Evidence report for local therapy vs. systemic therapy in patients with hepatocellular carcinoma and cirrhosis who are ineligible for resection or transplantation

Prepared by EBSCO Health Innovations and EBM Development Department (in alphabetic order):

Brian S. Alper, MD, MSPH, FAAFP, FAMIA, Chief Medical Knowledge Officer

Eric Manheimer, PhD, EBM Editor

Martin Mayer, DMSc, MS, PA-C, Senior Clinical/Medical Editor

Table of Contents

*Note that each *section* of the report also has *section-specific* page numbering at the bottom-right corner of the page. The above page numbers refer to the *report-specific* page number, which appears in bolded and italicized font in the top-right corner of the page.

Evidence report for local therapy vs. systemic therapy in patients with hepatocellular carcinoma and cirrhosis who are ineligible for resection or transplantation

Scoping

EBSCO Information Services

May 13, 2020
Updated May 27, 2020 after feedback from and discussion with UKSH
Final approval/agreement June 11, 2020

Prepared by EBSCO Health Innovations and EBM Development Department (in alphabetic order):

Brian S. Alper, MD, MSPH, FAAFP, FAMIA, Chief Medical Knowledge Officer

Eric Manheimer, PhD, EBM Editor

Martin Mayer, DMSc, MS, PA-C, Senior Clinical/Medical Editor

Abbreviations

General abbreviations that may appear in this report

ARD – absolute risk difference (sometimes simply abbreviated as "RD" for "risk difference")
CI – confidence interval
CrI – credible interval
HR – hazard ratio
IQR – interquartile range
ITT – intention-to-treat
MA – meta-analysis
MD – mean difference
NA, N/A – not applicable
NMA – network meta-analysis
NR – not reported
OR – odds ratio
RCT – randomized, controlled trial
RD – risk difference (sometimes abbreviated as "ARD" for "absolute risk difference")
RR – risk ratio, relative risk
RtR – rate ratio
SD – standard deviation
SE – standard error
SMD – standardized mean difference
SR – systematic review
SRMA – systematic review and meta-analysis

Abbreviations specific to this report that may appear in this report

BCLC – Barcelona-Clinic Liver Cancer
ECOG – Eastern Cooperative Oncology Group
EORTC QLQ-C30 – European Organization for Research and Treatment of Cancer Quality of Life Questionnaire
EQ-5D – EuroQol 5-dimension quality of life scale
HCC – hepatocellular carcinoma
RECIST – Response Evaluation Criteria in Solid Tumors
SBRT – stereotactic body radiotherapy
SIRT – selective internal radiotherapy; also referred to as transarterial radioembolization (TARE)
TACE – transarterial chemoembolization
TARE – transarterial radioembolization; also referred to as selective internal radiotherapy (SIRT)

Note: We use the terms "radioembolization", "selective internal radiation therapy", "selective internal radiotherapy", and "SIRT" interchangeably in this evidence report.

Methods

Using an established, structured template that delineates key elements for scoping (i.e., the population, interventions, and key outcomes to consider), we collaboratively agreed with UKSH on the scope of the present evidence report. Scoping was completed on May 13, 2020 and updated on May 27 after feedback from UKSH.

Where outcomes will be answered with systematic searches, critical appraisal, and evidence synthesis, we will state the intended approach after we state the outcome(s) of interest for that FAQ. For example, we might say: "The outcome of interest is all-cause mortality. We will obtain effect estimates from systematic reviews and meta-analyses of randomized, controlled trials or from individual trials if systematic reviews do not exist or meta-analysis is not the most appropriate method of synthesis." Otherwise, where we say: "We will include a description of", "We will include descriptive responses", or something similar, we will answer the FAQ with descriptive responses, which do not involve systematic evidence searches, critical appraisal, or evidence synthesis.

Results – Criteria Selection for Critical Appraisal

Population:

Inclusion criteria

Patients with intermediate to advanced HCC (BCLC stage B or C) and Child-Pugh A or B liver disease/cirrhosis who are not eligible for surgical resection or transplantation.

For patients with advanced liver cirrhosis or high tumor burden, there is insufficient healthy residual tissue for resection as a viable option. Liver transplantation is also contraindicated in advanced HCC because of tumor size, vascular invasion, distant metastases, and/or macrovascular invasion.

Exclusion criteria

Patients who do not have cirrhosis, otherwise have less severe disease than that noted in 'Inclusion criteria', or are eligible for surgical resection or transplantation.

Intervention and Comparator/Options:

Best supportive care

Two systematic reviews published in 2017 and 2018 (Roccarina 2017, Finn 2018) yield a total of 2 trials that compare a targeted intervention vs. placebo: Llovet 2008 and Cheng 2009. Given the use of best supportive care as a background intervention in such trials (it would be unethical to withhold best supportive care, irrespective of assignment), placebo assignment in these trials really constitutes receiving best supportive care only. In both trials, sorafenib was the targeted intervention, and a forward search from the closure of the search window for these systematic reviews finds no subsequent trials of sorafenib vs. placebo. Accordingly, and because (1) much of the comparative evidence for local therapy vs. systemic therapy (as defined below) focuses on sorafenib as the systemic therapy and (2) three-arm trial evidence is lacking, we will obtain estimates for the option of best supportive care from trials of sorafenib vs. placebo.

However, even if only a consequence of random variation/random error present in the available evidence, simply reporting the absolute outcomes in the best supportive care arms of the above-noted trials could lead to misleading comparisons if these estimates are then compared against the data for local therapy vs. systemic therapy. The framework of this evidence report and resultant decision aid suggests exactly this kind of comparison. Therefore, if necessary and appropriate to increase the comparability of the absolute results among the options, we may use ratio measures as part of obtaining absolute estimates (given ratio measures tend to have the advantage of being more 'portable' than absolute measures). For instance, consider a hypothetical scenario where the collective trial evidence for treatment A vs. placebo suggests an overall median survival of 10 months vs. 7 months. Consider also that the collective trial evidence for treatment B vs. treatment A suggests an overall median survival of 10 months vs. 14 months. Simply presenting the results of the best supportive care group (7 months) along with the data for the comparison of treatment B (10 months) vs. treatment A (14 months) would be misleading. Put simply, this is in large part due to the apparent variation in survival in the treatment A group in each body of evidence (10 months vs. 14 months), even though the patients in the treatment A group are presumably drawn from the same population; in the absence of systematic differences between these bodies of evidence, this would be attributable to random variation/random error. To help with this obstacle, one can utilize ratio measures. In the trial evidence comparing treatment A vs. best supportive care, the simple ratio of medians suggests those in the best supportive care group had 70% of the median survival that treatment A group had. Considering now the comparative evidence for treatment B vs. treatment A, one could use the median survival in the treatment A group (14 months) as the 'reference' point to estimate the median survival in the best supportive care group based on the ratio measure for the comparison of best supportive care vs. treatment A (70% = 0.7): 0.7*14 = 9.8 ≈ 10 months. Thus, one might estimate a median survival of 10 months, 14 months, and 10 months for best supportive care, treatment A, and treatment B, respectively. Of note, if one instead used the estimates directly (7 months, 14 months, and 10 months, respectively), this would be misleading – in this case, the effect is to potentially exaggerate the benefit of treatment A and treatment B. Lastly, in the preceding example, the median survival estimate of 14 months was used as the 'reference' for treatment A, but the above approach could be carried out with a different estimate or a range of estimates if the estimate of 14 months was for some reason felt not to be the best representation of the median survival one might expect with treatment A.

Local therapy

Local therapy is a minimally invasive, non-surgical treatment for HCC. It involves different locoregional liver-directed approaches that take a percutaneous (radiofrequency or microwave ablation) or transarterial (chemoembolization or radioembolization) approach. Of note, transarterial radioembolization is also often referred to as selective internal radiation therapy. Local therapy may also include stereotactic body radiotherapy (which may be referred to by different terms, such as linear accelerator [LINAC] radiotherapy, CyberKnife®, Varian®); it is also a percutaneous procedure in the strictest sense of the therapy occurring through the skin, it is categorically different from the traditional meaning of a 'percutaneous' procedure (there is no incision).

The comparison of interest is systemic therapy, as defined below.

Systemic therapy

Systemic therapy involves use of medications such as cytotoxic agents, molecularly targeted therapies, and immunotherapy agents. This may include such agents as sorafenib, regorafenib, lenvatinib, cabozantinib, ramucirumab, or atezolizumab + bevacizumab. Other therapies are also currently under evaluation.

The comparison of interest is local therapy as defined above.

Special consideration of the IMbrave150 trial separate from formally-scoped options above

Of note, the recently-published IMbrave150 trial compared sorafenib to atezolizumab + bevacizumab (Finn 2020). Finn 2020 notes sorafenib and levatinib are currently the first-line therapies on the basis of studies showing "modestly longer survival with sorafenib than with placebo and noninferiority of lenvatinib to sorafenib." Indeed, our scoping review suggests the randomized trial evidence focuses heavily on sorafenib. This makes the IMbrave150 trial important, because it has the potential to influence what is considered the first-line approach if electing to pursue systemic therapy. Accordingly, we will summarize the IMbrave150 trial separately within this evidence report as part of Critical Appraisal, and based on our appraisal of the trial, we may note its potential to impact the standard of care for systemic therapy within Evidence Synthesis; however, given the difference between the comparison in the IMbrave150 trial (systemic therapy vs. systemic therapy) and the comparisons of interest for this evidence report (as defined above), we will not attempt to make indirect comparisons within Evidence Synthesis based on the findings of the IMbrave150 trial.

Outcomes:

FAQ 1: What does the treatment involve?

We will include a description of the treatment or procedure, such as frequency of dosing, monitoring, and length of stay, as relevant.

Descriptive responses are acceptable, and this does not involve systematic evidence searches, critical appraisal, or evidence synthesis.

