized controlled trial comparing band ligation, propranolol, and isosorbide mononitrate. Gastroenterology. 2002; 123(3):735-744

128 Lupsor Platon M, Stefanescu H, Feier D, Maniu A, Badea R. Performance of unidimensional transient elastography in staging chronic hepatitis C. Results from a cohort of 1,202 biopsied patients from one single center. Journal of Gastrointestinal and Liver Diseases. 2013; 22(2):157-166

129 Lupsor M, Badea R, Stefanescu H, Sparchez Z, Branda H, Serban A et al. Performance of a new elastographic method (ARFI technology) compared to unidimensional transient elastography in the noninvasive assessment of chronic hepatitis C. Preliminary results. Journal of Gastrointestinal and Liver Diseases. 2009; 18(3):303-310

130 Macias J, Giron-Gonzalez JA, Gonzalez-Serrano M, Merino D, Cano P, Mira JA et al. Prediction of liver fibrosis in human immunodeficiency virus/hepatitis C virus coinfected patients by simple non-invasive indexes. Gut. 2006; 55(3):409-414

131 Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PC. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. Hepatology. 2000; 31(4):864-871

132 Martinez SM, Fernandez-Varo G, Gonzalez P, Sampson E, Bruguera M, Navasa M et al. Assessment of liver fibrosis before and after antiviral therapy by different serum marker panels in patients with chronic hepatitis C. Alimentary Pharmacology and Therapeutics. 2011; 33(1):138-148

133 Mas A, Rodes J, Sunyer L, Rodrigo L, Planas R, Vargas V et al. Comparison of rifaximin and lactitol in the treatment of acute hepatic encephalopathy: results of a randomized, double-blind, double-dummy, controlled clinical trial. Journal of Hepatology. 2003; 38(1):51-58

134 Merli M, Nicolini G, Angeloni S, Rinaldi V, De SA, Merkel C et al. Incidence and natural history of small esophageal varices in cirrhotic patients. Journal of Hepatology. 2003; 38(3):266-272

135 Miquel M, Sopena J, Vergara M, Gil M, Casas M, Sanchez-Delgado J et al. Factors related to survival in hepatocellular carcinoma in the geographic area of Sabadell (Catalonia, Spain). Revista Espanola De Enfermedades Digestivas. 2012; 104(5):242-247

136 Mueller S, Millonig G, Sarovska L, Friedrich S, Reimann FM, Pritsch M et al. Increased liver stiffness in alcoholic liver disease: differentiating fibrosis from steatohepatitis. World Journal of Gastroenterology. 2010; 16(8):966-972

137 Myburgh JA, Finfer S, Bellomo R, Billot L, Cass A, Gattas D et al. Hydroxyethyl starch or saline for fluid resuscitation in intensive care. New England Journal of Medicine. 2012; 367(20):1901-1911

138 Myers RP, Pomier-Layrargues G, Kirsch R, Pollett A, Duarte-Rojo A, Wong D et al. Feasibility and diagnostic performance of the FibroScan XL probe for liver stiffness measurement in overweight and obese patients. Hepatology. 2012; 55(1):199-208

139 Narahara Y, Kanazawa H, Fukuda T, Matsushita Y, Harimoto H, Kidokoro H et al. Transjugular intrahepatic portosystemic shunt versus paracentesis plus albumin in patients with refractory ascites who have good hepatic and renal function: a prospective randomized trial. Journal of Gastroenterology. 2011; 46(1):78-85

140 National Clinical Guideline Centre. Gastrointestinal bleeding: the management of acute upper gastrointestinal bleeding. NICE clinical guideline 141. London. National Clinical Guideline Centre, 2012. Available from: http://guidance.nice.org.uk/CG141

141 National Clinical Guideline Centre. Hepatitis B (chronic): diagnosis and management of chronic hepatitis B in children, young people and adults. NICE clinical guideline 165. London. National Clinical Guideline Centre, 2013. Available from: http://guidance.nice.org.uk/CG165

142 National Clinical Guideline Centre. Non-alcoholic fatty liver disease: assessment and management of non-alcoholic fatty liver disease (NAFLD) in adults, children and young people. NICE clinical guideline 49. London. National Clinical Guideline Centre, 2016. Available from: http://www.nice.org.uk/guidance/NG49

143 National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

144 National Institute for Health and Clinical Excellence. Social value judgements: principles for the development of NICE guidance. 2nd edition. London: National Institute for Health and Clinical Excellence; 2008. Available from: http://www.nice.org.uk/media/C18/30/SVJ2PUBLICATION2008.pdf

145 National Institute for Health and Clinical Excellence. The guidelines manual. London: National Institute for Health and Clinical Excellence; 2012. Available from: http://publications.nice.org.uk/the-guidelines-manual-pmg6/

146 National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal 2013. 2nd edition. London: National Institute for Health and Clinical Excellence; 2013. Available from: http://publications.nice.org.uk/pmg9

147 Ng FH, Wong SY, Loo CK, Lam KM, Lai CW, Cheng CS. Prediction of oesophagogastric varices in patients with liver cirrhosis. Journal of Gastroenterology and Hepatology. 1999; 14(8):785-790

148 NHS Choices. Cirrhosis. 2012. Available from: http://www.nhs.uk/conditions/cirrhosis/pages/introduction.aspx [Last accessed: 24 September 2015]

149 Norberto L, Polese L, Cillo U, Grigoletto F, Burroughs A, Neri D et al. A randomized study comparing ligation with propranolol for primary prophylaxis of variceal bleeding in candidates for liver transplantation. Liver Transplantation. 2007; 13(9):1272-1278

150 Novielli N, Cooper NJ, Abrams KR, Sutton AJ. How is evidence on test performance synthesized for economic decision models of diagnostic tests? A systematic appraisal of Health Technology Assessments in the UK since 1997. Value in Health. 2010; 13(8):952-957

151 Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2014. Available from: http://www.oecd.org/std/ppp [Last accessed: 15 January 2014]

152 Ortega R, Gines P, Uriz J, Cardenas A, Calahorra B, De Las HD et al. Terlipressin therapy with and without albumin for patients with hepatorenal syndrome: results of a prospective, nonrandomized study. Hepatology. 2002; 36(4 Pt 1):941-948

153 Pagliaro L, D'Amico G, Pasta L, Filippazzo G, Manenti F, Dobrilla G et al. Propranolol prevents first gastrointestinal bleeding in non-ascitic cirrhotic patients. Final report of a multicenter randomized trial. Journal of Hepatology. 1989; 9(1):75-83

154 Pagliaro L, D'Amico G, Pasta L, Filippazzo MG, Manenti F, Dobrilla G et al. Propranolol for prophylaxis of bleeding in cirrhotic patients with large varices: A multicenter, randomized clinical trial. Hepatology. 1988; 8(1):1-5

155 Pagliaro L, Pasta L, D'Amico G. A randomised controlled trial of propranolol for the prevention of initial bleeding in cirrhotic patients with portal hypertension. Preliminary results. Drugs. 1989; 37(Suppl.2):48-6

156 Paik YH, Lee KS, Han KH, Song KH, Kim MH, Moon BS et al. Comparison of rifaximin and lactulose for the treatment of hepatic encephalopathy: a prospective randomized study. Yonsei Medical Journal. 2005; 46(3):399-407

157 Parker MJ, Guha N, Stedman B, Hacking N, Wright M. Trans-jugular intrahepatic porto-systemic shunt placement for refractory ascites: a 'real-world' UK health economic evaluation. Frontline Gastroenterology. 2013; 4(3):182-186

158 Pascal JP, Cales P. Propranolol in the prevention of first upper gastrointestinal tract hemorrhage in patients with cirrhosis of the liver and esophageal varices. New England Journal of Medicine. 1987; 317(14):856-861

