

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Final appraisal document

Selective internal radiation therapies for treating hepatocellular carcinoma

1 Recommendations

- 1.1 The selective internal radiation therapy (SIRT) SIR-Spheres is recommended as an option for treating unresectable advanced hepatocellular carcinoma (HCC) in adults, only if:
- used for people with Child-Pugh grade A liver impairment when conventional transarterial therapies are inappropriate, and
 - the company provides SIR-Spheres according to the commercial arrangement (see section 2).
- 1.2 The SIRT TheraSphere is recommended as an option for treating unresectable advanced hepatocellular carcinoma (HCC) in adults, only if:
- used for people with Child-Pugh grade A liver impairment when conventional transarterial therapies are inappropriate, and
 - the company provides TheraSphere according to the commercial arrangement (see section 2).
- 1.3 The SIRT QuiremSpheres is not recommended, within its CE marking, for treating unresectable advanced hepatocellular carcinoma (HCC) in adults.
- 1.4 These recommendations are not intended to affect treatment with SIR-Spheres, TheraSphere and QuiremSpheres that was started in the NHS before this guidance was published. People having treatment outside these recommendations may continue without change to the funding arrangements in place for them before this guidance was

published, until they and their NHS clinician consider it appropriate to stop.

Why the committee made these recommendations

Treatment for HCC depends on the stage of the disease and liver function.

Treatment options include surgery, ablation, transarterial therapies, chemotherapy (such as sorafenib) and best supportive care. Treatment does not cure the disease for most people.

SIRTs are small radioactive beads that are injected into the liver's blood supply to treat liver cancer. QuiremSpheres, SIR-Spheres and TheraSphere are the 3 SIRTs considered in this appraisal. The clinical trial data for these SIRTs compared with other treatment options are limited. But, compared with sorafenib, SIRTs may have fewer and more manageable adverse effects, which can improve quality of life.

There is not enough evidence to consider SIRTs a cost-effective use of NHS resources for people with early and intermediate HCC.

For people with advanced HCC, QuiremSpheres is less clinically effective than sorafenib and costs more, so it is not recommended. SIR-Spheres and TheraSphere are slightly less clinically effective than sorafenib but cost less. The cost savings mean that SIR-Spheres and TheraSphere can be recommended as options for people with Child-Pugh grade A liver impairment when conventional transarterial therapies are inappropriate.

2 Information about QuiremSpheres, SIR-Spheres and TheraSphere

CE mark for QuiremSpheres

2.1 QuiremSpheres (Quirem Medical, CE mark holder; Terumo Europe, distributor) received its CE mark in April 2015. It is classified as an Active Implantable Medical Device by Council Directive 90/385/EEC. It is indicated for treating unresectable liver tumours.

Dosage in the CE mark for QuiremSpheres

2.2 QuiremSpheres is given through a catheter to the hepatic artery. The product is supplied as a customised, patient-specific dose. The maximum range of the emitted beta particles in tissue is 8.7 mm with a mean of 2.5 mm. Also, holmium-166 emits primary gamma photons (81 kilo electronvolt, KeV). The half-life is 26.8 hours, which means more than 90% of the radiation is given in the first 4 days after the procedure. At the moment of treatment, the activity per microsphere is 200 to 400 Becquerel (Bq). The number of particles implanted depends on the targeted liver volume and ranges, on average, from 10 to 30 million.

Price of QuiremSpheres

2.3 The company has stated that the acquisition cost of QuiremSpheres is £9,896 for a single treatment. The company has a commercial arrangement (simple discount patient access scheme), which would have applied if the technology had been recommended.

CE mark for SIR-Spheres

2.4 SIR-Spheres (Sirtex) received its CE mark as a class 3 active medical device in October 2002. It is indicated for treating advanced inoperable liver tumours.

Dosage in the CE mark for SIR-Spheres

2.5 SIR-Spheres is given through a catheter to the hepatic artery. It is supplied at 3 giga Bq yttrium-90 per vial in 5 ml water for injection in a shielded shipping vial. Each vial contains 40 to 80 million microspheres, ranging from 20 to 60 micrometres in diameter (median diameter 32.5 micrometres). The maximum range of beta emission in tissue is 11 mm with a mean of 2.5 mm. The average number of particles implanted is 30 million to 60 million.

Price for SIR-Spheres

2.6 The company has stated that the acquisition cost of SIR-Spheres is £8,000 for a single treatment. The company has a commercial arrangement (simple discount patient access scheme). This makes SIR-Spheres available to the NHS with a discount. The size of the discount is commercial in confidence. It is the company's responsibility to let relevant NHS organisations know details of the discount.

CE mark for TheraSphere

2.7 TheraSphere (BTG) received its CE mark as a class 3 active medical device in September 2014. It is indicated for treating hepatic neoplasia.

Dosage in the CE mark for TheraSphere

2.8 TheraSphere is given through a catheter to the hepatic artery. It is supplied in 6 dose sizes: 3 GBq, 5 GBq, 7 GBq, 10 GBq, 15 GBq or 20 GBq in 0.6 ml pyrogen-free water supplied in a 1 ml vial, inside an acrylic shield. Custom dose sizes are also available in increments of 0.5 GBq between 3 GBq and 20 GBq. A single treatment with TheraSphere contains 1.2 million to 8 million microspheres. The recommended dose to the liver is 80 Gy to 150 Gy.

Price for TheraSphere

2.9 The company has stated that the acquisition cost of TheraSphere is £8,000 for a single treatment. The company has a commercial arrangement (simple discount patient access scheme). This makes TheraSphere available to the NHS with a discount. The size of the discount is commercial in confidence. It is the company's responsibility to let relevant NHS organisations know details of the discount.

3 Committee discussion

The [appraisal committee](#) considered evidence from a number of sources. See the [committee papers](#) for full details of the evidence.