FAQ 2: How long will I live?

The outcome of interest is overall survival (OS).

We will obtain estimates from randomized, controlled trials or systematic reviews thereof. We will give preference to OS estimates expressed as median OS. We will consider OS rates at defined timepoints if median OS is not reported or if the preponderance of evidence reports OS rates at defined timepoints instead of median OS. As a complementary measure to median OS, we may also report OS rates at a defined timepoint (e.g., as assessed from reported OS rates or as extracted from OS curves) if sufficient evidence exists, with the timepoint depending on the nature of the data (e.g., if OS curves extend to 3 years but have a paucity of patients at risk at that point, we may instead extract information for an earlier timepoint when more patients are still at risk).

Although there may be an interest in cause-specific mortality, this is a less robust outcome measure than overall survival in that it suffers from greater methodologic and clinical issues than overall survival. For instance, an intervention may prevent one kind of cause-specific death while contributing to another; in other words, a patient could avoid death from cause X but die from cause Y where X is related to disease progression and Y is related to treatment. Overall survival intrinsically accounts for such tradeoffs. Additionally, our scoping review suggests cause-specific mortality is not typically reported in this literature base. For these reasons, we did not include it as an outcome.

FAQ 3: How long will my cancer stay controlled?

The outcome of interest is progression-free survival (PFS).

We will address this FAQ via the approach noted for OS. We note that, although PFS is a very common outcome in oncology literature, PFS is ultimately a surrogate marker; that is, a change in PFS is not necessarily linked to a change in outcomes like OS or quality of life. Accordingly, we will include language to this effect in our report. Another outcome that might be of interest within the broad construct this FAQ seeks to address is remission. However, akin to what we found for the outcome of cause-specific mortality during our scoping review, it appears the outcome of remission is not typically reported in the applicable literature. Additionally, we note that one would reasonably expect the outcome of remission to be related to overall survival and quality of life, and these outcomes may be more important for decision-making (i.e., the combination of these two outcomes may render unimportant the outcome of whether remission is achieved).

FAQ 4: Will the treatment change my quality of life?

The outcome of interest is quality of life (QoL)/health-related quality of life (HRQoL), ideally measured using validated scales such as the EQ-5D or EORTC QLQ-C30.

We will obtain estimates from randomized, controlled trials or systematic reviews thereof. We will focus on between-groups differences for assessment of comparative efficacy. Given the comparison in this report is not against placebo/best supportive care (but rather, against another specific treatment), patients may also be interested in how their quality of life may change over time (rather than just how the two options compare) if the available evidence cannot distinguish any difference in quality of life between the options. For instance, if quality of life goes down on average and the therapies do not have a clear difference in quality of life, this carries different meaning than if quality of life does not change on average and the therapies do not have a clear difference in quality of life. Accordingly, if permitted by the available evidence, we may try to give an indication of how quality of life changes from baseline if the available evidence cannot distinguish between the two options. However, we will not consider within-group differences from baseline in a comparative manner.

FAQ 5: What are the risks or side effects?

The outcomes of interest are treatment-related serious adverse events and treatment discontinuation due to adverse events. The relevance of the former may be more readily apparent, and the latter permits insight into the tolerability of the option. Of note, however, treatment discontinuation due to adverse events may not provide meaningful insights for the option of local therapy given the nature of its implementation (e.g., if an ablative or embolizing procedure is done once) compared to best supportive care and systemic therapy (which are ongoing therapies).

As part of informing the option of best supportive care, we will include named adverse events that occurred in ≥2% of the best supportive care group; to avoid misleading estimates and to provide insight into whether meaningful differences exist in these adverse events compared to targeted intervention, we will also assess the event rate in the intervention group for these named adverse events.

For adverse events with targeted interventions, we will include named adverse reactions that showed clear separation from the comparator.

We will obtain estimates from randomized, controlled trials or systematic reviews thereof. We may supplement with descriptive responses if needed.

FAQ 6: If I have other health issues, do they change how the treatment works?

The outcome of interest is whether comorbidities may lead to heterogeneity of treatment effect for either efficacy or harm.

We will obtain estimates from randomized, controlled trials or systematic reviews thereof. Given subgroup analyses sacrifice the virtues of randomization (barring a design that stratifies randomization on the subgroup variable of interest) and can often lead to spurious 'findings', they are to be viewed as hypothesis generating in the absence of rigorous verification (Oxman 1992, Sun 2012, Sun 2014, Wallach 2017). We will thus approach any subgroup analysis with commensurate scrutiny; the inferences drawn – if any – may be limited to qualitative statements and coupled with statements of uncertainty.

Cheng AL, Kang YK, Chen Z, Tsao CJ, Qin S, Kim JS, Luo R, Feng J, Ye S, Yang TS, Xu J, Sun Y, Liang H, Liu J, Wang J, Tak WY, Pan H, Burock K, Zou J, Voliotis D, Guan Z. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. Lancet Oncol. 2009 Jan;10(1):25-34. doi: 10.1016/S1470-2045(08)70285-7. Epub 2008 Dec 16. PubMed

Finn RS, Qin S, Ikeda M, et al. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. N Engl J Med. 2020;382(20):1894-1905. doi:10.1056/NEJMoa1915745. PubMed

Finn RS, Zhu AX, Farah W, et al. Therapies for advanced stage hepatocellular carcinoma with macrovascular invasion or metastatic disease: A systematic review and meta-analysis. Hepatology. 2018;67(1):422-435. doi:10.1002/hep.29486. PubMed

Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, de Oliveira AC, Santoro A, Raoul JL, Forner A, Schwartz M, Porta C, Zeuzem S, Bolondi L, Greten TF, Galle PR, Seitz JF, Borbath I, Häussinger D, Giannaris T, Shan M, Moscovici M, Voliotis D, Bruix J; SHARP Investigators Study Group. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med. 2008 Jul 24;359(4):378-90. doi: 10.1056/NEJMoa0708857. PubMed

Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med. 1992 Jan 1;116(1):78-84. doi: 10.7326/0003-4819-116-1-78. PubMed

Roccarina D, Majumdar A, Thorburn D, Davidson BR, Tsochatzis E, Gurusamy KS. Management of people with intermediate-stage hepatocellular carcinoma: an attempted network meta-analysis. Cochrane Database Syst Rev. 2017;3(3):CD011649. Published 2017 Mar 10. doi:10.1002/14651858.CD011649.pub2. PubMed

Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, Bala MM, Bassler D, Mertz D, Diaz-Granados N, Vandvik PO, Malaga G, Srinathan SK, Dahm P, Johnston BC, Alonso-Coello P, Hassouneh B, Walter SD, Heels-Ansdell D, Bhatnagar N, Altman DG, Guyatt GH. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ. 2012 Mar 15;344:e1553. doi: 10.1136/bmj.e1553. PubMed

Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. JAMA. 2014 Jan 22-29;311(4):405-11. doi: 10.1001/jama.2013.285063. PubMed

Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JP. Evaluation of Evidence of Statistical Support and Corroboration of Subgroup Claims in Randomized Clinical Trials. JAMA Intern Med. 2017 Apr 1;177(4):554-560. doi: 10.1001/jamainternmed.2016.9125. PubMed

Evidence report for local therapy vs. systemic therapy in patients with hepatocellular carcinoma and cirrhosis who are ineligible for resection or transplantation

Evidence Synthesis

EBSCO Information Services

July 17, 2020

Prepared by EBSCO Health Innovations and EBM Development Department (in alphabetic order):

Brian S. Alper, MD, MSPH, FAAFP, FAMIA, Chief Medical Knowledge Officer

Eric Manheimer, PhD, EBM Editor

Martin Mayer, DMSc, MS, PA-C, Senior Clinical/Medical Editor

Methods

Where evidence is very limited, or the FAQ is scoped as a descriptive response, such that no evidence synthesis or processing is applied, the descriptive summary/-ies will be given.

For each FAQ for which evidence synthesis or processing is considered, we will summarize the key concepts (results and factors affecting certainty of the results) from the included studies that are most relevant to evidence synthesis; we will prefer using a tabular format for this summary in most cases, but depending on the circumstances, we may adopt a non-tabular form instead. For each outcome, we will then articulate the method selected for evidence synthesis and justification of that method; again, we will prefer a tabular approach in most cases, but we may articulate the justification in a non-tabular form depending on the circumstances. We will do this for the effect estimate (i.e., answering the question "Is there a difference between the intervention and comparator?") and results will be presented with GRADE-based certainty of evidence ratings related to the certainty of effect estimates. For FAQs which require noncomparative/option-specific responses (i.e., the FAQ is "What will happen with this option?" rather than "How do things change with this option?") the evidence synthesis process will be extended to include the method to produce noncomparative responses.

For summary statements representing a plain language summary of evidence appraisal and synthesis
We will include the following 2 items immediately after the summary statement as a parenthetical, regardless of whether they occur in the Summary of Findings section or the individual FAQs in the body of the Evidence Synthesis document:

- our certainty in the summary statement (categorized as High, Moderate, Low, or Very low), and
- the number and type of studies informing the statement, as well as a label for the studies in "[Author] [Year]" format.

If the study informing the summary statement is a systematic review and meta-analysis, then unless there are compelling reasons to proceed in a different manner, we will cite the systematic review and meta-analysis as the source. We will do this even if the result on which our summary statement is based comes from a subset of studies in that systematic review and meta-analysis (e.g., a meta-analysis of 5 out of 10 studies, because only 5 of the 10 reported on the outcome of interest). If EBSCO Information Services conducts its own meta-analysis of studies to inform a summary statement, we will list the individual studies.