159 Pascal JP, Cales P. Propranolol in the primary prevention of upper gastrointestinal tract haemorrhage in patients with cirrhosis of the liver and oesophageal varices. Drugs. 1989; 37(Suppl.2):52-56

160 Pascual S, Irurzun J, Zapater P, Such J, Sempere L, Carnicer F et al. Usefulness of surveillance programmes for early diagnosis of hepatocellular carcinoma in clinical practice. Liver International. 2008; 28(5):682-689

161 Perez-Ayuso RM, Valderrama S, Espinoza M, Rollan A, Sanchez R, Otarola F et al. Endoscopic band ligation versus propranolol for the primary prophylaxis of variceal bleeding in cirrhotic patients with high risk esophageal varices. Annals of Hepatology. 2010; 9(1):15-22

162 Perez-Latorre L, Sanchez-Conde M, Rincon D, Miralles P, Aldamiz-Echevarria T, Carrero A et al. Prediction of liver complications in patients with hepatitis C virus-related cirrhosis with and without HIV coinfection: comparison of hepatic venous pressure gradient and transient elastography. Clinical Infectious Diseases. 2014; 58(5):713-718

163 Perner A, Haase N, Guttormsen AB, Tenhunen J, Klemenzson G, Aneman A et al. Hydroxyethyl starch 130/0.42 versus Ringer's acetate in severe sepsis. New England Journal of Medicine. 2012; 367(2):124-134

164 Piscaglia F, Salvatore V, Di Donato R, D'Onofrio M, Gualandi S, Gallotti A et al. Accuracy of VirtualTouch Acoustic Radiation Force Impulse (ARFI) imaging for the diagnosis of cirrhosis during liver ultrasonography. Ultraschall in Der Medizin. 2011; 32(2):167-175

165 Planas R, Montoliu S, Balleste B, Rivera M, Miquel M, Masnou H et al. Natural history of patients hospitalized for management of cirrhotic ascites. Clinical Gastroenterology and Hepatology. 2006; 4(11):1385-1394

166 Psilopoulos D, Galanis P, Goulas S, Papanikolaou IS, Elefsiniotis I, Liatsos C et al. Endoscopic variceal ligation vs. propranolol for prevention of first variceal bleeding: a randomized controlled trial. European Journal of Gastroenterology and Hepatology. 2005; 17(10):1111-1117

167 Pugh RN, Murray-Lyon I, Dawson JL, Pietroni MC, Williams R. Transection of the oesophagus for bleeding oesophageal varices. British Journal of Surgery. 1973; 60(8):646-649

168 Qamar AA, Grace N, Groszmann R, Garcia-Tsao G, Bosch J, Burroughs A et al. Platelet count is not a predictor of the presence or development of gastroesophageal varices in cirrhosis. Hepatology. 2008; 47(1):153-159

169 Rahimi RS, Singal AG, Cuthbert JA, Rockey DC. Lactulose vs polyethylene glycol 3350--electrolyte solution for treatment of overt hepatic encephalopathy: the HELP randomized clinical trial. JAMA Internal Medicine. 2014; 174(11):1727-1733

170 Regev A, Berho M, Jeffers LJ, Milikowski C, Molina EG, Pyrsopoulos NT et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. American Journal of Gastroenterology. 2002; 97(10):2614-2618

171 Reinhart K, Perner A, Sprung CL, Jaeschke R, Schortgen F, Johan Groeneveld AB et al. Consensus statement of the ESICM task force on colloid volume therapy in critically ill patients. Intensive Care Medicine. 2012; 38(3):368-383

172 Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt P, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinical Epidemiology. 2005; 58(10):982-990

173 Rizzo L, Calvaruso V, Cacopardo B, Alessi N, Attanasio M, Petta S et al. Comparison of transient elastography and acoustic radiation force impulse for non-invasive staging of liver fibrosis in patients with chronic hepatitis C. American Journal of Gastroenterology. 2011; 106(12):2112-2120

174 Robic MA, Procopet B, Metivier S, Peron JM, Selves J, Vinel JP et al. Liver stiffness accurately predicts portal hypertension related complications in patients with chronic liver disease: a prospective study. Journal of Hepatology. 2011; 55(5):1017-1024

175 Rolachon A, Cordier L, Bacq Y, Nousbaum JB, Franza A, Paris JC et al. Ciprofloxacin and long-term prevention of spontaneous bacterial peritonitis: results of a prospective controlled trial. Hepatology. 1995; 22(4 Pt.1):1171-1174

176 Rossi-Fanelli F, Cangiano C, Capocaccia L, Cascino A, Ceci F, Muscaritoli M et al. Use of branched chain amino acids for treating hepatic encephalopathy: clinical experiences. Gut. 1986; 27(Suppl.1):111-115

177 Rossi-Fanelli F, Cangiano C, Cascino A, Merli M, Riggio O, Stortoni M. Branched-chain amino acids in the treatment of severe hepatic encephalopathy. In: Capocaccia L, Fischer JE, Rossi-Fanelli F (eds), Hepatic encephalopathy in chronic liver failure, New York: Plenum Press, 1984: 335-344

178 Rossi-Fanelli F, Riggio O, Cangiano C, Cascino A, De Conciliis D, Merli M et al. Branched-chain amino acids vs lactulose in the treatment of hepatic coma: a controlled study. Digestive Diseases and Sciences. 1982; 27(10):929-935

179 Rossle M, Ochs A, Gulberg V, Siegerstetter V, Holl J, Deibert P et al. A comparison of paracentesis and transjugular intrahepatic portosystemic shunting in patients with ascites. New England Journal of Medicine. 2000; 342(23):1701-1707

180 Runyon BA. Introduction to the revised American Association for the Study of Liver Diseases Practice Guideline management of adult patients with ascites due to cirrhosis 2012. Hepatology. 2013; 57(4):1651-1653

181 Saab S, Nieto J, Lewis SK, Runyon BA. TIPS versus paracentesis for cirrhotic patients with refractory ascites. Cochrane Database of Systematic Reviews. 2006;(4):CD004889. DOI:10.1002/14651858.CD004889.pub2

182 Sabat M, Kolle L, Soriano G, Ortiz J, Pamplona J, Novella M et al. Parenteral antibiotic prophylaxis of bacterial infections does not improve cost-efficacy of oral norfloxacin in cirrhotic patients with gastrointestinal bleeding. American Journal of Gastroenterology. 1998; 93(12):2457-2462

183 Said A, Williams J, Holden J, Remington P, Gangnon R, Musat A et al. Model for end stage liver disease score predicts mortality across a broad spectrum of liver disease. Journal of Hepatology. 2004; 40(6):897-903

184 Salerno F, Gerbes A, Gines P, Wong F, Arroyo V. Diagnosis, prevention and treatment of hepatorenal syndrome in cirrhosis. Gut. 2007; 56(9):1310-1318

185 Salerno F, Merli M, Riggio O, Cazzaniga M, Valeriano V, Pozzi M et al. Randomized controlled study of TIPS versus paracentesis plus albumin in cirrhosis with severe ascites. Hepatology. 2004; 40(3):629-635

186 Sanchez-Conde M, Montes-Ramirez ML, Miralles P, Alvarez JMC, Bellon JM, Ramirez M et al. Comparison of transient elastography and liver biopsy for the assessment of liver fibrosis in HIV/hepatitis C virus-coinfected patients and correlation with noninvasive serum markers. Journal of Viral Hepatitis. 2010; 17(4):280-286

187 Sangiovanni A, Del Ninno E, Fasani P, De Fazio C, Ronchi G, Romeo R et al. Increased survival of cirrhotic patients with a hepatocellular carcinoma detected during surveillance. Gastroenterology. 2004; 126(4):1005-1014