Potential new treatment option

People with hepatocellular carcinoma would welcome a new treatment option

3.1 Hepatocellular carcinoma (HCC) is the most common form of liver cancer in England. Treatment depends on the location and stage of the cancer, and how well the liver is functioning. Treatment options include surgery or ablation in early-stage disease, transarterial therapies in intermediate-stage disease, and chemotherapy or systemic therapy in advanced-stage disease, as well as best supportive care. Treatment does not cure the disease for many people. The clinical experts explained that selective internal radiation therapy (SIRT) has also been used for HCC in England through compassionate schemes. Patient experts explained that HCC can have a substantial impact on quality of life. People with HCC and their carers live with uncertainty and hopelessness. Often people with HCC also live with stigma and isolation because of underlying causes of disease, such as alcohol. Clinical experts highlighted that people with advanced HCC have a poor prognosis with median life expectancy of less than 12 months. The committee concluded that people with HCC would welcome a new treatment option.

People with HCC and portal vein thrombosis are a relevant subgroup

3.2 The clinical experts explained that portal vein involvement, such as portal vein thrombosis (PVT), is a common comorbidity that might negatively affect prognosis. PVT happens when a blood clot narrows the vein that takes blood to the liver from the intestines. The committee understood that people with PVT were included in the NICE scope for this appraisal. It concluded that evidence for people with HCC and PVT should be considered.

This appraisal assesses 3 SIRTs for treating HCC

3.3 QuiremSpheres, SIR-Spheres and TheraSphere are SIRTs. These are small radioactive beads that are injected into the liver's blood supply to treat liver cancer. The 3 SIRTs are medical devices with CE marks for

their indications. QuiremSpheres is indicated for treating unresectable liver tumours, SIR-Spheres for treating advanced inoperable liver tumours and TheraSphere for treating hepatic neoplasia. The committee was aware that the scope for the appraisal was narrower than the CE marks, because it only included unresectable HCC, when SIRTs are most likely to be used. The committee agreed that the 3 SIRTs should be compared with each other and with available treatments to assess their cost effectiveness for treating HCC.

SIRTs are already used in the NHS for other cancers, but not for HCC

3.4 The clinical experts and NHS England explained that SIRTs are available in some specialist centres across England for other cancers (such as metastatic colorectal cancer). The committee understood that SIRTs are currently not commissioned for HCC in the NHS but that the infrastructure and knowledge for using SIRTs exists in some specialist centres.

Clinical management

Stage of cancer and liver function characterise the disease and people with HCC are a heterogenous population

3.5 There are different causes of HCC, including cirrhosis, alcohol, fatty liver disease and hepatitis. Therefore, people with HCC are a heterogenous population and their disease is characterised by both stage of cancer and liver function. Treatment choice is multifaceted because both the cancer and liver function affect treatment outcomes. Clinical experts advised that in England clinicians use the Barcelona Clinic Liver Cancer (BCLC) staging system and the Child-Pugh score to inform treatment decisions. BCLC staging looks at the number and size of tumours in the liver. There are 5 stages: very early stage (BCLC 0), early stage (BCLC A), intermediate stage (BCLC B), advanced stage (BCLC C) and terminal stage (BCLC D). The committee agreed that stages A, B and C align with the scope for this appraisal.

- The Child-Pugh score assesses liver function. It has 5 components: serum albumin levels, bilirubin levels, time for blood to clot, presence of ascites (fluid in the peritoneal cavity) and presence of hepatic encephalopathy. There are 3 classes: class A (the liver is working normally), class B (mild to moderate liver damage), class C (severe liver damage). People with BCLC A to C can have either good liver function (Child-Pugh A) or mild to moderate liver damage (Child-Pugh B).
- More recently an alternative measure, the albumin-bilirubin (ALBI) grade, was developed to look at liver function. The committee was aware that in previous NICE guidance for HCC, the Child-Pugh score was used as a criterion for treatment, but that ALBI was not. The clinical experts advised that ALBI is less frequently used for this purpose, and that Child-Pugh score is expected to be the measure of choice for the foreseeable future.

Treatment of HCC differs between the 3 BCLC stages and is influenced by Child-Pugh score

3.6 Treatment options include ablation and transplant in early disease, and conventional transarterial therapies (CTT) such as transarterial chemoembolisation (TACE) or transarterial embolisation (TAE) in intermediate disease. In advanced disease, treatment options are chemotherapy or systemic therapy with sorafenib, lenvatinib or regorafenib. For some people the aim of treatment might be to reduce the tumour size ('downstaging') to potentially allow subsequent transplantation that could cure the disease. The committee understood that people with HCC have different treatment options depending on the stage of their disease as assessed by BCLC and Child-Pugh score.

There are 3 distinct subgroups relevant to this appraisal

3.7 The committee concluded that there are 3 subgroups relevant for this appraisal:

- People for whom liver transplant is appropriate, including people with BCLC A and Child-Pugh A or B.
- People for whom CTT is appropriate, including people with BCLC B and Child-Pugh A or B.
- People for whom CTT is inappropriate, including people with BCLC C and Child-Pugh A or B.

In people with early disease, ablation and transplant are standard care in current NHS practice in England

3.8 Treatment options for early disease (BCLC A) are ablation and transplant. However, 1 clinical expert explained that transplants might not be available for people with good liver function (Child-Pugh A). The committee concluded that both ablation and transplant are standard care for people with early disease in clinical practice in England.

In people with intermediate disease, CTTs are standard care in current NHS practice in England

3.9 Treatments for intermediate disease (BCLC B) are CTTs, including transarterial chemoembolisation (TACE), drug-eluting bead transarterial chemoembolisation (DEB-TACE) and transarterial embolisation (TAE). The committee accepted that all CTTs available in the NHS in England are appropriate comparators for people with intermediate disease.

In people with advanced disease, sorafenib is standard care in current NHS practice in England

3.10 The systemic therapies sorafenib and lenvatinib are both recommended for advanced HCC (BCLC C) in people with Child-Pugh grade A liver impairment ([NICE technology appraisal guidance on sorafenib for treating advanced hepatocellular carcinoma](#) and [lenvatinib for untreated advanced hepatocellular carcinoma](#)). Regorafenib is only recommended after treatment with sorafenib ([NICE technology appraisal guidance on regorafenib for previously treated advanced hepatocellular carcinoma](#)). The committee understood that sorafenib is standard care in clinical

practice in England because there are subsequent treatments available after progression with sorafenib. The clinical expert confirmed that lenvatinib is now rarely used. The committee concluded that sorafenib is the most appropriate comparator for SIRTs in people with advanced disease and with Child-Pugh grade A liver impairment.