For summary statements representing a plain language summary of a descriptive response/summary
Certainty rating and provision of details about contributing studies do not apply. Accordingly, we will not include a parenthetical at the end of the summary statement. However, when descriptive summaries appear in the Summary of Findings section, we will include a general parenthetical citation of applicable references. We will place this at the end of the relevant FAQ in the Summary of Findings section. For summary statements occurring in the body of the Evidence Synthesis document, placement will depend on the context.

Results – Summary of Findings

FAQ 1: What does the treatment involve?

Radioembolization

Radioembolization tries to slow or kill the cancer. You will have some tests to make sure you can have this treatment. One of the tests is called an "angiogram". In this test, a small cut is made in your wrist or groin area. A small tube called a "catheter" is put through the cut into a blood vessel. The tube is moved through your blood vessels to your liver. X-rays help make sure the tube goes the right way. Once the tube is where it needs to be, the clinician will inject a dye to help make your blood vessels easier to see. This can make you feel warm and like you need to urinate. The test takes about an hour or so to do. The area where the tube is inserted will be numbed, and many people are given medicine to make them sleepy as well. You may have to stay in the hospital for one night after having this done.

About a week or two later, you will have the radioembolization treatment. This means having another angiogram. But this time, when the tube is in the right spot, the clinician will put small beads in the blood vessel that goes to the liver. The beads give off radiation to try to slow or kill the cancer. The level of radiation goes down a lot the farther away something is from the bead. This helps protect healthy tissue, but the healthy tissue may still get some radiation. You may need to stay in the hospital for one night again.

Depending on how much cancer is in your liver, you may only need one radioembolization treatment. But some people need two radioembolization treatments. If you need two treatments, they are often given about a month or so apart.

You might be told you should be careful around children or pregnant women for a week or so after having radioembolization. But this may not be needed. You can talk with your health care team more about this if you are thinking about this option.

You will also get the treatments in the "Best supportive care" option.

Sorafenib

Sorafenib is a medicine that tries to slow or kill the cancer. You will take a pill twice a day, usually about an hour or two after eating. You will also get the treatments in the "Best supportive care" option.

Best supportive care

Best supportive care does not try to slow or kill the cancer. Instead, it focuses on making sure your symptoms are treated as much as possible. This can include things like treating pain, anxiety, depression, and feeling tired. Treatment may use medication, therapy, or procedures, like having a blood transfusion. Best supportive care varies for each person, because each person has different needs. A person can have best supportive care alone or in addition to treatment that tries to slow or kill the cancer, and there is no "right" or "wrong" choice. Either choice is okay, and the best choice is the choice that you think best matches what you want.

(Sources considered for these descriptive responses include: Cancer Research UK 2019, Cedars-Sinai ND, Cleveland Clinic 2019, DynaMed 2020, Mayo Clinic 2019, Memorial Sloan Kettering Cancer Center 2019)

FAQ 2: How long will I live?

Radioembolization vs. sorafenib

On average, people who have treatment with radioembolization or sorafenib may live for about 9 months. Some people will live for a shorter time. Some people will live for a longer time. (Moderate certainty; based on 4 trials [Cheng 2009, Chow 2018, Llovet 2008, Vilgrain 2017])

There may be a treatment that is kind of like radioembolization that still works somewhat better than radioembolization. This treatment uses something called “chemoembolization”, and it also uses radiation outside the body. (Very low certainty; based on 1 trial [Yoon 2018]) There may also be a treatment that is kind of like sorafenib that still works somewhat better than sorafenib. This treatment uses atezolizumab and bevacizumab and works differently than sorafenib. (Moderate certainty; based on 1 trial [Finn 2020]) You can talk about this with your clinician if you are interested.

Best supportive care

On average, people who have treatment with best supportive care alone may live for about 6 months. Some people will live for a shorter time. Some people will live for a longer time. (Moderate certainty; based on 2 trials [Cheng 2009, Llovet 2008])

FAQ 3: How long will my cancer stay controlled?

Radioembolization vs. sorafenib

On average, people who have treatment with radioembolization or sorafenib may have their cancer controlled for about 4 to 5 months. Some people’s cancer will be controlled for a shorter time. Some people’s cancer will be controlled for a longer time. (Low certainty; based on 4 trials [Cheng 2009, Chow 2018, Llovet 2008, Vilgrain 2017]) Your cancer being controlled does not mean you will not have symptoms, but best supportive care treatments may be able to help with symptoms you may have.

There may be a treatment that is kind of like radioembolization that still works somewhat better than radioembolization. This treatment uses something called “chemoembolization”, and it also uses radiation outside the body. (Very low certainty; based on 1 trial [Yoon 2018]) There may also be a treatment that is kind of like sorafenib that still works somewhat better than sorafenib. This treatment uses atezolizumab and bevacizumab and works differently than sorafenib. (Very low certainty; based on 1 trial [Finn 2020]) You can talk about this with your clinician if you are interested.

Best supportive care

On average, people who have treatment with best supportive care alone may have their cancer controlled for about 2 months. Some people’s cancer will be controlled for a shorter time. Some people’s cancer will be controlled for a longer time. (Low certainty; based on 2 trials [Cheng 2009, Llovet 2008]) Your cancer being controlled does not mean you will not have symptoms, but best supportive care treatments may be able to help with symptoms you may have.

FAQ 4: Will the treatment change my quality of life?

Radioembolization vs. sorafenib vs. placebo

Some people may feel like their quality of life goes down some over time. But other people may not feel this way. Research does not show any difference in quality of life based on whether someone gets radioembolization, sorafenib, or best supportive care alone. But there is not much research in this area. (Low to Very low certainty; based on 2 trials [Cheng 2009, Chow 2018]

FAQ 5: What are the risks or side effects?

Radioembolization

In general, people often find this treatment to be tolerable. Part of this may be because it is often given in 1 or 2 sessions rather than something they do every day. Because of this, stopping treatment due to side effects may be less of a concern compared with sorafenib.

(Certainty not rated; descriptive response here based on adaptation of Vilgrain 2017)

Research gives different estimates about how often serious side effects occur. They may happen in less than 1 out of 100 people (<1%) or in as many as 16 out of 100 people (16%). One study suggests this number may be higher, but it is not clear how much of that number was really because of radioembolization.

(Moderate certainty; based on 3 trials [Cheng 2009, Chow 2018, Vilgrain 2017])

As with any treatment like this, there is a risk of infection and bleeding, but these things are not common. Other risks that can be serious but are not common include stomach ulcer and inflammation of the liver from the treatment. Some people can also have symptoms like fatigue or loss of appetite, but these things may be due to the illness itself and are probably less likely to happen compared with someone who takes sorafenib. You can talk more about these and other possible risks or side effects with your clinician.

(Certainty not rated; based on Cheng 2009, Chow 2018, Llovet 2008, and Vilgrain 2017)

Sorafenib

Out of 100 people who take sorafenib, about 2 to 11 (2% to 11%) may stop taking it because of side effects. One study suggests this number may be higher, but it is not clear how much of that number was really because of sorafenib.

(Low certainty; based on 2 trials [Llovet 2008, Vilgrain 2017])

Research gives different estimates about how often serious side effects occur. They may happen in 1 out of 100 people (1%) or in as many as 13 out of 100 people (13%). One study suggests this number may be higher, but it is not clear how much of that number was really because of sorafenib.

(Moderate certainty; based on 3 trials [Cheng 2009, Chow 2018, Vilgrain 2017])

Sorafenib may cause loss of appetite, diarrhea, fatigue, hair thinning or hair loss, increased blood pressure, and different kinds of skin rashes or reactions. Sometimes the skin rashes or reactions can be serious. Although some people may have fatigue or loss of appetite from the illness itself, these symptoms probably occur more often in people taking sorafenib. You can talk more about these and other possible risks or side effects with your clinician.

(Certainty not rated; based on Cheng 2009, Chow 2018, Llovet 2008, and Vilgrain 2017)

Best supportive care

Your cancer may spread more easily without treatment that tries to kill or slow it. Because of this, some people might initially want best supportive care alone but later decide they want specific treatment. If this happens, you can discuss your options with your clinician at that time.

People may have loss of appetite or fatigue from the illness itself, but these things are probably less likely to happen compared with taking sorafenib.

(Certainty not rated; based on Cheng 2009, Chow 2018, Llovet 2008, and Vilgrain 2017)

FAQ 6: If I have other health issues, do they change how the treatment works?

Radioembolization vs. sorafenib vs. placebo

If you have other health issues, you can talk with your clinician about whether one option might be better or worse for you based on things like the side effects or possible problems with other medications you might take.

(Certainty not rated; based on absence of evidence for heterogeneity for treatment effect based on comorbidities from 4 trials [Cheng 2009, Chow 2018, Llovet 2008, Vilgrain 2017] and basic clinical practice paradigm of taking side effects and medication interactions into account when helping a patient decide on the best treatment option for him/her)

Note: Because the sample decision aid was already lengthy given the complexity of this topic and because this bolded synthesis statement did not seem to add substantively to the sample decision aid beyond what a clinician could discuss at the bedside in general, we did not progress this outcome to the sample decision aid.