188 Santi V, Trevisani F, Gramenzi A, Grignaschi A, Mirici-Cappa F, Del Poggio P et al. Semiannual surveillance is superior to annual surveillance for the detection of early hepatocellular carcinoma and patient survival. Journal of Hepatology. 2010; 53(2):291-297

189 Santos J, Planas R, Pardo A, Durandez R, Cabre E, Morillas R et al. Spironolactone alone or in combination with furosemide in the treatment of moderate ascites in nonazotemic cirrhosis. A randomized comparative study of efficacy and safety. Journal of Hepatology. 2003; 39(2):187-192

190 Sanyal AJ, Gennings C, Reddy KR, Wong F, Kowdley K, Benner K et al. The North American study for the treatment of refractory ascites. Gastroenterology. 2003; 124(3):634-641

191 Sarin SK, Guptan RK, Jain AK, Sundaram KR. A randomized controlled trial of endoscopic variceal band ligation for primary prophylaxis of variceal bleeding. European Journal of Gastroenterology and Hepatology. 1996; 8(4):337-342

192 Sarin SK, Lamba GS, Kumar M, Misra A, Murthy NS. Comparison of endoscopic ligation and propranolol for the primary prevention of variceal bleeding. New England Journal of Medicine. 1999; 340(13):988-993

193 Sarin SK, Mishra SR, Sharma P, Sharma BC, Kumar A. Early primary prophylaxis with beta-blockers does not prevent the growth of small esophageal varices in cirrhosis: A randomized controlled trial. Hepatology International. 2013; 7(1):248-256

194 Sarin S, Lamba G, Kumar M, Misra A, Murthy N. Randomized trial of propranolol vs endoscopic variceal ligation in the primary prophylaxis of bleeding from high risk varices in cirrhosis: an interim analysis. Hepatology. 1997; 26(4 Pt.2):360A

195 Schcpka M, Goebel C, Nuernberg D, Willert J, Koch L, Sauerbruch T. Endoscopic banding ligation versus propranolol for the primary prevention of variceal bleeding in cirrhosis: a randomized controlled multicenter trial. Hepatology. 2003; 38(4 S1):218

196 Schepke M, Kleber G, Nuernberg D, Willert J, Koch L, Veltzke-Schlieker W et al. Ligation versus propranolol for the primary prophylaxis of variceal bleeding in cirrhosis. Hepatology. 2004; 40(1):65-72

197 Schult A, Eriksson H, Wallerstedt S, Kaczynski J. Overweight and hypertriglyceridemia are risk factors for liver cirrhosis in middle-aged Swedish men. Scandinavian Journal of Gastroenterology. 2011; 46(6):738-744

198 Shah HA, Azam Z, Rauf J, Abid S, Hamid S, Jafri W et al. Carvedilol vs. esophageal variceal band ligation in the primary prophylaxis of variceal hemorrhage: a multicentre randomized controlled trial. Journal of Hepatology. 2014; 60(4):757-764

199 Sharma BC, Sharma P, Lunia MK, Srivastava S, Goyal R, Sarin SK. A randomized, double-blind, controlled trial comparing rifaximin plus lactulose with lactulose alone in treatment of overt hepatic encephalopathy. American Journal of Gastroenterology. 2013; 108(9):1458-1463

200 Shehab H, Elattar I, Elbaz T, Mohey M, Esmat G. CUFA algorithm: assessment of liver fibrosis using routine laboratory data. Journal of Viral Hepatitis. 2014; 21(12):956-964

201 Sigal SH, Stanca CM, Fernandez J, Arroyo V, Navasa M. Restricted use of albumin for spontaneous bacterial peritonitis. Gut. 2007; 56(4):597-599

202 Silva Junior RG, Schmillevitch J, Nascimento MdFA, Miranda ML, Brant PEAC, Schulz PO et al. Acoustic radiation force impulse elastography and serum fibrosis markers in chronic hepatitis C. Scandinavian Journal of Gastroenterology. 2014; 49(8):986-992

203 Singh B, Saxena PD, Rohtagi V, Kumar V. Comparison of endoscopic variceal ligation and propranolol for the primary prevention of variceal bleeding. Journal, Indian Academy of Clinical Medicine. 2012; 13(3):214-217

204 Singh N, Gayowski T, Yu VL, Wagener MM. Trimethoprim-sulfamethoxazole for the prevention of spontaneous bacterial peritonitis in cirrhosis: a randomized trial. Annals of Internal Medicine. 1995; 122(8):595-598

205 Sirli R, Sporea I, Bota S, Popescu A, Cornianu M. A comparative study of non-invasive methods for fibrosis assessment in chronic HCV infection. Hepatitis Monthly. 2010; 10(2):88-94

206 Song IH, Shin JW, Kim IH, Choi J, Lim CY, Kim JW et al. A prospective randomized trial between the prophylactic endoscopicvariceal ligation and propranolol administration for prevention offirst bleeding in cirrhotic patients with high-risk esophageal varices. Journal of Hepatology. 2000; 32(Suppl.2):41

207 Sorensen TIA. Prophylaxis of first hemorrhage from esophageal varices by sclerotherapy, propranolol or both in cirrhotic patients: A randomized multicenter trial. Hepatology. 1991; 14(6):1016-1024

208 Soriano G, Guarner C, Teixido M, Such J, Barrios J, Enriquez J et al. Selective intestinal decontamination prevents spontaneous bacterial peritonitis. Gastroenterology. 1991; 100(2):477-481

209 Spanish Group for the Study of Bacterial Infections in Cirrhosis. Norfloxacin versus ofloxacin in the prophylaxis of infection in cirrhotic patients with gastrointestinal hemorrhage. Journal of Hepatology. 1998; 28(Suppl.1):80

210 Sporea I, Sirli RL, Deleanu A, Popescu A, Focsa M, Danila M et al. Acoustic radiation force impulse elastography as compared to transient elastography and liver biopsy in patients with chronic hepatopathies. Ultraschall in Der Medizin. 2011; 32(Suppl.1):S46-S52

211 Sporea I, Bota S, Peck-Radosavljevic M, Sirli R, Tanaka H, Iijima H et al. Acoustic Radiation Force Impulse elastography for fibrosis evaluation in patients with chronic hepatitis C: an international multicenter study. European Journal of Radiology. 2012; 81(12):4112-4118

212 Sporea I, Sirli R, Popescu A, Bota S, Badea R, Lupsor M et al. Is it better to use two elastographic methods for liver fibrosis assessment? World Journal of Gastroenterology. 2011; 17(33):3824-3829

213 Steadman R, Myers RP, Leggett L, Lorenzetti D, Noseworthy T, Rose S et al. A health technology assessment of transient elastography in adult liver disease. Canadian Journal of Gastroenterology. 2013; 27(3):149-158

214 Stevenson M, Lloyd-Jones M, Morgan MY, Wong R. Non-invasive diagnostic assessment tools for the detection of liver fibrosis in patients with suspected alcohol-related liver disease: a systematic review and economic evaluation. Health Technology Assessment. 2012; 16(4):1-174

215 Stibbe KJM, Verveer C, Francke J, Hansen BE, Zondervan PE, Kuipers EJ et al. Comparison of non-invasive assessment to diagnose liver fibrosis in chronic hepatitis B and C patients. Scandinavian Journal of Gastroenterology. 2011; 46(7-8):962-972

216 Strauss E, da Costa MF. The importance of bacterial infections as precipating factors of chronic hepatic encephalopathy in cirrhosis. Hepato-Gastroenterology. 1998; 45(21):900-904