Clinical evidence

The systematic review included non-RCT evidence when not enough RCT evidence was identified

3.11 The assessment group (AG) did a systematic review of the clinical evidence on SIRTs and comparators. Randomised controlled trials (RCTs) were eligible for inclusion in the review. The AG had identified all the RCTs that were also identified by the companies in their submissions. The committee was aware of non-RCT evidence and noted that typically the risk of bias in non-RCT evidence is higher than in RCT evidence. It agreed with the AG's approach to only include non-RCT evidence in the review when there was not enough RCT evidence. The committee understood that some studies might include a mixed population containing all 3 subgroups of interest. It agreed to exclude studies from the network meta-analyses if they did not provide separate results for the 3 subgroups (see section 3.7). The committee used the AG's analysis for its decision making. This was because it included evidence for all 3 SIRTs and so was more comprehensive than the companies' submissions.

There is not enough evidence to assess the clinical effectiveness of QuiremSpheres in the 3 subgroups relevant to this appraisal

3.12 The clinical evidence for QuiremSpheres came from 1 retrospective case series including 9 people that showed a 56% response rate. A mixed population was included, and results were only presented for the whole study population. The committee concluded that the single, small retrospective study did not provide enough data to assess the clinical

effectiveness of QuiremSpheres in any of the 3 subgroups relevant to this appraisal (see section 3.7).

There is limited randomised clinical evidence for TheraSphere compared with TACE when transplant is appropriate

3.13 Two small RCTs (PREMIERE and Kulik et al. 2014) for TheraSphere were identified that included people for whom transplant was appropriate (see section 3.7). The committee was also aware of 10 non-RCT studies, including 7 prospective comparative studies, that included people from the 3 subgroups relevant to this appraisal. The PREMIERE study was done in the US and included 45 people for whom transplant was appropriate. It compared TheraSphere with TACE as an alternative to prepare for transplant. The AG advised that PREMIERE had a high risk of bias because of concerns with randomisation and potential deviations from the intended interventions. Also, the baseline characteristics were different in the 2 arms so people in the TACE arm had better prognosis than people in the TheraSphere arm. Overall survival of people who had a transplant was numerically, but not statistically, significantly longer in the TheraSphere arm. The median overall survival for TheraSphere was 18.6 months (95% confidence interval [CI] 7.4 to 32.5) compared with 17.7 months (95% CI 7.4 to 32.5) for TACE. The committee concluded that there was limited evidence, with a high risk of bias, to establish whether TheraSphere was better or worse than TACE when transplant is appropriate.

There is limited evidence for TheraSphere compared with TheraSphere plus sorafenib when transplant is appropriate

3.14 The study by Kulik et al. (2014) was done in the US and included 20 people for whom transplant was appropriate. It compared TheraSphere with TheraSphere plus sorafenib. The AG had some concerns with the randomisation process, treatments received and measurement of outcomes. The baseline characteristics were different in the 2 arms so people in the TheraSphere plus sorafenib arm had a better prognosis.

There was no evidence of a difference in overall survival between the 2 arms (3 deaths in the TheraSphere arm, 2 deaths in the combination arm). The committee was aware that TheraSphere plus sorafenib was not included in sorafenib's marketing authorisation or TheraSphere's CE mark. The committee concluded that there was limited evidence, with high risk of bias, to establish whether TheraSphere is better or worse than TheraSphere with sorafenib when transplant is appropriate.

Non-randomised evidence comparing TheraSphere with non-SIRT treatments is not robust enough for decision making

3.15 Of the 7 prospective comparative non-RCTs, only 4 reported overall survival or progression-free survival. Of these, 2 compared TheraSphere with TACE or DEB-TACE across the 3 subgroups. The AG suggested that both studies had high risk of bias and differences in baseline characteristics such as age, tumour size and number of tumours. The committee concluded that results from these studies might be unreliable for decision making. Another study compared TheraSphere with TheraSphere plus sorafenib, in people for whom CTT is inappropriate. The remaining prospective study was done in people for whom CTT is inappropriate. This compared TheraSphere in people with PVT with TheraSphere in people without PVT and best supportive care. The AG advised that this study had a high risk of bias, and that the people in the treatment arms had very different baseline characteristics. The committee recognised that the large volume of non-randomised evidence might be useful for tentative conclusions, but it remained aware of the limitations of non-RCT studies. Therefore, it agreed that they should not be used for decision making. Also, there was not enough evidence to establish whether TheraSphere is better or worse than other treatments in people for whom CTT is appropriate and in people for whom CTT is inappropriate.

There are insufficient data to establish the clinical effectiveness of SIR-Spheres compared with non-SIRT treatments when transplant is appropriate

3.16 The AG identified 1 RCT comparing SIR-Spheres with TACE (SIR-TACE) that included people for whom transplant was appropriate. SIR-TACE was done in Germany and Spain, and included 28 people with early, intermediate and late-stage disease. Only overall results for the mixed population were available. The AG assessed that the study had a high risk of bias because of the randomisation process, missing outcome data and measurement of the outcome. The committee concluded that there are insufficient data to establish whether SIR-Spheres are better than TACE when transplant is appropriate.

It is unclear whether SIR-Spheres is better than DEB-TACE or TACE when CTT is appropriate

3.17 The AG identified 2 RCTs that compared SIR-Spheres with TACE (SIR-TACE) or DEB-TACE (Pitton et al. 2015) that included people for whom CTT is appropriate in their trial populations. SIR-TACE is described in section 3.16. Pitton et al. (2015) was done in Germany and included 24 people with intermediate-stage disease (BCLC B). Overall survival and progression-free survival were longer in the DEB-TACE arm compared with the SIR-Spheres arm, but this was not statistically significant (788 days compared with 592 days and 216 days compared with 180 days, respectively). Because of this and the small sample size, the committee concluded that it could not establish whether SIR-Spheres was better than TACE or DEB-TACE when CTT is appropriate.