Search Summary

Number of articles screened and selected

	Screened	Selected*
1st-round (DynaMed)	73	6
2nd-round (Systematic review searches)	51	3
3rd-round (Original article searches)	102	1
4th-round (Citation tracing)	6	0
Total	*232*	*10†*

*Number shown is after deduplication

†Only 7 of the 10 appear in Critical Appraisal and Evidence Synthesis, as the utility of 3 (Finn 2018, Roccarina 2017, and Shao 2015) were strictly for citation tracing and/or serving as a timepoint for forward searching for more current systematic reviews and/or randomized trials

	Articles selected for evaluation (Author Year)
Systematic reviews	4 (Abdel-Rahman 2020, Finn 2018*, Roccarina 2017*, Shao 2015*)
Randomized, controlled trials	6 (Cheng 2009, Chow 2018, Finn 2020, Llovet 2008, Vilgrain 2017, Yoon 2018)
Other article types	0

*This reference only appears in Search and Selection, as its only purpose for this evidence report was for citation tracing and/or serving as a timepoint for forward searching for more current systematic reviews and/or randomized trials

Results – Synthesis Details

Background information

This evidence report was scoped to examine the evidence for the FAQs of interest for three broad approaches to management: watchful waiting/best supportive care, systemic therapy, and local therapy. As can be appreciated from Scoping, these approaches included various treatment options within each broader option, particularly for systemic therapy and local therapy. However, we also specified in Scoping that we would obtain evidence from randomized trials or systematic reviews thereof, specifically those comparing local therapy vs. systemic therapy and those comparing systemic therapy vs. placebo (with additional specification in the latter case that we would limit consideration to sorafenib vs. placebo based on the findings from our scoping review). In our full literature searches, we indeed found the randomized trial evidence in line with Scoping was comprised of sorafenib as the systemic therapy option and radioembolization/selective internal radiotherapy as the local therapy option. The exception to this pattern is Yoon 2018, which compared sorafenib vs. transarterial chemoembolization (TACE) plus external beam radiotherapy (EBRT). Importantly, however, Yoon 2018 was a very small (N = 90), single-center trial and its results have – as of yet – not been replicated. Consequently, the local therapy of TACE plus EBRT comprises a small (and as-of-yet unreplicated) minority of the randomized trial evidence investigating the broad comparison of local therapy vs. systemic therapy (which has invariably been sorafenib in trials to date). Accordingly, Yoon 2018 is featured in relatively minor ways throughout the evidence report; likewise, we restrict our descriptive summaries in FAQ 1 to radioembolization, sorafenib, and best supportive care, and in our bolded synthesis statements, we refer to local therapy as radioembolization and systemic therapy as sorafenib. To do otherwise (e.g., to leave the options broadly as "systemic therapy" and "local therapy" in the synthesis statements and resultant decision aid) would arguably be misleading. However, in select areas, we do allude to the fact that there may be other local therapies that patients might talk about with their clinician. For instance, we synthesize Yoon 2018's findings for overall survival (FAQ 2: How long will I live?) and progression-free survival (FAQ 3: How long will my cancer stay controlled?) in this manner.

The above notes do not address the IMbrave150 trial (Finn 2020) – a head-to-head comparison of atezolizumab plus bevacizumab vs. sorafenib – which we included on special request as an effort separate from the remainder of the formally-scoped efforts for this evidence report. We specified in Scoping that we would not consider Finn 2020 within Evidence Synthesis beyond noting its potential to impact the standard of care for systemic therapy. However, we did expand this scope somewhat to permit generic statements in the FAQs for overall survival (FAQ 2: How long will I live?) and progression-free survival (FAQ 3: How long will my cancer stay controlled?) given this seemed like a fairly natural extension in light of how we synthesized Yoon 2018's findings for these FAQs. We also note there is additional evidence examining the comparison of two modalities within a given broader option and/or the utility of adding broader options together (e.g., Katsanos 2017, Kudo 2018), evidence very recently presented at the 2020 virtual American Society of Clinical Oncology meeting (e.g., Bi 2020, Ikeda 2020), and evidence currently under study (e.g., the HIMALAYA trial, NCT03298451), but all such literature is beyond the scope of this evidence report.

Finally, please note we use the terms "radioembolization", "selective internal radiation therapy", "selective internal radiotherapy", and "SIRT" interchangeably in this evidence report.

FAQ 1: What does the treatment involve?

Given the nature of this FAQ and the approach to descriptive summaries, we give a single collective reference at the end of this FAQ.

Local therapy

Descriptive Summary (pre-synthesis and synthesis efforts not applicable for descriptive summaries)

Although there are various types of therapy for the intended population of this evidence report, as noted above (see "Background information"), the preponderance of the evidence concerns radioembolization, which is also referred to as transarterial radioembolization (TARE) and selective internal radiotherapy (SIRT). A closely related variant, called transarterial chemoembolization essentially does the same thing, but does so by using chemical agents rather than radioactive agents.

In radioembolization, the clinician uses a catheter to deliver radioactive microspheres (essentially radioactive beads) into an artery that feeds the liver. The microspheres become lodged in smaller vessels closer to the liver and then emit radiation locally. The radiation rapidly dissipates as a function of distance from the microsphere; this helps concentrate the radioactivity at the cancerous tissue while minimizing exposure of healthy surrounding tissue (e.g., stomach, intestines). However, no treatment is perfect, so surrounding healthy tissues will still be exposed to some degree of radiation.

Prior to having the radioembolization procedure, patients undergo an evaluation to plan for the radioembolization and ensure they are a good candidate for it. Among other tests, patients will have an angiogram. The catheter will be inserted via the wrist or thigh (groin) and advanced to the liver under fluoroscopic guidance. When in place, a dye will be injected to help visualize the vasculature; this often makes patients feel warm and may make them feel like they need to urinate. Patients are often sedated for the procedure, and they will also receive local anesthetic at the site of the catheter insertion. The angiogram itself may take about an hour or so on average, but in some cases may take longer. Depending on the patient's general health and the nature of the evaluation, patients often spend a night in the hospital.

The radioembolization procedure often occurs a week or so after the evaluation angiogram. The radioembolization procedure involves a repeat angiogram, with the treatment occurring as noted above. Again, it may take about an hour or so. People often stay in the hospital overnight again, and some may get a repeat imaging study the next day prior to discharge to make sure the microspheres are in place.

As noted above, the radiation dissipates rapidly as the distance from the microspheres increases; however, patients may understandably be cautioned about contact with others for the first week or so after treatment, perhaps especially for more vulnerable populations (e.g., children, women who are pregnant).

Radioembolization is often given in one treatment, but in some cases (e.g., if there is cancer on both sides of the liver), radioembolization may be given in two treatments (with some period of time between the treatments, perhaps about 4 to 6 weeks).

Sorafenib

Descriptive Summary (pre-synthesis and synthesis efforts not applicable for descriptive summaries)

Sorafenib is taken at a typical dose of 400 mg orally twice a day on a continuous basis, with each dose often being given an hour or two after eating (unless one determines there is no benefit or the benefit:toxicity ratio is undesirable).

Best supportive care

Descriptive Summary (pre-synthesis and synthesis efforts not applicable for descriptive summaries)

Best supportive care is a broad concept constituting any therapy that is not specific to the disease *per se* but that can help support the patient in terms of palliation or management of symptoms or consequences of the disease. For instance, it can involve procedural and/or pharmacological treatment of pain, pharmacotherapy and/or psychotherapy for anxiety or depression, and the like, but it could also potentially involve something like a blood transfusion if the patient's disease was causing severe anemia; during best supportive care, treatment of any other comorbidities would also progress in the normal fashion (e.g., still actively managing a person's heart failure, diabetes, etc.). Best supportive care should not be confused with hospice; likewise, it is critical to note that best supportive care can – and should – be occurring in tandem with any disease-specific therapy that might be pursued. For instance, if a patient wished to pursue sorafenib, s/he should also receive best supportive care in tandem. For this reason, we often refer to the option of best supportive care as operationalized within this evidence report as best supportive care *alone*, with *alone* signifying that no disease-specific management is being pursued. Crucially, the word *alone* should not be misconstrued as receiving a lower quality of care: Patients should always be empowered by their health care team to know that pursuing best supportive care alone is a perfectly valid choice if their values and preferences best align with best supportive care alone; it does not constitute a "lesser" path, just a different one.

Descriptive Summary Results

Radioembolization

Radioembolization tries to slow or kill the cancer. You will have some tests to make sure you can have this treatment. One of the tests is called an "angiogram". In this test, a small cut is made in your wrist or groin area. A small tube called a "catheter" is put through the cut into a blood vessel. The tube is moved through your blood vessels to your liver. X-rays help make sure the tube goes the right way. Once the tube is where it needs to be, the clinician will inject a dye to help make your blood vessels easier to see. This can make you feel warm and like you need to urinate. The test takes about an hour or so to do. The area where the tube is inserted will be numbed, and many people are given medicine to make them sleepy as well. You may have to stay in the hospital for one night after having this done.

About a week or two later, you will have the radioembolization treatment. This means having another angiogram. But this time, when the tube is in the right spot, the clinician will put small beads in the blood vessel that goes to the liver. The beads give off radiation to try to slow or kill the cancer. The level of radiation goes down a lot the farther away something is from the bead. This helps protect healthy tissue, but the healthy tissue may still get some radiation. You may need to stay in the hospital for one night again.

Depending on how much cancer is in your liver, you may only need one radioembolization treatment. But some people need two radioembolization treatments. If you need two treatments, they are often given about a month or so apart.

You might be told you should be careful around children or pregnant women for a week or so after having radioembolization. But this may not be needed. You can talk with your health care team more about this if you are thinking about this option.

You will also get the treatments in the "Best supportive care" option.

Sorafenib

Sorafenib is a medicine that tries to slow or kill the cancer. You will take a pill twice a day, usually about an hour or two after eating. You will also get the treatments in the "Best supportive care" option.

Best supportive care

Best supportive care does not try to slow or kill the cancer. Instead, it focuses on making sure your symptoms are treated as much as possible. This can include things like treating pain, anxiety, depression, and feeling tired. Treatment may use medication, therapy, or procedures, like having a blood transfusion. Best supportive care varies for each person, because each person has different needs. A person can have best supportive care alone or in addition to treatment that tries to slow or kill the cancer, and there is no "right" or "wrong" choice. Either choice is okay, and the best choice is the choice that you think best matches what you want.

(Cancer Research UK 2019, Cedars-Sinai ND, Cleveland Clinic 2019, DynaMed 2020, Mayo Clinic 2019, Memorial Sloan Kettering Cancer Center 2019)

FAQ 2: How long will I live?