217 Strauss E, dos Santos WR, Silva EC, Lacet CM, Capacci MLL, Bernardini AP. Treatment of hepatic encephalopathy: A randomizes clinical trial comparing a branched chain enriched amino acid solution to oral neomycin. Nutritional Support Services. 1986; 6(7):18-21

218 Strauss E, Tramote R, Silva EP, Caly WR, Honain NZ, Maffei RA et al. Double-blind randomized clinical trial comparing neomycin and placebo in the treatment of exogenous hepatic encephalopathy. Hepato-Gastroenterology. 1992; 39(6):542-545

219 Stroffolini T, Trevisani F, Pinzello G, Brunello F, Tommasini MA, Iavarone M et al. Changing aetiological factors of hepatocellular carcinoma and their potential impact on the effectiveness of surveillance. Digestive and Liver Disease. 2011; 43(11):875-880

220 Sushma S, Dasarathy S, Tandon RK, Jain S, Gupta S, Bhist MS. Sodium benzoate in the treatment of acute hepatic encephalopathy: a double-blind randomized trial. Hepatology. 1992; 16(1):138-144

221 Svoboda P, Kantorova I, Ochmann J, Kozumplik L, Marsova J. A prospective randomized controlled trial of sclerotherapy vs ligation in the prophylactic treatment of high-risk esophageal varices. Surgical Endoscopy. 1999; 13(6):580-584

222 Tandon P, Garcia-Tsao G. Bacterial infections, sepsis, and multiorgan failure in cirrhosis. Seminars in Liver Disease. 2008; 28(1):26-42

223 Tellez-Avila F, Sifuentes-Osornio J, Barbero-Becerra V, Franco-Guzman A, Ruiz-Cordero R, Alfaro-Lara R et al. Primary prophylaxis with ciprofloxacin in cirrhotic patients with ascites: a randomized, double blind study. Annals of Hepatology. 2013; 13(1):65-74

224 Terg R, Fassio E, Guevara M, Cartier M, Longo C, Lucero R et al. Ciprofloxacin in primary prophylaxis of spontaneous bacterial peritonitis: a randomized, placebo-controlled study. Journal of Hepatology. 2008; 48(5):774-779

225 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 Technology Assessment. 2007; 11(34):1-206

226 Thompson-Coon J, Rogers G, Hewson P, Wright D, Anderson R, Jackson S et al. Surveillance of cirrhosis for hepatocellular carcinoma: a cost utility analysis. British Journal of Cancer. 2008; 98(7):1166-1175

227 Thuluvath PJ, Maheshwari A, Jagannath S, Arepally A. A randomized controlled trial of beta-blockers versus endoscopic band ligation for primary prophylaxis: a large sample size is required to show a difference in bleeding rates. Digestive Diseases and Sciences. 2005; 50(2):407-410

228 Trevisani F, Cantarini MC, Labate AM, De Notariis S, Rapaccini G, Farinati F et al. Surveillance for hepatocellular carcinoma in elderly Italian patients with cirrhosis: effects on cancer staging and patient survival. American Journal of Gastroenterology. 2004; 99(8):1470-1476

229 Trevisani F, De Notariis S, Rapaccini G, Farinati F, Benvegnu L, Zoli M et al. Semiannual and annual surveillance of cirrhotic patients for hepatocellular carcinoma: effects on cancer stage and patient survival (Italian experience). American Journal of Gastroenterology. 2002; 97(3):734-744

230 Trevisani F, Magini G, Santi V, Morselli-Labate AM, Cantarini MC, Di Nolfo MA et al. Impact of etiology of cirrhosis on the survival of patients diagnosed with hepatocellular carcinoma during surveillance. American Journal of Gastroenterology. 2007; 102(5):1022-1031

231 Trevisani F, Santi V, Gramenzi A, Di Nolfo MA, Del Poggio P, Benvegnu L et al. Surveillance for early diagnosis of hepatocellular carcinoma: is it effective in intermediate/advanced cirrhosis? American Journal of Gastroenterology. 2007; 102(11):2448-2457

232 Triantos C, Vlachogiannakos J, Armonis A, Saveriadis A, Kougioumtzian A, Leandro G et al. Primary prophylaxis of variceal bleeding in cirrhotics unable to take beta-blockers: a randomized trial of ligation. Alimentary Pharmacology and Therapeutics. 2005; 21(12):1435-1443

233 Trinchet JC, Chaffaut C, Bourcier V, Degos F, Henrion J, Fontaine H et al. Ultrasonographic surveillance of hepatocellular carcinoma in cirrhosis: a randomized trial comparing 3- and 6-month periodicities. Hepatology. 2011; 54(6):1987-1997

234 Tripathi D, Ferguson J, Kochar N, Leithead JA, Therapondos G, McAvoy NC et al. Randomized controlled trial of carvedilol versus variceal band ligation for the prevention of the first variceal bleed. Hepatology. 2009; 50(3):825-833

235 Tsochatzis EA, Bosch J, Burroughs A. Liver cirrhosis. The Lancet. 2014; 383(9930):1749-1761

236 Uribe M, Berthier JM, Lewis H, Mata JM, Sierra JG, Garcia-Ramos G et al. Lactose enemas plus placebo tablets vs. neomycin tablets plus starch enemas in acute portal systemic encephalopathy. A double-blind randomized controlled study. Gastroenterology. 1981; 81(1):101-106

237 Uribe M, Campollo O, Vargas F, Ravelli GP, Mundo F, Zapata L et al. Acidifying enemas (lactitol and lactose) vs. nonacidifying enemas (tap water) to treat acute portal-systemic encephalopathy: a double-blind, randomized clinical trial. Hepatology. 1987; 7(4):639-643

238 Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine. 2002; 21(4):589-624

239 Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. Statistics in Medicine. 1993; 12(24):2273-2284

240 Vilstrup H, Gluud C, Hardt F, Kristensen M, Kohler O, Melgaard B et al. Branched chain enriched amino acid versus glucose treatment of hepatic encephalopathy. A double-blind study of 65 patients with cirrhosis. Journal of Hepatology. 1990; 10(3):291-296

241 Wahren J, Denis J, Desurmont P, Eriksson LS, Escoffier JM, Gauthier AP et al. Is intravenous administration of branched chain amino acids effective in the treatment of hepatic encephalopathy? A multicenter study. Hepatology. 1983; 3(4):475-480

242 Wang JH, Chuah SK, Lu SN, Hung CH, Kuo CM, Tai WC et al. Baseline and serial liver stiffness measurement in prediction of portal hypertension progression for patients with compensated cirrhosis. Liver International. 2014; 34(9):1340-1348

243 Wong VW-S, Vergniol J, Wong GL-H, Foucher J, Chan AWH, Chermak F et al. Liver stiffness measurement using XL probe in patients with nonalcoholic fatty liver disease. American Journal of Gastroenterology. 2012; 107(12):1862-1871

244 Wong VW-S, Vergniol J, Wong GL-H, Foucher J, Chan HLY, Le Bail B et al. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. Hepatology. 2010; 51(2):454-462

245 Yamada H, Ebara M, Yamaguchi T, Okabe S, Fukuda H, Yoshikawa M et al. A pilot approach for quantitative assessment of liver fibrosis using ultrasound: preliminary results in 79 cases. Journal of Hepatology. 2006; 44(1):68-75

246 Yang Z, Han G, Wu Q, Ye X, Jin Z, Yin Z et al. Patency and clinical outcomes of transjugular intrahepatic portosystemic shunt with polytetrafluoroethylene-covered stents versus bare sents: a meta-analysis. Journal of Gastroenterology and Hepatology. 2010; 25(11):1718-1725