People in SARAH had poorer prognosis than people seen in clinical practice in England

3.18 The AG identified 2 RCTs comparing SIR-Spheres with sorafenib (SARAH and SIRveNIB) in people for whom CTT was inappropriate. SARAH was done in France between 2011 and 2015 and included a heterogeneous population of people with HCC. This included, for example, people with advanced HCC, people with HCC who had previous treatment with

2 treatments of TACE, and people with Child-Pugh A or B liver impairment. There was no difference in overall survival or progression-free survival between the treatment arms. The median overall survival was 8.0 months (95% CI 6.7 to 9.9) for SIR-Spheres and 9.9 months (95% CI 8.7 to 11.4) for sorafenib. The hazard ratios (HRs) were 1.15 (95% CI 0.94 to 1.41) for the intention-to-treat (ITT) population and 0.99 (95% CI 0.79 to 1.24) for the per-protocol population. The median progression-free survival was 4.1 months (95% CI 3.8 to 4.6) for SIR-Spheres and 3.7 months (95% CI 3.3 to 5.4) for sorafenib. The HR was 1.03 (95% CI 0.85 to 1.25) for the ITT population. More adverse events were reported with sorafenib than SIR-Spheres. A post-hoc analysis of SARA focused on people with ALBI grade 1 and low tumour burden (25% or less tumour burden). Again, there was no difference in overall or progression-free survival between the treatment arms. The median overall survival was 21.9 months (95% CI 15.2 to 32.5) for SIR-Spheres and 17.0 months (95% CI 11.6 to 20.8) for sorafenib. The HR was 0.73 (95% CI 0.44 to 1.21). The median progression-free survival HR was 0.65 (95% CI 0.41 to 1.02). The clinical experts advised that the SARA trial had more people with a high tumour burden, PVT and impaired liver function than people seen in clinical practice in England. The committee concluded that people in the SARA trial had poorer prognosis than people seen in clinical practice in England.

The results from SIRveNIB may not be fully generalisable to the NHS

- 3.19 SIRveNIB was done in the Asia-Pacific region between 2010 and 2018. The clinical experts explained that results from SIRveNIB might not be generalisable to the NHS in England. This was because in the Asia-Pacific region HCC is often caused by hepatitis B and C, whereas in the UK fatty liver disease and alcohol are the most common causes. There was no difference in overall survival or progression-free survival between the treatment arms. The median overall survival was 8.8 months for SIR-Spheres and 10.0 months for sorafenib. The HRs were 1.12 (95% CI 0.9 to 1.4) for the ITT population and 0.86 (95% CI 0.7 to 1.1) for the per-

protocol population. The median progression-free survival was 5.8 months for SIR-Spheres and 5.1 months for sorafenib. The HRs were 0.89 (95% CI 0.7 to 1.1) for the ITT population and 0.73 (95% CI 0.6 to 0.9) for the per-protocol population. More adverse events were reported with sorafenib than SIR-Spheres. The committee concluded that results from SIRveNIB may not be fully generalisable to people seen in the NHS.

The evidence from SARAH and SIRveNIB is preferable to non-RCT evidence for decision making when CTT is inappropriate

3.20 The committee considered including non-RCT evidence identified by the AG. The AG assessed the 3 non-RCT studies as having a high risk of bias. So, the committee concluded that the RCT evidence from SARAH and SIRveNIB was preferable for decision making in people for whom CTT was inappropriate.

There is no evidence to compare the 3 SIRTs' effectiveness when transplant or CTT is appropriate

3.21 The clinical evidence for comparative effectiveness of the 3 SIRTs came from 6 retrospective studies that reported overall survival or progression-free survival. Of these, 5 compared SIR-Spheres with TheraSphere and 1 small study of 30 people compared all 3 SIRTs. The AG advised that most of these studies had a high risk of bias because of selection and performance bias. None of the studies included people for whom transplant was appropriate. The study comparing all 3 SIRTs may have included people for whom CTTs were appropriate but there were no results presented for this subgroup. The committee concluded that there was no evidence identified for people when transplant or CTT was appropriate.

There is not enough direct evidence to compare the 3 SIRTs' effectiveness when CTT is inappropriate, so a mixed treatment comparison is considered

3.22 The AG identified 5 retrospective studies that included people for whom CTT is inappropriate (see section 3.21). The study comparing all 3 SIRTs

also included people for whom CTTs were appropriate, but no results for subgroups were presented. The committee was aware that the populations were different across these studies and acknowledged that this meant results were difficult to compare. The committee was also aware that the baseline characteristics were different in most studies, and that this might affect prognosis and outcomes between the arms. In 2 studies that compared TheraSphere with SIR-Spheres, there was no difference in overall survival. In van der Gucht et al. (2017; n=77), the median overall survival was 7.0 months for TheraSphere (95% CI 1.6 to 12.4) compared with 7.7 months for SIR-Spheres (95% CI 7.2 to 8.2). In Bhangoo et al. (2015; n=17) the median overall survival for TheraSphere was 8.4 months (95% CI 1.3 to 21.1) compared with 7.8 months for SIR-Spheres (95% CI 2.3 to 12.5). In 2 studies (Biederman et al. 2015 and Biederman et al. 2016) that compared TheraSphere with SIR-Spheres in people with PVT, overall survival was better in the TheraSphere arm than the SIR-Spheres arm. The committee concluded that there was not enough direct evidence to establish the relative effectiveness of the 3 SIRTs in people with HCC, and so decided to consider mixed treatment comparisons for decision making.

There was not enough robust evidence to establish the clinical effectiveness of SIRTs compared with non-SIRT treatments for people with PVT

3.23 The clinical expert explained that people with PVT (see section 3.2) have poorer prognosis and limited treatment options. Often the only available treatment is sorafenib because people with PVT do not tolerate TACE. Therefore, the committee agreed that people with PVT might benefit more than others from treatment with SIRTs. It considered the evidence that included people with PVT (see section 3.15 and section 3.22). There was no new evidence presented specifically for this subgroup at consultation. The committee concluded there was not enough robust evidence to establish the clinical effectiveness of SIRTs compared with non-SIRT treatments for people with PVT.