As noted in Scoping, we considered Finn 2020 on special request despite it falling outside the formally-defined scope of this evidence report. We acknowledge that Finn 2020 may change the landscape of systemic therapy in the intended population of this evidence report (see Critical Appraisal for details). Accordingly, although we do not attempt to make any indirect comparisons based on Finn 2020 (in line with Scoping), we do broadly incorporate the findings from Finn 2020 as part of the bolded synthesis statement for this FAQ; specifically, we include a qualitative statement that there may be an option other than sorafenib that might work somewhat better than sorafenib and that patients can talk about this with their clinician if they are interested. In line with our appraisal of Finn 2020, we consider this statement to have Moderate certainty for this outcome (downgrade due to inability to assess consistency of findings; see Critical Appraisal for details).

Sorafenib vs. placebo

Evidence Review

Overall survival, comparing sorafenib vs. placebo in ITT population

Study	overall survival (months median, 95% CI)		HR (95% CI)
	sorafenib group	placebo group	
Llovet 2008	10.7 (9.4 to 13.3)	7.9 (6.8 to 9.1)	0.69 (0.55 to 0.87)
Cheng 2009	6.5 (5.6 to 7.6)	4.2 (3.8 to 5.5)	0.68 (0.50 to 0.93)

Local therapy vs. sorafenib

Evidence Review

Overall survival, comparing sorafenib vs. placebo in ITT population

Study	median overall survival (95% CI), months		HR (95% CI)
	local therapy group*	sorafenib group	
Chow 2018	8.8 (7.5 to 10.8)	10.0 (8.6 to 13.8)	1.12 (0.9 to 1.4)
Vilgrain 2017	8.0 (6.7 to 9.9)	9.9 (8.7 to 11.4)	1.15 (0.94 to 1.41)
Yoon 2018	12.7 (6.7 to 18.7)†	9.9 (6.0 to 13.8)†	0.61 (0.38 to 0.98)

*Local therapy was radioembolization in Chow 2018 and Vilgrain 2017 and transarterial chemoembolization and external beam radiotherapy in Yoon 2018

†Yoon 2018 presented results in weeks rather than months, so we generated estimates in months via first estimating an average of (30*4+28+31*7)/12 ≈ 30.4 days/month, and then estimating Yoon 2018's weekly estimate in terms of months via: (estimate in weeks*7 days/week)/(30.4 days/month). The estimates in weeks are 55.0 (28.9 to 81.2) and 43.0 (25.9 to 60.1) for the local therapy and sorafenib groups, respectively.

Evidence Synthesis

Evidence Synthesis Process

Both Cheng 2009 and Llovet 2008 provide consistent evidence that sorafenib improves overall survival by approximately 3 months. We rate this finding as having Moderate certainty due to imprecision.

Both Chow 2018 and Vilgrain 2017 provide consistent evidence of no clear difference between radioembolization and sorafenib for overall survival. We rate this finding as having Moderate certainty due to imprecision.

Yoon 2018 finds that transarterial chemoembolization plus external beam radiotherapy (EBRT) may improve overall survival compared with sorafenib. However, we consider this finding to have only Very low certainty: We are concerned that this is a single-center, small (N = 90) trial; that all patients were required to have evidence of macroscopic vascular invasion (this constitutes a narrower inclusion than our inclusion/exclusion criteria in Scoping, with a group that may be sicker on average); and that Yoon 2018's findings have – as of yet – not been replicated. Accordingly, we have serious concern for indirectness (due to possible difficulty generalizing these findings to the intended population given the inclusion criterion of macroscopic vascular invasion and the single-center nature of the trial), and serious concerns for inability to assess consistency (due to the findings being limited to a single, small trial). We also note one might reasonably raise additional concern for imprecision based on the small nature of the trial and the reported median survivals in each group: the point estimates are separated by about 2.8 months, and the associated HR has a 95% CI that is compatible with smaller differences. One can also indirectly approximate a concern for imprecision based on the 95% CIs for the median survival of each group; of note, however, such an assessment is inherently limited, because the more appropriate assessment would come from directly-reported differences in median survivals (and the CI for the difference), but such metrics are not reported by the authors. Nevertheless, despite our concerns about Yoon 2018's data, we do acknowledge – albeit out of scope for this evidence report – there may be evidence that transarterial chemoembolization is more efficacious than radioembolization (e.g., as presented in Katsanos 2017), and we also acknowledge that the addition of EBRT could also yield additional efficacy beyond a transarterial approach alone (which is also explored in Katsanos 2017). These considerations could very well explain why Yoon 2018's results are different from the results of Chow 2018 and Vilgrain 2017.

Considering the above and in consideration of the fact that the overwhelming majority of evidence in this report considers radioembolization as the local therapy (vs. sorafenib), **we synthesize Yoon 2018 separately from the other evidence for this outcome**. We do this in a manner comparable to how we handle Finn 2020, namely with a **qualitative statement that there may be other approaches that help more than radioembolization** (or, in the case of Finn 2020, than sorafenib). As above, we consider this to have only **Very low certainty, with downgrades due to concerns regarding indirectness, imprecision, and inability to assess consistency of findings**.

Accordingly, if considering the median survival estimates for sorafenib in the above evidence (again, excluding Yoon 2018), this gives a range of 6.5 to 10.7 months (based on the estimates of 6.5 months [Cheng 2009], 9.9 months [Vilgrain 2017], 10 months [Chow 2018], and 10.7 months [Llovet 2008]), with a mean of approximately 9.3 months and a median of approximately 10 months. Doing the same for placebo gives a range of 4.2 to 7.9 months (based on these two values from Cheng 2009 and Llovet 2008, respectively) and a mean and median of 6.1 months. For radioembolization, the range is 8 to 8.8 months (based on these two values from Vilgrain 2017 and Chow 2018, respectively) with a mean and median of 8.4 months; importantly,

however, we note Chow 2018 and Vilgrain 2017 did not show any clear difference between radioembolization and sorafenib. If one then considered these radioembolization estimates in conjunction with the sorafenib estimates, one would have a range of 6.5 to 10.7 months with either sorafenib or radioembolization, with a mean of approximately 9 months and a median of approximately 9.4 months. We consider an **estimate of a median of 6 months with best supportive care** and an **estimate of a median of 9 months with either sorafenib or radioembolization** to be representative of the totality of these data (with **Moderate certainty due to imprecision**, as noted above), including reasonably maintaining the relative improvement in overall survival one can see in the sorafenib vs. placebo data: The sorafenib arm in Llovet 2008 had an approximate 35% relative increase in median survival time compared with the placebo arm ([10.7-7.9]/7.9), and for Cheng 2009, the relative increase was about 55% ([6.5-4.2]/4.2). Estimates of a median of 6 months and 9 months would constitute a relative 50% increase in median survival time; to the extent one desired more precise approximation via the relative increases just described, this could be estimated by applying those relative increases in conjunction with either a best estimate for sorafenib and radioembolization or a best estimate for the placebo group. Below, we demonstrate this estimation for the sake of transparency and clarity:

<u>If starting with an estimate for the placebo group (median of 6.1 months, as above) and estimating median overall survival for sorafenib or radioembolization based on a relative increase of approximately 35% to 55%:</u>

6.1 + 6.1*((10.7-7.9)/7.9) ≈ 6.1 + 6.1*0.35 ≈ 8.26 (or 8.24 if rounding to 0.35 in the calculation)
6.1 + 6.1*((6.5-4.2)/4.2) ≈ 6.1 + 6.1*0.55 ≈ 9.44 (or 9.46 if rounding to 0.55 in the calculation)

<u>If starting with an estimate for the sorafenib or radioembolization group (median of 9.4 months, as above) and estimating median overall survival for placebo based on a relative increase of approximately 35% to 55%:</u>

9.4/(10.7/7.9) ≈ 9.4/1.35 ≈ 6.94 (or 6.96 if rounding to 1.35 in the calculation)
9.4/(6.5/4.2) ≈ 9.4/1.55 ≈ 6.07 (or 6.06 if rounding to 1.35 in the calculation)

In these scenarios, we see that one might reasonably argue for presentation of approximately 8 to 9 months for sorafenib or radioembolization and approximately 6 months with placebo *or* approximately 9 months for sorafenib or radioembolization and approximately 6 to 7 months for placebo. The subtle difference in these presentations versus presenting 6 months and 9 months as noted above is the implied inference that the overall survival benefit may be about 3 months on average vs. about 2 to 3 months on average. Given the overall data and the how we present the bolded synthesis statement (namely, explicitly including statements that some people may live for shorter or longer periods of time), we opt for simplicity here in presenting single numbers rather than a range.

Evidence Synthesis Results

Radioembolization vs. sorafenib

On average, people who have treatment with radioembolization or sorafenib may live for about 9 months. Some people will live for a shorter time. Some people will live for a longer time. (Moderate certainty; based on 4 trials [Cheng 2009, Chow 2018, Llovet 2008, Vilgrain 2017])

There may be a treatment that is kind of like radioembolization that still works somewhat better than radioembolization. This treatment uses something called "chemoembolization", and it also uses radiation outside the body. (Very low certainty; based on 1 trial [Yoon 2018]) **There may also be a treatment that is kind of like sorafenib that still works somewhat better than sorafenib. This treatment uses atezolizumab and bevacizumab and works differently than sorafenib.** (Moderate certainty; based on 1 trial [Finn 2020]) **You can talk about this with your clinician if you are interested.**

Best supportive care

On average, people who have treatment with best supportive care alone may live for about 6 months. Some people will live for a shorter time. Some people will live for a longer time. (Moderate certainty; based on 2 trials [Cheng 2009, Llovet 2008])

FAQ 3: How long will my cancer stay controlled?