247 Yoneda M, Yoneda M, Mawatari H, Fujita K, Endo H, Iida H et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with nonalcoholic fatty liver disease (NAFLD). Digestive and Liver Disease. 2008; 40(5):371-378

248 Yoneda M, Suzuki K, Kato S, Fujita K, Nozaki Y, Hosono K et al. Nonalcoholic fatty liver disease: US-based acoustic radiation force impulse elastography. Radiology. 2010; 256(2):640-647

249 Zarski JP, Sturm N, Guechot J, Paris A, Zafrani ES, Asselah T et al. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: the ANRS HCEP-23 study. Journal of Hepatology. 2012; 56(1):55-62

17 Acronyms and abbreviations

Acronym or abbreviation	Description
AFP	Alpha-fetoprotein
ALD	Alcohol-related liver disease
ALT	Alanine transaminase
APRI	AST to platelet ratio index
ARD	Absolute risk difference
ARFI	Acoustic radiation force impulse imaging
AST	Aspartate transaminase
AUC	Area under the curve
BCAA	Branch chain amino acids
BCLC	Barcelona Clinic Liver Cancer
BMI	Body Mass Index
BNF	British National Formulary
CEAC	Cost-effectiveness acceptability curve
CHB	Chronic hepatitis B
CHC	Chronic hepatitis C
CI	Confidence intervals
CLD	Chronic liver disease
ED	Emergency department
EEG	Electroencephalography
ELF	Enhanced liver fibrosis
EQ-5D	EuroQol 5-dimension
EVL	Endoscopic variceal ligation
FN	False negatives
FP	False positives
GDG	Guideline development group
GI	Gastrointestinal
GRADE	Grading of recommendations assessment, development and evaluation
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HR	Hazard ratio
HRQoL	Health-related quality of life
HVPG	Hepatovenous portal pressure gradient
ICER	Incremental cost-effectiveness ratio
IV	Intravenous
LETR	Linking evidence to recommendations
LVP	Large-volume paracentesis
MELD	Model for End-Stage Liver Disease
MID	Minimally important differences
Na	Sodium
NAFLD	Non-alcoholic fatty liver disease

Acronym or abbreviation	Description
NASH	Non-alcoholic steatohepatitis
NGC	National Guideline Centre
NCT	Number connection test
NICE	National Institute for Health and Care Excellence
NPV	Negative predictive value
OECD	Organisation for Economic Co-operation and Development
OGD	Oesophagogastroduodenoscopy
OR	Odds ratio
PBC	Primary biliary cholangitis
PICO	Population Intervention Comparison Outcomes
PPV	Positive predictive value
PROBAST	Prediction study risk of bias assessment tool
PSC	Primary sclerosing cholangitis
PSE	Portal systemic encephalopathy
QALYS	Quality-adjusted life years
RCT	Randomised controlled trial
ROC	Receiver operating characteristic
RR	Risk ratio (relative risk)
SBP	Spontaneous bacterial peritonitis
SEIQOL-DW	Schedule for the evaluation of individual quality of life-direct weighting
SF-36	36-Item Short Form Health Survey
TE	Transient elastography
TIPS	Transjugular intrahepatic portosystemic shunt
TN	True negatives
TP	True positives

18 Glossary

The NICE Glossary can be found at www.nice.org.uk/glossary.

18.1 Guideline-specific terms

Term	Definition
Acoustic radiation force impulse (ARFI) imaging	An ultrasound-based elastography method enabling quantitative measurement of tissue stiffness to be made.
Alcohol-related liver disease (ALD)	Damage to the liver caused by excessive alcohol consumption.
Ascites	Accumulation of fluid in the abdominal cavity.
Child-Pugh score	A scoring system which employs 3 clinical and 2 laboratory variables and is used to classify the severity of hepatic dysfunction in people with cirrhosis and to assess the prognosis.
Chronic liver disease	Disease of the liver which has lasted for over 6 months. It consists of a range of liver pathologies which include inflammation (chronic hepatitis) and cirrhosis.
Compensated cirrhosis	Cirrhosis with preserved functional capacity and no evident complications of either portal hypertension or hepatocellular dysfunction.
Decompensated cirrhosis	Cirrhosis with impaired functional capacity.
Diuretic-resistant ascites	Ascites that is no longer responsive to diuretics.
Endoscopic variceal band ligation	A procedure in which an enlarged vein or a varix in the oesophagus is tied off or ligated by a rubber band delivered via an endoscope.
Enhanced liver fibrosis (ELF) test	A minimally-invasive blood test for measuring liver fibrosis.
Fibrosis	Scar tissue which form in the liver following inflammation and cell death. Fibrosis can take a variable time to develop and, even with scar tissue present, the liver can continue to function well. However, continued build-up of scar tissue following further episodes of persistence of inflammation and cell death may eventually result in the development of cirrhosis.
Harmful drinking	A pattern of drinking which is likely to cause physical or psychological harm. Men who drink more than 50 units in the course of a week are classified as harmful drinkers, as are women who consume over 35 units.
Hazardous drinking	A pattern of drinking which brings about the risk of physical or psychological harm. Men who drink more than 21 units in the course of a week are classified as hazardous drinkers, as are women who consume over 14 units.
Hepatic encephalopathy	A syndrome of neuropsychiatric changes which arises as a complication of liver disease.
Hepatitis	Inflammation of the liver, caused by infectious or toxic agents. Inflammation arising over a period of days or weeks is termed acute hepatitis while inflammation lasting for 6 or more months is termed chronic hepatitis.
Hepatocellular carcinoma (HCC)	The most common type of malignant tumour of the liver. It develops most frequently in the UK as a long-term complication of cirrhosis but can also develop in people with chronic hepatitis B or chronic hepatitis C infection who do not have cirrhosis; very occasionally it can develop in people with other types of non-cirrhotic liver disease.
Hepatology	The study, prevention, diagnosis and management of diseases that affect the liver, gallbladder, biliary tree and pancreas.

Term	Definition
Hepatorenal syndrome	One of several causes of acute renal failure in patients with advanced cirrhosis with portal hypertension which is associated with a particularly poor prognosis.
Large-volume paracentesis (LVP)	The removal of large volumes of ascitic fluid via a drainage catheter inserted through the abdominal wall.
Liver biopsy	A diagnostic test in which a sample of tissue is removed from the liver most commonly via a needle inserted through the abdominal wall.
Model for End-Stage Liver Disease (MELD) score	A scoring system for assessing the severity of chronic liver disease.
Non-alcoholic fatty liver disease (NAFLD)	Excess fat in the liver (steatosis) in the absence of excessive alcohol consumption or any of the other secondary causes of steatosis.
Primary biliary cholangitis	A long-term liver disease in which the bile ducts in the liver become damaged.
Primary biliary cirrhosis	See primary biliary cholangitis.
Refractory ascites	Ascites that cannot be mobilised or the early recurrence of which cannot be prevented by medical therapy.
Sclerotherapy	A method of treating oesophageal varices. The vein is injected with an irritant solution, which causes inflammation in the vessel lining, leading to scar tissue formation and the obliteration of the vein.
Specialist hepatology service	Provides additional interventions for the complications of cirrhosis (for example TIPS), as well as assessment for liver transplantation.
Spontaneous bacterial peritonitis (SBP)	Bacterial infection of the ascitic fluid.
Transient elastography (TE)	A non-invasive test for the assessment of liver fibrosis.
Transjugular intrahepatic portosystemic shunt (TIPS or TIPPS)	A procedure in which a shunt (tube) is placed between the portal and hepatic veins in order to reduce the pressure in the portal venous system.
Varices	Dilated blood vessels in the oesophagus or stomach which arise as a result of portal hypertension.
Variceal bleeding	Blood loss from a ruptured variceal vein.
Vasoactive	Medication that affects the diameter of blood vessels (and hence blood pressure).