There is not enough robust data to establish whether SIRT is better or worse than sorafenib or TACE in people with large tumours

3.24 After consultation, the committee considered the evidence for people with 1 or more large tumours (5 cm or larger) with or without PVT. This was because this subgroup might benefit more than others from treatment with SIRT. The committee understood that in the UK this group currently has sorafenib or TACE. Clinical evidence showed that TACE is not very effective and there are substantial adverse events with sorafenib. The committee saw data from 1 study in this group. This study, DOSISPHERE-01, compared TheraSphere personalised dosimetry with TheraSphere standard dosimetry. Personalised dosimetry improved the response rate and overall survival compared with standard dosimetry (overall survival for personalised dosimetry 26.6 months, 95% CI 11.7 months to not reached; compared with standard dosimetry 10.7 months, 95% CI 6.0 months to 16.8 months; $p=0.0096$). The committee understood that people in the 2 arms of the study might not be similar and therefore the results may have selection bias. Also, there were no data comparing SIRT with sorafenib or TACE in this group. The committee acknowledged that personalised dosimetry could improve the effectiveness of SIRT. It concluded that there were not enough robust data to establish whether SIRT is better or worse than sorafenib or TACE in people with 1 or more large tumours.

People who are unable to tolerate sorafenib might benefit from treatment with SIRT but there is no comparative evidence

3.25 Clinical expert comments provided during consultation advised that people who are unable to tolerate sorafenib do not have alternative treatment options. This means they have best supportive care. The committee understood that there is some clinical experience in England of this group having treatment with SIRT, with promising outcomes. It also acknowledged that this group is not included in the RCTs because of their characteristics (for example, older age and comorbidities). Despite the

lack of evidence, the committee concluded that people who are unable to tolerate sorafenib might benefit from treatment with SIRTs.

Most of the RCT evidence is in people with advanced disease with Child-Pugh A grade liver impairment, which is the relevant subgroup

3.26 The committee recalled that current treatments for advanced HCC are only recommended for people with Child-Pugh grade A liver impairment (see section 3.10). People with Child-Pugh grade B liver impairment have best supportive care. It noted that there was no best supportive care arm in the RCTs (SARAH and SIRveNIB), which compared SIR-Spheres with sorafenib only. The committee acknowledged that most people in the trials had Child-Pugh grade A liver impairment (83% in SARAH and 90% in SIRveNIB). Therefore, it concluded that the trial results were acceptable for decision making in people with Child-Pugh grade A liver impairment. However, it could not establish whether SIRTs were effective in people with Child-Pugh grade B liver impairment, because of the lack of evidence comparing SIRTs with the relevant comparator for that group. Therefore, in people with advanced HCC, the subgroup with Child-Pugh A liver impairment was appropriate for decision making.

SIRTs have fewer and less severe side effects than other treatment options

3.27 The clinical and patient experts stated that there were fewer and less severe side effects with SIRTs than with other treatments. Also, side effects from SIRTs have a shorter duration, whereas side effects from chemotherapies such as sorafenib can continue for the whole treatment course. After the second committee meeting, the committee invited companies and stakeholders to submit additional data on adverse event severity and duration. The committee considered adverse event data from SARAH for SIR-Spheres and non-RCT studies for TheraSphere. The data included adverse event rates and durations for all severity grades. They showed that SIRTs and sorafenib have different adverse event profiles. The committee was aware that data on event duration were averaged across all severity grades and both study arms in SARAH. The committee

concluded that SIRTs were likely to have fewer and less severe side effects than sorafenib, and that this benefit may be important to patients. The committee agreed that this should be captured in the cost-effectiveness analysis and taken into account during decision making.

Mixed treatment comparisons

Data are not robust enough to provide a meaningful comparison between treatment options when transplant is appropriate

3.28 The AG considered the feasibility of a mixed treatment comparison to estimate comparative effectiveness between available treatment options for people when transplant is appropriate. There are 2 RCTs that could be included in this analysis. Both were done in the US and compared TheraSphere with TACE (n=45) or with a combination of TheraSphere and sorafenib (n=20). Also, the committee recalled that ablation or transplant was the most relevant comparator for people for whom transplant is appropriate (see section 3.8). Because of limited data, results from the mixed treatment comparison would be very uncertain. The committee concluded that a mixed treatment comparison in this population would not help decision making for the subgroup for whom transplant is appropriate.

Estimates comparing effectiveness for treatment options in people for whom CTT is appropriate are very uncertain, and are not suitable for decision making

3.29 After consultation on the assessment report, the AG did a mixed treatment comparison in people for whom CTT was appropriate. There were 6 RCTs that could be included in this analysis: 5 compared different CTTs with each other and 1 compared SIR-Spheres with DEB-TACE (n=24). The AG also included 1 retrospective study that compared SIR-Spheres with TheraSphere (n=77). From this study, only a subgroup of 35 people with early or intermediate HCC could be included in the analysis. The study had a high risk of bias because its 2 treatment groups were not similar at baseline (people with small tumour volumes were preferentially treated

with TheraSphere). The committee agreed that there was little evidence to link SIR-Spheres and TheraSphere to the network of treatments. Results from the mixed treatment comparison for overall survival and progression-free survival were uncertain, with wide credible intervals that included a HR of 1 (no statistical difference between treatment options). The committee concluded that the results from the mixed treatment comparison in this population were uncertain. Also, there was not enough evidence in this population to compare SIR-Spheres with TheraSphere, or compare the SIRTs with TACE, DEB-TACE and TAE.

The comparative effectiveness estimates of the 3 SIRTs in people for whom CTT is inappropriate are uncertain

3.30 The AG did a mixed treatment comparison to estimate comparative effectiveness between available treatment options in people when CTT was inappropriate. There were 3 RCTs included in this analysis. Of these, 1 RCT compared lenvatinib with sorafenib and 2 compared sorafenib with SIR-Spheres. To include TheraSphere in the network, 2 retrospective studies comparing TheraSphere with SIR-Spheres were included in sensitivity analyses. There were no data for QuiremSpheres to be included in the analysis. In the main analysis, when CTT is inappropriate and people have Child-Pugh grade A liver impairment, there was no evidence of a difference between SIR-Spheres and sorafenib. In the ITT population for SIR-Spheres compared with sorafenib, the HR was 1.13 (95% CI 0.96 to 1.32). A value of less than 1 indicates better overall survival. When the retrospective evidence was included, TheraSphere was shown to be more effective than SIR-Spheres and sorafenib. In the ITT population the HR for SIR-Spheres compared with sorafenib was 1.13 (95% credible interval [CrI] 0.96 to 1.32). For TheraSphere compared with sorafenib or SIR-Spheres, the HRs were 0.47 (95% CrI 0.21 to 0.88) or 0.41 (95% CrI 0.20 to 0.77) respectively. In an alternative analysis with a wider population, SIR-Spheres was less effective than sorafenib. In the corresponding sensitivity analysis including the retrospective evidence, TheraSphere was again more effective than SIR-Spheres and sorafenib.