As noted in Scoping, we considered Finn 2020 on special request despite it falling outside the formally-defined scope of this evidence report. We acknowledge that Finn 2020 may change the landscape of systemic therapy in the intended population of this evidence report (see Critical Appraisal for details). Accordingly, although we do not attempt to make any indirect comparisons based on Finn 2020 (in line with Scoping), we do broadly incorporate the findings from Finn 2020 as part of the bolded synthesis statement for this FAQ; specifically, we include a qualitative statement that there may be an option other than sorafenib that might work somewhat better than sorafenib and that patients can talk about this with their clinician if they are interested. In line with our appraisal of Finn 2020, we consider this statement to have Very low certainty for this outcome (downgrades due to indirectness, imprecision, and inability to assess consistency of findings; see Critical Appraisal for details).

Sorafenib vs. placebo

Evidence Review

Progression-free survival*, comparing sorafenib vs. placebo in ITT population

Study	median progression-free survival (95% CI), months		HR (95% CI)
	sorafenib group	placebo group	
Cheng 2009	2.8 (2.63 to 3.58)	1.4 (1.35 to 1.55)	0.57 (0.42 to 0.79)
Llovet 2008	5.5 (4.1 to 6.9)	2.8 (2.7 to 3.9)	0.58 (0.45 to 0.74)

*Time to radiologic progression as defined by RECIST criteria; we simplify this to "Progression-free survival" in preparation for synthesis and to align with terminology used in the trials comparing local therapy vs. sorafenib

Local therapy vs. sorafenib

Evidence Review

Progression-free survival, comparing local therapy vs. sorafenib in ITT population

Study	median progression-free survival (95% CI), months		HR (95% CI)
	local therapy group*	sorafenib group	
Chow 2018	5.8 (3.7 to 6.3)	5.1 (3.9 to 5.6)	0.89 (0.7 to 1.1)
Vilgrain 2017	4.1 (3.8 to 4.6)	3.7 (3.3 to 5.4)	1.03 (0.85 to 1.25)
Yoon 2018	6.9 (5.3 to 8.5)†	2.6 (1.0 to 4.2)†	0.27 (0.17 to 0.44)

*Local therapy was radioembolization in Chow 2018 and Vilgrain 2017 and transarterial chemoembolization and external beam radiotherapy in Yoon 2018

†Yoon 2018 presented results in weeks rather than months, so we generated estimates in months via first estimating an average of (30*4+28+31*7)/12 ≈ 30.4 days/month, and then estimating Yoon 2018's weekly estimate in terms of months via: (estimate in weeks*7 days/week)/(30.4 days/month). The estimates in weeks are 30.0 (22.9 to 37.1) and 11.3 (4.4 to 18.2) for the local therapy and sorafenib groups, respectively.

Evidence Synthesis

Evidence Synthesis Process

Both Cheng 2009 and Llovet 2008 provide consistent evidence that sorafenib delays progression of the cancer by about 2 to 3 months. We rate this finding as having Low certainty due to indirectness (progression-free survival is a surrogate outcome) and imprecision.

Both Chow 2018 and Vilgrain 2017 provide consistent evidence of no clear difference between radioembolization and sorafenib for delaying the progression of the cancer. We rate this finding as having Low certainty due to indirectness (progression-free survival is a surrogate outcome) and imprecision.

Yoon 2018 finds that transarterial chemoembolization plus external beam radiotherapy (EBRT) may delay progression of the cancer compared with sorafenib. In addition to concerns about the use of progression-free survival (PFS) as an outcome *per se* (as noted above and in Scoping, this is a surrogate outcome), we are additionally concerned that this is a single-center, small (N = 90) trial; that all patients were required to have evidence of macroscopic vascular invasion (this constitutes a narrower inclusion than our inclusion/exclusion criteria in Scoping, with a group that may be sicker on average); and that Yoon 2018's findings have – as of yet – not been replicated. We thus have very serious concern for indirectness (due to the surrogate nature of PFS combined with possible difficulty generalizing these findings to the intended population given the inclusion criterion of macroscopic vascular invasion and the single-center nature of the trial), and serious concerns for inability to assess consistency (due to the findings being limited to a single, small trial). We thus have only Very low certainty in the findings from Yoon 2018. We also note one might reasonably raise additional concern for imprecision based on the small nature of the trial and the reported median survivals in each group. The point estimates are separated by about 4.3 months, but unlike the situation with overall survival (as addressed in FAQ 2: Will I live longer?), the 95% CI for the HR for PFS leaves one uncertain about the degree to which the results might be imprecise (to the extent one considers a 4.3-month difference in PFS to be meaningful, this would be an important consideration). We note the confidence intervals for median survival for each group suggest the data are compatible with a larger or smaller difference, including a smaller difference that might be of questionable benefit. Of note, however, such an assessment is inherently limited, because the more appropriate assessment would come from directly-reported differences in median survivals (and the CI for the difference), but such metrics are not reported by the authors. Although there exists a proposed simple method to approximate median survivals from the HR (and this is even incorporated into some statistical software), we note the method ultimately treats the HR as a RR. This is apparent in the proposed estimation of the hazard rate via *ln(2)/median survival time*, which is then incorporated into the estimation of the HR via *hazard rate 1*/*hazard rate 2*; accordingly, *ln(2)* is functionally removed from the estimation, which means the estimate of the HR becomes a function only of the median survival times in a simple ratio, akin to a RR. One can see the potential pitfall in this approach – at least when considered for Yoon 2018 – by returning to Yoon 2018's data: The median PFS times are 6.9 vs. 2.6 months, and the corresponding HR is 0.27, but 2.6 months/6.9 months ≈ 0.38 (or, alternatively, 6.9 months*0.27 ≈ 1.9 months). We thus do not employ such approximation here. Nevertheless, despite our concerns about Yoon 2018's data, we do acknowledge – albeit out of scope for this evidence report – there may be evidence that transarterial chemoembolization is more efficacious than radioembolization (e.g., as presented in Katsanos 2017), and we also acknowledge that the addition of EBRT could also yield additional efficacy beyond a transarterial approach alone (which is also explored in Katsanos 2017). These considerations could very well explain why Yoon 2018's results are different from the results of Chow 2018 and Vilgrain 2017.

Considering the above and in consideration of the fact that the overwhelming majority of evidence in this report considers radioembolization as the local therapy (vs. sorafenib), **we synthesize Yoon 2018 separately from the other evidence for this outcome**. We do this in a manner comparable to how we handle Finn 2020, namely with a **qualitative statement that there may be other approaches that help more than radioembolization** (or, in the case of Finn 2020, than sorafenib). As above, we consider this to have only **Very low certainty, with downgrades due to concerns regarding indirectness, imprecision, and inability to assess consistency of findings**.

Accordingly, if considering the median survival estimates for sorafenib in the above evidence (again, excluding Yoon 2018), this gives a range of 2.8 to 5.5 months (based on the estimates of 2.8 [Cheng 2009], 3.7 [Vilgrain 2017], 5.1 [Chow 2018], and 5.5 [Llovet 2008]), with a mean of approximately 4.3 months and a median of approximately 4.4 months. Doing the same for placebo gives a range of 1.4 to 2.8 months (based on these two values from Cheng 2009 and Llovet 2008, respectively) and a mean and median of 2.1 months. For radioembolization, the range is 4.1 to 5.8 months (based on these two values from Vilgrain 2017 and Chow 2018, respectively) with a mean and median of approximately 5 months; importantly, however, we note Chow 2018 and Vilgrain 2017 did not show any clear difference between radioembolization and sorafenib. If one then considered these radioembolization estimates in conjunction with the sorafenib estimates, one would have a range of 2.8 to 5.8 months with either sorafenib or radioembolization, with a mean of 4.5 months and a median of 4.6 months. We consider an **estimate of a median of 2.1 months with best supportive care** and an **estimate of a median of 4.6 months with either sorafenib or radioembolization** to be very representative of the totality of these data, including reasonably maintaining the relative improvement in PFS one can see in the sorafenib vs. placebo data (the sorafenib arms had an approximate doubling of the PFS time seen in the placebo arm). As above, we consider these statements to be of **Low certainty, with downgrades for indirectness and imprecision**.

Evidence Synthesis Results

Radioembolization vs. sorafenib

On average, people who have treatment with radioembolization or sorafenib may have their cancer controlled for about 4 to 5 months. Some people's cancer will be controlled for a shorter time. Some people's cancer will be controlled for a longer time. (Low certainty; based on 4 trials [Cheng 2009, Chow 2018, Llovet 2008, Vilgrain 2017]) **Your cancer being controlled does not mean you will not have symptoms, but best supportive care treatments may be able to help with symptoms you may have.**

There may be a treatment that is kind of like radioembolization that still works somewhat better than radioembolization. This treatment uses something called "chemoembolization", and it also uses radiation outside the body. (Very low certainty; based on 1 trial [Yoon 2018]) **There may also be a treatment that is kind of like sorafenib that still works somewhat better than sorafenib. This treatment uses atezolizumab and bevacizumab and works differently than sorafenib.** (Very low certainty; based on 1 trial [Finn 2020]) **You can talk about this with your clinician if you are interested.**

Best supportive care

On average, people who have treatment with best supportive care alone may have their cancer controlled for about 2 months. Some people's cancer will be controlled for a shorter time. Some people's cancer will be controlled for a longer time. (Low certainty; based on 2 trials [Cheng 2009, Llovet 2008]) **Your cancer being controlled does not mean you will not have symptoms, but best supportive care treatments may be able to help with symptoms you may have.**

FAQ 4: Will the treatment change my quality of life?

Sorafenib vs. placebo

Evidence Review

Cheng 2009 reported only that there was no between-group difference on a quality-of-life measure (Functional Assessment of Cancer Therapy–Hepatobiliary [FACT–HP] questionnaire) but provided no further detail. Llovet 2008 did not report on a quality of life outcome.