18.2 General terms

Term	Definition
Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper.
Algorithm (in guidelines)	A flow chart of the clinical decision pathway described in the guideline, where decision points are represented with boxes, linked with arrows.
Allocation concealment	The process used to prevent advance knowledge of group assignment in an RCT. The allocation process should be impervious to any influence by the individual making the allocation, by being administered by someone who is not responsible for recruiting participants.
Applicability	How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.
Arm (of a clinical study)	Subsection of individuals within a study who receive one particular intervention, for example placebo arm.
Association	Statistical relationship between 2 or more events, characteristics or other variables. The relationship may or may not be causal.

Term	Definition
Base case analysis	In an economic evaluation, this is the main analysis based on the most plausible estimate of each input. In contrast, see Sensitivity analysis.
Baseline	The initial set of measurements at the beginning of a study (after run-in period where applicable), with which subsequent results are compared.
Bayesian analysis	A method of statistics, where a statistic is estimated by combining established information or belief (the 'prior') with new evidence (the 'likelihood') to give a revised estimate (the 'posterior').
Before-and-after study	A study that investigates the effects of an intervention by measuring particular characteristics of a population both before and after taking the intervention, and assessing any change that occurs.
Bias	Influences on a study that can make the results look better or worse than they really are. (Bias can even make it look as if a treatment works when it does not.) Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.
Blinding	A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot influence the results. The best way to do this is by sorting patients into study groups randomly. The purpose of 'blinding' or 'masking' is to protect against bias. A single-blinded study is one in which patients do not know which study group they are in (for example whether they are taking the experimental drug or a placebo). A double-blinded study is one in which neither patients nor the researchers and doctors know which study group the patients are in. A triple blind study is one in which neither the patients, clinicians or the people carrying out the statistical analysis know which treatment patients received.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Case–control study	A study to find out the cause(s) of a disease or condition. This is done by comparing a group of patients who have the disease or condition (cases) with a group of people who do not have it (controls) but who are otherwise as similar as possible (in characteristics thought to be unrelated to the causes of the disease or condition). This means the researcher can look for aspects of their lives that differ to see if they may cause the condition. For example, a group of people with lung cancer might be compared with a group of people the same age that do not have lung cancer. The researcher could compare how long both groups had been exposed to tobacco smoke. Such studies are retrospective because they look back in time from the outcome to the possible causes of a disease or condition.
Case series	Report of a number of cases of a given disease, usually covering the course of the disease and the response to treatment. There is no comparison (control) group of patients.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.
Clinical effectiveness	How well a specific test or treatment works when used in the 'real world' (for example, when used by a doctor with a patient at home), rather than in a carefully controlled clinical trial. Trials that assess clinical effectiveness are sometimes called management trials. Clinical effectiveness is not the same as efficacy.

Term	Definition
Clinician	A healthcare professional who provides patient care. For example, a doctor, nurse or physiotherapist.
Cochrane Review	The Cochrane Library consists of a regularly updated collection of evidence-based medicine databases including the Cochrane Database of Systematic Reviews (reviews of randomised controlled trials prepared by the Cochrane Collaboration).
Cohort study	A study with 2 or more groups of people – cohorts – with similar characteristics. One group receives a treatment, is exposed to a risk factor or has a particular symptom and the other group does not. The study follows their progress over time and records what happens. See also observational study.
Comorbidity	A disease or condition that someone has in addition to the health problem being studied or treated.
Comparability	Similarity of the groups in characteristics likely to affect the study results (such as health status or age).
Concordance	This is a recent term whose meaning has changed. It was initially applied to the consultation process in which doctor and patient agree therapeutic decisions that incorporate their respective views, but now includes patient support in medicine taking as well as prescribing communication. Concordance reflects social values but does not address medicine-taking and may not lead to improved adherence.
Confidence interval (CI)	There is always some uncertainty in research. This is because a small group of patients is studied to predict the effects of a treatment on the wider population. The confidence interval is a way of expressing how certain we are about the findings from a study, using statistics. It gives a range of results that is likely to include the 'true' value for the population. The CI is usually stated as '95% CI', which means that the range of values has a 95 in a 100 chance of including the 'true' value. For example, a study may state that "based on our sample findings, we are 95% certain that the 'true' population blood pressure is not higher than 150 and not lower than 110". In such a case the 95% CI would be 110 to 150. A wide confidence interval indicates a lack of certainty about the true effect of the test or treatment – often because a small group of patients has been studied. A narrow confidence interval indicates a more precise estimate (for example, if a large number of patients have been studied).
Confounding factor	Something that influences a study and can result in misleading findings if it is not understood or appropriately dealt with. For example, a study of heart disease may look at a group of people that exercises regularly and a group that does not exercise. If the ages of the people in the 2 groups are different, then any difference in heart disease rates between the 2 groups could be because of age rather than exercise. Therefore age is a confounding factor.
Consensus methods	Techniques used to reach agreement on a particular issue. Consensus methods may be used to develop NICE guidance if there is not enough good quality research evidence to give a clear answer to a question. Formal consensus methods include Delphi and nominal group techniques.
Control group	A group of people in a study who do not receive the treatment or test being studied. Instead, they may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo). The results for the control group are compared with those for a group receiving the treatment being tested. The aim is to check for any differences. Ideally, the people in the control group should be as similar as possible to those in the treatment group, to make it as easy as possible to detect any

Term	Definition
	effects due to the treatment.
Cost-benefit analysis (CBA)	Cost-benefit analysis is one of the tools used to carry out an economic evaluation. The costs and benefits are measured using the same monetary units (for example, pounds sterling) to see whether the benefits exceed the costs.
Cost-consequences analysis (CCA)	Cost-consequences analysis is one of the tools used to carry out an economic evaluation. This compares the costs (such as treatment and hospital care) and the consequences (such as health outcomes) of a test or treatment with a suitable alternative. Unlike cost-benefit analysis or cost-effectiveness analysis, it does not attempt to summarise outcomes in a single measure (like the quality-adjusted life year) or in financial terms. Instead, outcomes are shown in their natural units (some of which may be monetary) and it is left to decision-makers to determine whether, overall, the treatment is worth carrying out.
Cost-effectiveness analysis (CEA)	Cost-effectiveness analysis is one of the tools used to carry out an economic evaluation. The benefits are expressed in non-monetary terms related to health, such as symptom-free days, heart attacks avoided, deaths avoided or life years gained (that is, the number of years by which life is extended as a result of the intervention).
Cost-effectiveness model	An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.
Cost-utility analysis (CUA)	Cost-utility analysis is one of the tools used to carry out an economic evaluation. The benefits are assessed in terms of both quality and duration of life, and expressed as quality-adjusted life years (QALYs). See also utility.
Credible interval (CrI)	The Bayesian equivalent of a confidence interval.
Decision analysis	An explicit quantitative approach to decision-making under uncertainty, based on evidence from research. This evidence is translated into probabilities, and then into diagrams or decision trees which direct the clinician through a succession of possible scenarios, actions and outcomes.
Deterministic analysis	In economic evaluation, this is an analysis that uses a point estimate for each input. In contrast, see Probabilistic analysis
Diagnostic odds ratio	The diagnostic odds ratio is a measure of the effectiveness of a diagnostic test. It is defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease.
Discounting	Costs and perhaps benefits incurred today have a higher value than costs and benefits occurring in the future. Discounting health benefits reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.
Disutility	The loss of quality of life associated with having a disease or condition. See Utility
Dominance	A health economics term. When comparing tests or treatments, an option that is both less effective and costs more is said to be 'dominated' by the alternative.
Drop-out	A participant who withdraws from a trial before the end.
Economic evaluation	An economic evaluation is used to assess the cost-effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits – health effects – relative to the resources available. It should be used to inform and support