Final appraisal document – Selective internal radiation therapies for treating hepatocellular carcinoma

The committee recalled the AG's assessment that the retrospective studies had a high risk of bias and uncertain results (see section 3.15). The committee agreed that the retrospective studies should not be included in the analysis because of the risk of bias. It concluded that the comparative effectiveness results based on RCT evidence from SIR-Spheres could be used in a cost-effectiveness analysis. The committee also concluded that, because its preferred network meta-analysis only had evidence for 1 SIRT (SIR-Spheres), the comparative effectiveness of the 3 SIRTs was uncertain.

Cost-effectiveness evidence

The AG's model is used for decision making

3.31 Two companies included economic analyses in their evidence submissions. For SIR-Spheres, the company submitted a cost-minimisation analysis for people for whom CTT was appropriate, and a cost-utility analysis for people for whom CTT was inappropriate. The cost-utility analysis was restricted to people with ALBI grade 1 and low tumour burden, a subpopulation from the SARA trial. For TheraSphere, the company submitted 2 cost-utility analyses, 1 for people for whom CTT was appropriate and 1 for people for whom CTT was inappropriate. The committee acknowledged the submission of the companies' models. It noted that the AG model used a similar structure (see section 3.32) as the companies' cost-utility analyses. Also, the AG used inputs from the companies' models, such as costs and treatment frequency. The committee concluded that there was not enough evidence to support an economic analysis in people for whom CTT was appropriate (see section 3.29). When CTT was inappropriate, the AG model was the most suitable for decision making because it included all 3 SIRTs as specified in the NICE scope (see section 3.3).

The structure of the AG model for people for whom CTT is inappropriate is acceptable for decision making

3.32 The AG did a cost-utility analysis for people with unresectable intermediate (BCLC stage B) or advanced (BCLC stage C) HCC, when CTT was inappropriate, with or without macroscopic vascular invasion but without extrahepatic disease. The model consisted of a decision tree and partitioned survival model with 3 health states. The decision tree represented the outcome of the work-up procedure that happens before SIRT. The partitioned survival model was like that used by the companies. The interventions were SIR-Spheres, TheraSphere and QuiremSpheres, which were assumed to have equal effectiveness in the base case (see section 3.333). The comparators were initially sorafenib and lenvatinib. Because sorafenib and lenvatinib are recommended only for people with Child-Pugh grade A liver impairment, the base-case analysis was restricted to this population. The committee concluded that the model structure was acceptable for decision making.

Cost-effectiveness results assuming all SIRTs are equally effective have been considered, but this is uncertain for QuiremSpheres

3.33 The AG's economic analysis assumed that the 3 SIRTs were equally effective. Most data used in the model, such as clinical effectiveness and adverse event data, were from the SARAH trial for SIR-Spheres. There was very little evidence for QuiremSpheres to inform the model (see section 3.12), and the evidence for TheraSphere was less certain than the evidence for SIR-Spheres (see section 3.15). The committee noted that there was not enough evidence to establish whether the 3 SIRTs had different effectiveness (see section 3.22 and section 3.30). The committee considered whether it was appropriate to assume the 3 SIRTs were equally effective. It noted that the technologies used different beads to give treatment, and QuiremSpheres used a different isotope to the other SIRTs. It agreed that these differences might result in different effectiveness and adverse event profiles, to an unknown extent. In the

absence of better evidence, the committee concluded that it would consider the cost effectiveness of the 3 SIRTs by assuming they were equally effective, generalising the SIR-Spheres data to the other 2 SIRTs. It also concluded that by doing so, the cost-effectiveness estimates for QuiremSpheres would be more uncertain than those for TheraSphere and substantially more uncertain than for SIR-Spheres. It took this uncertainty into consideration in its decision making.

Sorafenib is the only relevant comparator for assessing the cost effectiveness of SIRTs in people for whom CTT is inappropriate

3.34 In line with the NICE scope, the AG initially included sorafenib and lenvatinib as comparators in the model. The AG used the hazard ratio from the mixed treatment comparison to include lenvatinib in the model and assumed proportional hazards over time. Therefore, it chose the Weibull function to model overall survival and progression-free survival, even though the Weibull was not the best-fitting function. After consultation on the AG report, sorafenib was considered to be the only relevant comparator (see section 3.10). The generalised gamma was used to fit overall survival and progression-free survival in the revised base case, because the proportional hazards assumption was no longer needed. The committee also recalled that the trial evidence could be generalised to people with Child-Pugh A liver impairment, who can have sorafenib in current practice, but not to people with Child-Pugh B liver impairment, who have best supportive care (see section 3.26). It concluded that sorafenib was the only appropriate comparator, and that the best-fitting function (generalised gamma) should be used to estimate overall survival and progression-free survival.

There are not enough robust data for the ALBI grade 1 and low tumour burden subgroup for decision making

3.35 The AG presented scenario analyses that restricted the population to people with ALBI grade 1 and low tumour burden. The clinical experts explained that ALBI grade could be a more objective measure than

Child-Pugh score for liver impairment and that people with ALBI grade 1 have good liver function. However, this measure is not routinely used in the NHS, and the Child-Pugh score is expected to be the standard assessment method for liver impairment for the foreseeable future (see section 3.5). The committee was aware that clinical outcomes for the ALBI grade 1, low tumour burden subgroup came from a post-hoc analysis of the SARAH trial (n=85, section 3.188). It agreed that this analysis was not robust because the subgroup was not prespecified and the numbers were small. It was not presented with additional evidence after consultation. It concluded that it had not seen sufficiently robust data in this subgroup, but agreed that more evidence may be useful for decision making.

Usually, only 1 lobe is treated at a time in people with bilobar disease

3.36 HCC can be unilobar (tumour in 1 lobe of the liver) or bilobar (tumours in both lobes of the liver). The clinical experts explained that people with bilobar disease have a higher risk of liver impairment, and therefore usually only 1 lobe is treated at a time. The same lobe might be treated twice to reduce the size of the tumour. The committee concluded that it is not appropriate for a model to assume that both lobes are treated simultaneously in bilobar disease.