Local therapy vs. sorafenib

Evidence Review

Chow 2018 reported no significant differences in the EuroQol-5 Dimension (EQ-5D) questionnaire comparing selective internal radiation therapy vs. sorafenib groups throughout the study (EQ-5D collected at each visit) in either an ITT or as-treated populations. The results at 9 months for the ITT population are presented in the table below. Selection of this time point was informed by median overall survival estimates and the 9-month timepoint reporting having 95% CIs (next closest timepoint reporting 95% CIs was at 12 months).

Quality of life as assessed by the EuroQol-5 Dimension in Chow 2018*

Population and timing of assessment	Selective internal radiation therapy (mean [95% CI])	Sorafenib (mean [95% CI])
Baseline	83 (79 to 87)	85 (82 to 88)
9 months	72 (61 to 83)	69 (58 to 80)

*As estimated from Fig A5 in data supplement for Chow 2018. The EuroQol-5 Dimension (EQ-5D) assesses 5 dimensions of health status: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Depending on the subtype of EQ-5D a person uses, these may be assessed slightly differently (e.g., on a 3-level scale vs. a 5-level scale). Higher scores are better. The authors present the overall EQ-5D scores on a scale from 0 to 1. We have multiplied the data by 100 for easier comprehension.

Vilgrain 2017 used the European Organisation for Research and Treatment of Cancer (EORTC) quality of life questionnaire (QLQ-C30; version 3) and the specific hepatocellular carcinoma module (QLQ-HCC18) to assess quality of life, but it only reported the "global health status subscore", finding it to be statistically-significantly better in the selective internal radiation therapy group than in the sorafenib group (group effect $p = 0.0048$; time effect $p < 0.0001$) and that the between-group difference increased with time (group-time interaction $p = 0.0447$). Although Vilgrain 2017 noted the remainder of the quality of life findings would be "reported elsewhere", we note there is no signal of follow-up publications in the Abdel-Rahman 2020 Cochrane review. Furthermore, there is evidence in Vilgrain 2017 of substantial loss to follow-up over time (see Critical Appraisal for details). Coupling our serious concern for loss to follow-up with our very serious concern for selective outcome reporting and serious concern for imprecision (see Critical Appraisal for details), we did not carry Vilgrain 2017 forward to Evidence Synthesis for this outcome. Alternatively, one could note the Vilgrain 2017 data have Very low certainty (with extremely serious concern for risk of bias and serious concern for imprecision). Given the data from Vilgrain 2017

highlight a statistically-significant finding, this might be viewed as roughly 'inconsistent' with the findings of no clear difference in Chow 2018 (though we note full assessment of consistency would involve more than mere assessment of statistical significance or lack thereof); in any case, in the setting of inconsistent findings that also differ in certainty, GRADE calls for concluding on the higher-certainty evidence, which would be Chow 2018.

Evidence Synthesis

Evidence Synthesis Process

The data from Chow 2018 show no clear difference between SIRT and sorafenib for quality of life as assessed by the EuroQoL-5 Dimension (EQ-5D). However, we note these data are somewhat limited in quantity and come from a single trial. It also remains possible that a more disease-specific quality of life scale might capture some meaningful difference, thus raising concern for indirectness with the use of EQ-5D. Accordingly, we have concerns regarding imprecision, an **inability to assess consistency of findings, and indirectness, giving us only Low certainty of no difference in quality of life between SIRT and sorafenib**. In Scoping, we also specified that, if permitted by the available evidence, we might try to give an indication of how quality of life changes from baseline if the available evidence could not distinguish between options. Although we did not carry out formal statistical analysis on the change within groups over time in Chow 2018, we consider it reasonable to infer that their data do suggest a **potential for quality of life to decrease over time irrespective of treatment group; importantly, however, the degree to which this potential decrease would be meaningful or perceptible to the patient is unclear**. We thus rate this inference as having **Low certainty, with the same downgrades as we gave for the between-groups comparison for Chow 2018**.

Based on the above observation and the extremely limited reporting from Cheng 2009 suggesting no difference in quality of life between sorafenib vs. placebo, one could also potentially infer that **quality of life may decrease in the placebo group over time**. However, this inference would bring **additional indirectness, thus yielding only a Very low certainty inference**.

Combining these considerations leads us to our bolded synthesis statement below, and we explicitly incorporate uncertainty into the synthesis statement via the final sentence.

Evidence Synthesis Results

Radioembolization vs. sorafenib vs. placebo

Some people may feel like their quality of life goes down some over time. But other people may not feel this way. Research does not show any difference in quality of life based on whether someone gets radioembolization, sorafenib, or best supportive care alone. But there is not much research in this area.

(Low to Very low certainty; based on 2 trials [Cheng 2009, Chow 2018])

FAQ 5: What are the risks or side effects?

Treatment-related vs. treatment-emergent

In general, adverse events can be either 'treatment-emergent' or 'treatment-related'. Being 'treatment-emergent' signifies the adverse event *emerged* after treatment began, but there is no effort to ascertain whether there is a reasonable link between the adverse event and the study therapy (other than occurring in someone who received that study therapy). For this reason, 'treatment-emergent' adverse events are sometimes called 'all-cause' adverse events or adverse events 'from any cause'. In contrast, 'treatment-related' signifies the adverse event both emerged after treatment began *and* was judged to be a direct result of taking the study therapy (ideally, such judgment should be made by people blinded to what therapy the patients are receiving).

In Scoping, we specified our primary outcomes for this FAQ as serious adverse events and treatment withdrawal/discontinuation due to adverse events. Our intention here was to extract information on treatment-related serious adverse events and treatment withdrawal/discontinuation due to treatment-related adverse events, only using the more generic treatment-emergent versions if the former were lacking. However, we did not make this explicit in Scoping. Even as such, we feel it is very reasonable and even desirable to focus attention on the treatment-related version of these outcomes rather than the treatment-emergent version of these outcomes whenever possible. This is reflected throughout Evidence Synthesis for this FAQ. If desired, further detail is available in Critical Appraisal for some of the trials for which we only present treatment-related adverse events here.

Specific adverse events

Per Scoping, we included specific adverse events to help contextualize the primary outcomes (as above, treatment-related serious adverse events and treatment withdrawal due to treatment-related adverse events). However, we did not assess certainty for specific adverse events given their role in this evidence report. In line with the section above, we also focus on adverse events specified to be treatment-related whenever possible.

Comparative statistics

In many cases, no comparative statistics were reported beyond the authors (1) sometimes reporting the rates were 'similar' (or using comparable verbiage) and/or (2) sometimes providing p values for the comparison between groups. Unless specified otherwise, comparative statistics other than p values (e.g., relative risks, confidence intervals) were calculated by EBSCO Information Services. We use method 10 of Newcombe 1998 (also recommended by Altman 2000) for calculating the confidence interval for the absolute risk difference where applicable given 'traditional' methods may underperform with certain data.

Providing quantitative context for adverse events vs. placebo

In some cases, one can contextualize adverse event outcomes in the placebo group as adverse events that occur even though no specific treatment is taken. Consideration of placebo group rates is important to provide a proper quantitative framework for interpreting the results in the intervention group (i.e., to permit insight into how much of the risk in the intervention group is conceivably attributable to the intervention). For instance, if the risk of outcome X is 12% in the intervention group and 5% in the placebo group, it could be misleading to present the 12% in isolation (whereas providing it in conjunction with the 5% risk helps people appreciate the magnitude of risk attributable to the intervention, which in the example is 7%). However, direct presentation of placebo group risks is less meaningful in the context of this evidence report for reasons we specify in greater detail below; nevertheless, we still take account of the placebo group risk in our synthesis to avoid misleading inferences.

Discontinuation due to treatment-related adverse events and serious treatment-related adverse events

Sorafenib vs. placebo

Evidence Review

Cheng 2009 reported treatment-related serious adverse events and treatment discontinuations due to adverse events (without specification of whether the adverse events leading to discontinuation were judged to be treatment-related).

Llovet 2008 reported discontinuations both as a treatment-emergent version and as a treatment-related version.

Results from Llovet 2008 and Cheng 2009 are included in the tables below.

Proportion of patients having adverse events in the safety population* (Cheng 2009)

Outcome	Sorafenib (n = 149), % (n)	Placebo (n = 75), % (n)	Comparative statistics (95% CI)
treatment discontinuation due to adverse events	19.5% (29)	13.3% (10)	ARD 6.1% (-4.9% to 15.4%)
≥1 drug-related serious adverse event†	8.7% (13)	1.3% (1)	ARD 7.4% (0.6% to 13.1%)

*Safety analysis included all patients who received ≥ 1 dose of study drug; no comparative statistics were reported

†Defined as those that were life-threatening, resulted in death, required patient hospitalization or prolongation of hospitalization, or resulted in a persistent or significant disability or incapacity

Proportion of patients having adverse events in the safety population* (Llovet 2008)

outcome	Sorafenib (n = 297), % (n)	Placebo (n = 302), % (n)	Comparative statistics (95% CI)
discontinuations due to adverse events	29.0% (86)	29.8% (90)	ARD -0.8% (-8.1% to 6.4%)
discontinuations due to adverse events considered to be drug-related	11.4% (34)	5.0% (15)	ARD 6.5% (2.1% to 11%)

*Safety population refers to patients who received ≥ 1 dose of a study drug

Local therapy vs. sorafenib

Evidence Review

Three trials compared local therapy vs. sorafenib, two of which (Vilgrain 2017, Chow 2018) used radioembolization (RE)/selective internal radiotherapy (SIRT) for local therapy and one of which (Yoon 2018) used transarterial chemoembolization plus external beam radiotherapy (TACE + EBRT). Vilgrain 2017 and Chow 2018 were multi-center trials with 467 and 360 patients respectively, whereas Yoon 2018 was conducted at a single center in Korea with only 90 patients. Yoon 2018 also appears to give focus to treatment-emergent adverse events rather than treatment-related adverse events (see Critical Appraisal for further detail). The much smaller sample size, less desirable approach to adverse event handling, and single-center nature of the study render the data from Yoon 2018 of questionable utility and certainty compared to the data available via Vilgrain 2017 and Chow 2018. Considering this, we did not carry Yoon 2018's adverse event data forward to Evidence Synthesis.