Term	Definition
	the decision-making process; it is not supposed to replace the judgement of healthcare professionals. There are several types of economic evaluation: cost-benefit analysis, cost-consequences analysis, cost-effectiveness analysis, cost-minimisation analysis and cost-utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
Epidemiological study	The study of a disease within a population, defining its incidence and prevalence and examining the roles of external influences (for example, infection, diet) and interventions.
EQ-5D (EuroQol 5 dimensions)	A standardised instrument used to measure health-related quality of life. It provides a single index value for health status.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients).
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.
Exclusion criteria (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Extended dominance	If Option A is both more clinically effective than Option B and has a lower cost per unit of effect, when both are compared with a do-nothing alternative then Option A is said to have extended dominance over Option B. Option A is therefore more cost-effective and should be preferred, other things remaining equal.
Extrapolation	An assumption that the results of studies of a specific population will also hold true for another population with similar characteristics.
Follow-up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health-related variables.
Generalisability	The extent to which the results of a study hold true for groups that did not participate in the research. See also external validity.
Gold standard	A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease.
GRADE, GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.

Term	Definition
Health economics	Study or analysis of the cost of using and distributing healthcare resources.
Health-related quality of life (HRQoL)	A measure of the effects of an illness to see how it affects someone's day-to-day life.
Heterogeneity or Lack of homogeneity	The term is used in meta-analyses and systematic reviews to describe when the results of a test or treatment (or estimates of its effect) differ significantly in different studies. Such differences may occur as a result of differences in the populations studied, the outcome measures used or because of different definitions of the variables involved. It is the opposite of homogeneity.
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.
Inclusion criteria (literature review)	Explicit criteria used to decide which studies should be considered as potential sources of evidence.
Incremental analysis	The analysis of additional costs and additional clinical outcomes with different interventions.
Incremental cost	The extra cost linked to using one test or treatment rather than another. Or the additional cost of doing a test or providing a treatment more frequently.
Incremental cost-effectiveness ratio (ICER)	The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for one treatment compared with another.
Incremental net benefit (INB)	The value (usually in monetary terms) of an intervention net of its cost compared with a comparator intervention. The INB can be calculated for a given cost-effectiveness (willingness to pay) threshold. If the threshold is £20,000 per QALY gained then the INB is calculated as: (£20,000 × QALYs gained) – Incremental cost.
Indirectness	The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison and outcome).
Intention-to-treat analysis (ITT)	An assessment of the people taking part in a clinical trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully complied with the treatment or switched to an alternative treatment. Intention-to-treat analyses are often used to assess clinical effectiveness because they mirror actual practice: that is, not everyone complies with treatment and the treatment people receive may be changed according to how they respond to it.
Intervention	In medical terms this could be a drug treatment, surgical procedure, diagnostic or psychological therapy. Examples of public health interventions could include action to help someone to be physically active or to eat a more healthy diet.
Intraoperative	The period of time during a surgical procedure.
Kappa statistic	A statistical measure of inter-rater agreement that takes into account the agreement occurring by chance.
Length of stay	The total number of days a participant stays in hospital.
Licence	See 'Product licence'.
Life years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.
Likelihood ratio	The likelihood ratio combines information about the sensitivity and specificity. It tells you how much a positive or negative result changes the likelihood that a patient would have the disease. The likelihood ratio of a positive test result (LR+) is sensitivity divided by (1 minus specificity).

Term	Definition
Long-term care	Residential care in a home that may include skilled nursing care and help with everyday activities. This includes nursing homes and residential homes.
Logistic regression or Logit model	In statistics, logistic regression is a type of analysis used for predicting the outcome of a binary dependent variable based on one or more predictor variables. It can be used to estimate the log of the odds (known as the 'logit').
Loss to follow-up	A patient, or the proportion of patients, actively participating in a clinical trial at the beginning, but whom the researchers were unable to trace or contact by the point of follow-up in the trial
Markov model	A method for estimating long-term costs and effects for recurrent or chronic conditions, based on health states and the probability of transition between them within a given time period (cycle).
Meta-analysis	A method often used in systematic reviews. Results from several studies of the same test or treatment are combined to estimate the overall effect of the treatment.
Multivariate model	A statistical model for analysis of the relationship between 2 or more predictor (independent) variables and the outcome (dependent) variable.
Negative predictive value (NPV)	In screening or diagnostic tests: A measure of the usefulness of a screening or diagnostic test. It is the proportion of those with a negative test result who do not have the disease, and can be interpreted as the probability that a negative test result is correct. It is calculated as follows: TN/(TN+FN)
Net monetary benefit (NMB)	The value in monetary terms of an intervention net of its cost. The NMB can be calculated for a given cost-effectiveness threshold. If the threshold is £20,000 per QALY gained then the NMB for an intervention is calculated as: (£20,000 × mean QALYs) − mean cost. The most preferable option (that is, the most clinically effective option to have an ICER below the threshold selected) will be the treatment with the highest NMB.
Number needed to treat (NNT)	The average number of patients who need to be treated to get a positive outcome. For example, if the NNT is 4, then 4 patients would have to be treated to ensure 1 of them gets better. The closer the NNT is to 1, the better the treatment. For example, if you give a stroke prevention drug to 20 people before 1 stroke is prevented, the number needed to treat is 20. See also number needed to harm, absolute risk reduction.
Observational study	Individuals or groups are observed or certain factors are measured. No attempt is made to affect the outcome. For example, an observational study of a disease or treatment would allow 'nature' or usual medical care to take its course. Changes or differences in one characteristic (for example, whether or not people received a specific treatment or intervention) are studied without intervening. There is a greater risk of selection bias than in experimental studies.
Odds ratio	Odds are a way to represent how likely it is that something will happen (the probability). An odds ratio compares the probability of something in one group with the probability of the same thing in another. An odds ratio of 1 between 2 groups would show that the probability of the event (for example a person developing a disease, or a treatment working) is the same for both. An odds ratio greater than 1 means the event is more likely in the first group. An odds ratio less than 1 means that the event is less likely in the first group. Sometimes probability can be compared across more than 2 groups – in this case, one of the groups is chosen as the 'reference category', and the

Term	Definition
	odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, risk ratio.
Opportunity cost	The loss of other healthcare programmes displaced by investment in or introduction of another intervention. This may be best measured by the health benefits that could have been achieved had the money been spent on the next best alternative healthcare intervention.
Outcome	The impact that a test, treatment, policy, programme or other intervention has on a person, group or population. Outcomes from interventions to improve the public's health could include changes in knowledge and behaviour related to health, societal changes (for example, a reduction in crime rates) and a change in people's health and wellbeing or health status. In clinical terms, outcomes could include the number of patients who fully recover from an illness or the number of hospital admissions, and an improvement or deterioration in someone's health, functional ability, symptoms or situation. Researchers should decide what outcomes to measure before a study begins.
P value	The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing 2 treatments found that one seems more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
Perioperative	The period from admission through surgery until discharge, encompassing the preoperative and postoperative periods.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had – over and above any placebo effect caused because someone has received (or thinks they have received) care or attention.
Polypharmacy	The use or prescription of multiple medications.
Posterior distribution	In Bayesian statistics this is the probability distribution for a statistic based after combining established information or belief (the prior) with new evidence (the likelihood).
Positive predictive value (PPV)	In screening or diagnostic tests: A measure of the usefulness of a screening or diagnostic test. It is the proportion of those with a positive test result who have the disease, and can be interpreted as the probability that a positive test result is correct. It is calculated as follows: TP/(TP+FP)
Postoperative	Pertaining to the period after patients leave the operating theatre, following surgery.
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.