Downstaging of HCC might benefit some people with advanced HCC, but the proportion of people and subsequent outcomes are uncertain

3.37 The clinical experts explained that downstaging might be a treatment aim for some people who have SIRT, because they then might be able to have a liver transplant. For some people downstaging might have a large impact on quality of life. This is because of the potential for curative treatment. Both clinical experience and limited trial evidence (for example SARAH) show that downstaging is rare in advanced HCC. The committee understood that people whose tumour downstages have different subsequent treatments, and few might have a liver transplant. It was unclear whether people who have a liver transplant after downstaging of their tumour have similar outcomes to those who have a liver transplant

without the need for downstaging. The committee reconsidered downstaging after consultation and during its third meeting. It concluded that downstaging may be an option for a small proportion of people with advanced HCC. However, the proportion of people who have tumours that downstage, and the subsequent outcomes, are uncertain. Therefore, downstaging was not included in the base-case model.

SIRTs may have fewer and less severe adverse events than sorafenib and these have not been captured in the economic modelling

3.38 Both the SARA and SIRveNIB trials collected data on health-related quality of life. SARA used the European Organisation for Research and Treatment of Cancer Quality-of-Life Questionnaire Core 30 (EORTC-QLQ-C30) questionnaire. The company mapped this onto the EQ-5D scale using the Longworth et al. algorithm. The AG used these estimates in its model. The committee noted that utility values were similar between SIRTs and sorafenib for the following disease states: progression-free survival, progressive disease and after transplant. There were only small differences in utilities between progression-free survival and progressive disease. The clinical experts explained that people who had sorafenib for a long time may have a long-lasting negative effect on their quality of life. SIRTs are given in 1 procedure, meaning there is a shorter duration of effect on health-related quality of life. The committee was concerned that the potential important differences in long-term quality of life might not be captured in clinical trial results because quality-of-life data are collected at fixed time points (3, 6, 9 and 12 months after randomisation). It noted that in SARA, EORTC-QLQ-C30 values in the SIR-Spheres arm were relatively constant over the 12 months from randomisation. Values for people in the sorafenib arm worsened for 6 months then improved. This was not seen in the mapped EQ-5D values. The committee acknowledged that this might be because the EORTC-QLQ-C30 scale is more sensitive than the EQ-5D to adverse events associated with sorafenib (fatigue, diarrhoea and skin reactions).

The committee was also aware that the mapping algorithm did not include

data from people with HCC, meaning that differences important to people with HCC might not accurately translate across to the EQ-5D. The committee recalled its conclusion that SIRT may have fewer and less severe adverse events than sorafenib (see section 3.27). It concluded that some aspects of health-related quality of life might not be captured in the utility values.

Adverse event disutility values should be included in the model to capture differences in quality of life between SIRT and sorafenib

3.39 The clinical experts advised that the side-effect profiles of SIRT and sorafenib were different and should result in improved health-related quality of life for SIRT compared with sorafenib. The committee understood that some people stop taking sorafenib because of intolerable adverse events. After consultation and additional analysis by the AG, the committee considered analyses applying disutility values for adverse events of grade 3 and above, and for adverse events of any grade. Various assumptions were included about the effect of less severe (grade 1 and 2) events. SARAH provided data on adverse event rates and pooled event duration (see section 3.27). The disutility values were informed by previous NICE technology appraisals. The committee understood that these values came from primary studies of variable quality, including vignette studies which are less robust. In these additional analyses, the smallest incremental quality-adjusted life year (QALY) gain from adverse events for SIRT compared with sorafenib resulted in no additional total QALYs, when assuming that health-state utility values adequately captured all adverse effects. The biggest gain was 0.120 QALYs, when applying event-specific disutility values for events regardless of their severity (grades 1 and above). The committee agreed that there is some QALY gain with SIRT resulting from the fewer and less severe adverse events. However, it also agreed that it was inappropriate to assume grade 1 and 2 events have the same effect on quality of life as grade 3 and 4 events. It also noted that typically, only grade 3 or 4 events are included in cost-effectiveness analyses.

Final appraisal document – Selective internal radiation therapies for treating hepatocellular carcinoma

Therefore, it agreed that a gain of 0.120 QALYs would be too optimistic. It agreed that an intermediate adverse event QALY gain would be appropriate. The committee concluded that an adverse event-related QALY gain of 0.047 for SIRTs compared with sorafenib might be plausible, and should be included in the base-case analysis. It also concluded that there was high uncertainty associated with this estimate and that the uncertainty was highest for QuiremSpheres because of its limited data.

Cost-effectiveness results

SIR-Spheres and TheraSphere are a cost-effective use of NHS resources for HCC

3.40 The committee agreed that its preferred approach to modelling included:

- identical procedure-related administration costs for all SIRTs
- individual participant data from SARA for duration of sorafenib and dose and duration of regorafenib
- an additional SIRT QALY gain of 0.047 to account for differences in adverse events compared with sorafenib.

The economic analysis included the committee's preferred assumptions and confidential patient access schemes for QuiremSpheres, SIR-Spheres, TheraSphere, regorafenib and sorafenib. It assumed that the 3 SIRTs had the same effectiveness. It showed that all SIRTs were less effective than sorafenib despite the additional SIRT QALY gains to account for differences in adverse events, giving 0.029 fewer QALYs overall. QuiremSpheres was more costly than sorafenib. SIR-Spheres and TheraSphere were less costly than sorafenib and provided fewer QALYs. Because of confidential discounts for interventions, comparator and follow-on therapies, exact cost-effectiveness results cannot be reported here. The AG also presented extensive scenario analyses during the first committee meeting and after consultation. This included:

- alternative functions to model overall survival and progression-free survival (see section 3.34)
- alternative costs and utility values
- ALBI grade 1 and low tumour burden subpopulation (see section 3.35)
- retrospective studies with high risk of bias (see section 3.15)
- downstaging (see section 3.37).

Alternative functions, costs and utility values did not have a great effect on the incremental cost-effectiveness ratios (ICERs). The committee agreed that scenarios that restricted the population to people with ALBI grade 1 and low tumour burden were not plausible because this population is not relevant to NHS practice in England (see section 3.35). It also agreed that retrospective studies should not be included because of high risk of bias and uncertainty of the data (see section 3.15). Additionally, downstaging should not be included in the committee's preferred base case because the proportion of people who have tumours that downstage, and subsequent outcomes, are uncertain (see section 3.37). The committee concluded that in the probabilistic base-case analysis, QuiremSpheres was less effective and more costly than sorafenib. This meant sorafenib dominated QuiremSpheres (that is, it was more effective and less costly). SIR-Spheres and TheraSphere were less effective and less costly than sorafenib. The cost savings were sufficient to offset the QALY loss at a £30,000 saved per QALY lost level. The committee also recalled that the model assumed that SIRTs were equally effective, and that this was a highly uncertain assumption for QuiremSpheres because of its very limited evidence base compared with SIR-Spheres and TheraSphere (see section 3.33). The committee also recalled that personalised dosimetry could improve the effectiveness of SIRTs, which may increase their QALYs (see section 0). It considered that this would not meaningfully affect the cost-effectiveness estimate for QuiremSpheres or offset the uncertainty in its evidence base. Because of its higher costs compared with sorafenib and its limited clinical evidence, the committee considered QuiremSpheres not to be a cost-effective use of NHS resources for

treating HCC. Because of the cost savings per QALYs lost, the committee considered that both SIR-Spheres and TheraSphere are cost-effective use of NHS resources.

End of life

The end of life criteria are not met

3.41 The committee considered the advice about life-extending treatments for people with a short life expectancy in [NICE's guide to the methods of technology appraisal](#).

- When transplant or CTT is appropriate, people have a life expectancy of more than 24 months. This means that the life-expectancy criterion (that is, the treatment is indicated for patients with a short life expectancy, normally less than 24 months) was not met for these subgroups.
- When CTT is inappropriate, in advanced disease, people have a poor prognosis with a life expectancy of less than 24 months. Therefore, the short life-expectancy criterion was met for this subgroup.
- In all plausible scenarios, there was no increase in the modelled undiscounted life expectancy with SIRTs compared with sorafenib. The committee concluded that the life-extending criterion (that is, there is sufficient evidence that the treatment could extend life, normally by a mean value of at least an additional 3 months, compared with current NHS treatment) was not met.

Because both parts of the criteria were not met, the committee concluded that the end-of-life criteria were not met.

Innovation

No evidence was identified showing additional benefits of SIRT, above those captured in the cost-effectiveness analysis

3.42 The companies considered SIRTs to be innovative because they offer a more personalised treatment option. The patient experts stated that SIRTs would be a substantial change in treating HCC because they could offer a chance for subsequent curative treatment for people who would not otherwise have this option. The committee concluded it had not seen evidence of any additional benefits that were not captured in the measurement of QALYs in its preferred model.

Conclusion

SIR-Spheres and TheraSphere, but not QuiremSpheres, are recommended for treating HCC

3.43 In the committee's preferred analysis, SIRTs were less effective than sorafenib with an incremental quality-adjusted life years loss of 0.029 QALYs. In clinical trials, SIR-Spheres did not improve survival compared with sorafenib. There was very limited clinical evidence to compare the effectiveness of QuiremSpheres and TheraSphere with sorafenib and for all the SIRT technologies. The committee also considered that there were limited data for QuiremSpheres and TheraSphere, and the effectiveness estimates and resulting ICERs were more uncertain than those of SIR-Spheres. However, the committee considered that the adverse event profiles of SIRTs and sorafenib are different and people with HCC would welcome new treatment options (see section 3.1). Taking this into account, the committee concluded that QuiremSpheres was less effective and more costly than sorafenib and was not considered a cost-effective use of NHS resources. QuiremSpheres was not recommended for HCC. SIR-Spheres and TheraSphere were less costly than sorafenib and the estimated cost savings outweigh the loss of QALYs after taking into account the

uncertainty associated with the clinical effectiveness. Therefore, the committee considered SIR-Spheres and TheraSphere to be a cost-effective use of NHS resources, and recommended both as options for treating advanced HCC for people with Child-Pugh grade A liver impairment for whom CTT is inappropriate.

4 Implementation

- 4.1 [Section 7\(6\) of the National Institute for Health and Care Excellence \(Constitution and Functions\) and the Health and Social Care Information Centre \(Functions\) Regulations 2013](#) requires clinical commissioning groups, NHS England and, with respect to their public health functions, local authorities to comply with the recommendations in this appraisal within 3 months of its date of publication.
- 4.2 The Welsh ministers have issued directions to the NHS in Wales on implementing NICE technology appraisal guidance. When a NICE technology appraisal recommends the use of a drug or treatment, or other technology, the NHS in Wales must usually provide funding and resources for it within 2 months of the first publication of the final appraisal document.
- 4.3 When NICE recommends a treatment 'as an option', the NHS must make sure it is available within the period set out in the paragraphs above. This means that, if a patient has hepatocellular carcinoma and the doctor responsible for their care thinks that SIR-Spheres or TheraSphere is the right treatment, it should be available for use, in line with NICE's recommendations.

5 Review of guidance

- 5.1 The guidance on this technology will be considered for review 3 years after publication. The guidance executive will decide whether the technology should be reviewed based on information gathered by NICE, and in consultation with consultees and commentators.

Stephen O'Brien
Chair, appraisal committee
February 2021

6 Appraisal committee members and NICE project team

Appraisal committee members

The 4 technology appraisal committees are standing advisory committees of NICE. This topic was considered by [committee C](#).

Committee members are asked to declare any interests in the technology to be appraised. If it is considered there is a conflict of interest, the member is excluded from participating further in that appraisal.

The [minutes of each appraisal committee meeting](#), which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Committee members are asked to declare any interests in the technology to be appraised. If it is considered there is a conflict of interest, the member is excluded from participating further in that appraisal.

NICE project team

Each technology appraisal is assigned to a team consisting of 1 or more health technology analysts (who act as technical leads for the appraisal), a technical adviser and a project manager.

Verena Wolfram

Technical lead

Jamie Elvidge

Technical adviser

Louise Jafferally

Project manager

ISBN: [to be added at publication]