Term	Definition
Prediction study risk of bias assessment tool (PROBAST)	Checklist for assessing the risk of bias and applicability (indirectness) of evidence from prognostic risk tool studies.
Preoperative	The period before surgery commences.
Pre-test probability	In diagnostic tests: The proportion of people with the target disorder in the population at risk at a specific time point or time interval. Prevalence may depend on how a disorder is diagnosed.
Prevalence	See Pre-test probability.
Prior distribution	In Bayesian statistics this is the probability distribution for a statistic based on previous evidence or belief.
Primary care	Healthcare delivered outside hospitals. It includes a range of services provided by GPs, nurses, health visitors, midwives and other healthcare professionals and allied health professionals such as dentists, pharmacists and opticians.
Primary outcome	The outcome of greatest importance, usually the one in a study that the power calculation is based on.
Probabilistic analysis	In economic evaluation, this is an analysis that uses a probability distribution for each input. In contrast, see Deterministic analysis.
Product license	An authorisation from the MHRA to market a medicinal product.
Prognosis	A probable course or outcome of a disease. Prognostic factors are patient or disease characteristics that influence the course. Good prognosis is associated with low rate of undesirable outcomes; poor prognosis is associated with a high rate of undesirable outcomes.
Prospective study	A research study in which the health or other characteristic of participants is monitored (or 'followed up') for a period of time, with events recorded as they happen. This contrasts with retrospective studies.
Publication bias	Publication bias occurs when researchers publish the results of studies showing that a treatment works well and don't publish those showing it did not have any effect. If this happens, analysis of the published results will not give an accurate idea of how well the treatment works. This type of bias can be assessed by a funnel plot.
Quality of life	See 'Health-related quality of life'.
Quality-adjusted life year (QALY)	A measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life. One QALY is equal to 1 year of life in perfect health. QALYS are calculated by estimating the years of life remaining for a patient following a particular treatment or intervention and weighting each year with a quality of life score (on a scale of 0 to 1). It is often measured in terms of the person's ability to perform the activities of daily life, freedom from pain and mental disturbance.
Randomisation	Assigning participants in a research study to different groups without taking any similarities or differences between them into account. For example, it could involve using a random numbers table or a computer-generated random sequence. It means that each individual (or each group in the case of cluster randomisation) has the same chance of receiving each intervention.
Randomised controlled trial (RCT)	A study in which a number of similar people are randomly assigned to 2 (or more) groups to test a specific drug or treatment. One group (the experimental group) receives the treatment being tested, the other (the comparison or control group) receives an alternative treatment, a dummy treatment (placebo) or no treatment at all. The groups are followed up to see how effective the experimental treatment was. Outcomes are measured at specific times and any difference in response between the

Term	Definition
	groups is assessed statistically. This method is also used to reduce bias.
RCT	See 'Randomised controlled trial'.
Receiver operated characteristic (ROC) curve	A graphical method of assessing the accuracy of a diagnostic test. Sensitivity is plotted against 1 minus specificity. A perfect test will have a positive, vertical linear slope starting at the origin. A good test will be somewhere close to this ideal.
Reference standard	The test that is considered to be the best available method to establish the presence or absence of the outcome – this may not be the one that is routinely used in practice.
Reporting bias	See 'Publication bias'.
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Retrospective study	A research study that focuses on the past and present. The study examines past exposure to suspected risk factors for the disease or condition. Unlike prospective studies, it does not cover events that occur after the study group is selected.
Review question	In guideline development, this term refers to the questions about treatment and care that are formulated to guide the development of evidence-based recommendations.
Risk ratio (RR)	The ratio of the risk of disease or death among those exposed to certain conditions compared with the risk for those who are not exposed to the same conditions (for example, the risk of people who smoke getting lung cancer compared with the risk for people who do not smoke). If both groups face the same level of risk, the risk ratio is 1. If the first group had a risk ratio of 2, subjects in that group would be twice as likely to have the event happen. A risk ratio of less than 1 means the outcome is less likely in the first group. The risk ratio is sometimes referred to as relative risk.
Secondary outcome	An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes.
Selection bias	Selection bias occurs if: a) The characteristics of the people selected for a study differ from the wider population from which they have been drawn, or b) There are differences between groups of participants in a study in terms of how likely they are to get better.
Sensitivity	How well a test detects the thing it is testing for. If a diagnostic test for a disease has high sensitivity, it is likely to pick up all cases of the disease in people who have it (that is, give a 'true positive' result). But if a test is too sensitive it will sometimes also give a positive result in people who don't have the disease (that is, give a 'false positive'). For example, if a test were developed to detect if a woman is 6 months pregnant, a very sensitive test would detect everyone who was 6 months pregnant, but would probably also include those who are 5 and 7 months pregnant. If the same test were more specific (sometimes referred to as having higher specificity), it would detect only those who are 6 months pregnant, and someone who was 5 months pregnant would get a negative result (a 'true negative'). But it would probably also miss some people who were 6 months pregnant (that is, give a 'false negative'). Breast screening is a 'real-life' example. The number of women who are recalled for a second breast screening test is relatively high because the test is very sensitive. If it were made more specific, people who don't have the disease would be less likely to be called back for a second test but

Term	Definition
	more women who have the disease would be missed.
Sensitivity analysis	A means of representing uncertainty in the results of economic evaluations. Uncertainty may arise from missing data, imprecise estimates or methodological controversy. Sensitivity analysis also allows for exploring the generalisability of results to other settings. The analysis is repeated using different assumptions to examine the effect on the results. One-way simple sensitivity analysis (univariate analysis): each parameter is varied individually in order to isolate the consequences of each parameter on the results of the study. Multi-way simple sensitivity analysis (scenario analysis): 2 or more parameters are varied at the same time and the overall effect on the results is evaluated. Threshold sensitivity analysis: the critical value of parameters above or below which the conclusions of the study will change are identified. Probabilistic sensitivity analysis: probability distributions are assigned to the uncertain parameters and are incorporated into evaluation models based on decision analytical techniques (for example, Monte Carlo simulation).
Significance (statistical)	A result is deemed statistically significant if the probability of the result occurring by chance is less than 1 in 20 ($p<0.05$).
Specificity	The proportion of true negatives that are correctly identified as such. For example in diagnostic testing the specificity is the proportion of non-cases correctly diagnosed as non-cases. See related term 'Sensitivity'. In terms of literature searching a highly specific search is generally narrow and aimed at picking up the key papers in a field and avoiding a wide range of papers.
Stakeholder	An organisation with an interest in a topic that NICE is developing a clinical guideline or piece of public health guidance on. Organisations that register as stakeholders can comment on the draft scope and the draft guidance. Stakeholders may be: • manufacturers of drugs or equipment • national patient and carer organisations • NHS organisations • organisations representing healthcare professionals.
State transition model	See Markov model
Systematic review	A review in which evidence from scientific studies has been identified, appraised and synthesised in a methodical way according to predetermined criteria. It may include a meta-analysis.
Time horizon	The time span over which costs and health outcomes are considered in a decision analysis or economic evaluation.
Transition probability	In a state transition model (Markov model), this is the probability of moving from one health state to another over a specific period of time.
Treatment allocation	Assigning a participant to a particular arm of a trial.
Univariate	Analysis which separately explores each variable in a data set.
Utility	In health economics, a 'utility' is the measure of the preference or value that an individual or society places upon a particular health state. It is generally a number between 0 (representing death) and 1 (perfect health). The most widely used measure of benefit in cost-utility analysis is the quality-adjusted life year, but other measures include disability-adjusted life years (DALYs) and healthy year equivalents (HYEs).