Chow 2018 and Vilgrain 2017 both report on treatment-related serious adverse events, but neither reported on discontinuations due to treatment-related adverse events in a comparative manner. We note this may be a natural consequence of the comparison being investigated by these trials; specifically, discontinuation due to treatment-related adverse events may be less of a consideration for the option of local therapy. This may be largely due to two factors: (1) local therapy may occur in a single treatment session or possibly one or two treatment sessions, and (2) local therapy appears, in general, to be well-tolerated. To the extent these things are true, any measurement of discontinuation due to treatment-related adverse events could in effect be mostly or entirely capturing the proportion of patients who discontinue sorafenib due to treatment-related adverse events (given sorafenib is administered daily). One sees at least indirect evidence of this in Vilgrain 2017, which reported on treatment discontinuation, but only for the sorafenib group (reporting permanent discontinuation of sorafenib due to drug-related toxicity in 108/216 [50%]). We see additional engagement with this concept in Vilgrain 2017's discussion, where they note SIRT constitutes "one or two" treatments whereas sorafenib is a "daily treatment that should, ideally, not be discontinued".

Although Chow 2018 and Vilgrain 2017 both report on treatment-related serious adverse events, there are complexities surrounding the way in which these populations were defined such that Chow 2018 provides the clearest assurance of reliable estimates. In brief, Vilgrain 2017 only provides estimates for treatment-related serious adverse events in its safety population, but its safety population includes a sizeable number of people who never received their allocated treatment; specifically, the safety population for the SIRT group includes 42 people who never received SIRT.

The results for Chow 2018 and Vilgrain 2017 are included below.

Proportion of patients having treatment-related serious adverse events in the as-treated or safety* population

Outcome	Source	Selective internal radiation therapy, % (n/N)	Sorafenib, % (n/N)	Comparative statistics (95% CI)
≥1 treatment-related serious adverse event	Chow 2018†	4.6% (6/130)	9.3% (15/162)	RR 0.5 (0.2 to 1.25) ARD -4.6% (-10.6% to 1.6%)
	Vilgrain 2017‡	15.9% (36/226)§¶	25.9% (56/216)§	RR 0.61 (0.42 to 0.89) ARD -10% (-17.5% to -2.4%)
		14.9% (26/174)**	22.3% (46/206)**	RR 0.67 (0.43 to 1.04) ARD -7.4% (-15% to 0.6%)
		20.7% (36/174)††	27.2% (56/206)††	RR 0.76 (0.53 to 1.1) ARD -6.5% (-14.9% to 2.2%)
		14.9% (26/174)‡‡	27.2% (56/206)‡‡	RR 0.55 (0.36 to 0.84) ARD -12.2% (-20.1% to -4%)
		20.7% (36/174)§§	22.3% (46/206)§§	RR 0.93 (0.63 to 1.36) ARD -1.6% (-9.8% to 6.7%)

*Chow 2018 defined its as-treated population as patients who received the allocated intervention; Vilgrain 2017 uses a similar definition for its per-protocol population, but unfortunately, they present estimates for this outcome only for the safety population, which includes a sizeable portion of patients who did not receive the treatment to which they were allocated. See footnotes § and following (as well as Critical Appraisal) for additional detail if desired.

†Chow 2018 report the p value for this comparison as 0.17. Per their protocol (text that follows is largely verbatim, with very minimal revision for clarity as a footnote), a serious adverse event was defined as an adverse event that led to: (1) death; (2) a serious deterioration in health that (2a) resulted in a life-threatening illness or injury, (2b) resulted in a permanent impairment of a body structure or a body function, (2c) required in-patient hospitalization or prolongation of existing hospitalization, or (2d) resulted in medical or surgical intervention to prevent permanent impairment to body structure or a body function; or (3) led to fetal distress or death or a congenital abnormality or birth defect.

‡Defined as "Any adverse event or reaction that results in the death, or endangers the life, of an individual participating in a research study, requires hospitalisation or prolonged hospitalisation, causes a significant or long-term disability or incapacity, or results in a congenital anomaly or malformation".

§This approach sets the numerator as the number of events considered attributable to the intervention and the denominator as all the people in the safety population. However, the denominator is problematic – or at best uncertain – because it includes 42 people who did not receive SIRT. This larger denominator could downwardly-bias the proportion in the SIRT group. It is also unclear in this approach what number of the events in each group would be excluded from a comparable per-protocol analysis (10 people in each group were excluded for protocol violations not related to whether they received the allocated treatment, giving a final per-protocol population of 174 in the SIRT group and 206 in the sorafenib group). Thus, a range of possibilities exists between the number of events being 10 lower in each group, 0 lower in each group, or some mixture thereof.

¶The authors also report this as 20% (45/226), but they too note only 36 of these events were related to SIRT, with 9 being related to sorafenib.

**Based on the situation delineated in footnote §, this approach assumes 10 events would be lost from the numerator in both groups in the per-protocol analysis.

††Based on the situation delineated in footnote §, this approach assumes no events would be lost from the numerator in both groups in the per-protocol analysis.

‡‡Based on the situation delineated in footnote §, this approach assumes 10 events would be lost from the numerator in the SIRT group and no events would be lost in the sorafenib group in the per-protocol analysis.

§§Based on the situation delineated in footnote §, this approach assumes no events would be lost from the numerator in the SIRT group and 10 events would be lost in the sorafenib group in the per-protocol analysis.

Evidence Synthesis

Evidence Synthesis Process

Outcome	Approach taken and justification
discontinuations due to treatment-related adverse events	***sorafenib vs. placebo*** Both trials comparing sorafenib vs. placebo (Llovet 2008, Cheng 2009) report on discontinuations due to adverse events, and both find an increase with sorafenib (although only reaching conventional statistical significance in the larger trial, Llovet 2008). Of note, however, only Llovet 2008 clearly provides estimates for discontinuations due to treatment-related adverse events; Cheng 2009 appears only to provide estimates for discontinuations due to treatment-emergent adverse events. Accordingly, we utilize Llovet 2008 for evidence synthesis for this outcome, giving: Risk in sorafenib group: 11.4% Risk in placebo group: 5.0% ARD 6.5% (2.1% to 11%) We consider this outcome to have **Moderate certainty with a downgrade for imprecision**. For the bolded synthesis statement, we incorporate this imprecision explicitly by using the risk in the placebo group as the baseline and adding the lower and upper bounds of the 95% CI for the ARD to express the risk in the sorafenib group, **giving an estimate of 5% for the placebo group and 7.1% to 16% for the sorafenib group. However, presentation of a placebo group risk for this outcome within the option of best supportive care would not make sense considering no placebo is given in the real world when someone opts to pursue best supportive care alone. Accordingly, for ease of understanding, we simply subtract the placebo group risk from the sorafenib group risk for the simplified synthesis statement, giving an estimate of 2.1% to 11% of people in the sorafenib group (note: this amounts to simply using the lower and upper bounds of the 95% CI for the ARD and is in effect the 'attributable risk' for this outcome for sorafenib).** ***selective internal radiotherapy vs. sorafenib*** Of note, discontinuation due to treatment-related adverse events was not reported in a comparative manner in trials comparing selective internal radiotherapy vs. sorafenib. However, Vilgrain 2017 does report discontinuation of sorafenib due to drug-related toxicity in 108/216 (50%). This is markedly inconsistent with the estimates offered by head-to-head comparison against placebo in Llovet 2008, and without a corresponding placebo-based estimate from Vilgrain 2017, we find it difficult to reconcile these estimates, including entertaining the possibility that the estimate from Vilgrain 2017 is biased by the open-label nature of the study (thus possibly leading to a greater tendency to attribute side effects to sorafenib). Other possibilities exist as well, which may be acting instead or in concert, such as the population in Vilgrain 2017 possibly being sicker (e.g., somewhat more patients with worse grade of cirrhosis and considerably more with vascular extension, the latter of which Vilgrain 2017 notes carries a worse prognosis than extrahepatic

	spread). We acknowledge that Llovet 2008 was conducted in 121 centers in 21 countries across Europe, North America, South America, and Australasia, whereas Vilgrain 2017 was conducted in 25 centers in 1 country (France), but we are uncertain whether this by itself could offer satisfactory explanation for the apparent inconsistency in the results. Given the above considerations, we are reluctant to conclude in a manner that incorporates the 50% from Vilgrain 2017 in a quantitative manner, especially without a corresponding placebo risk. As can be appreciated from the Llovet 2008 data, the attributable risk is importantly different from the absolute risk. If one were to consider the point estimate for the RR for the comparison of sorafenib vs. placebo in Llovet 2008, it would be 2.3; there is no way to know what a comparable estimate would be for Vilgrain 2017 given the nature of the trial, but even if using the point estimate for the RR from Llovet 2008 as a rough approximation, the 50% risk in the sorafenib group in Vilgrain 2017 would translate to an anticipated (hypothetical) placebo group risk of approximately 21.7% (50%/2.3). If this were true, this would have substantial implications for the risk we present for the placebo group, but we only present this here as a thought experiment to make it clear why incorporation of the 50% estimate from Vilgrain 2017 would be potentially problematic. **Accordingly, we incorporate Vilgrain 2017 into our synthesis statement by saying one study may have found a higher risk of stopping sorafenib due to side effects, but it is not clear how much of this risk was due to the medication**. This is a valid statement given we cannot be certain about the degree of attributable risk in Vilgrain 2017. Additionally, given the uncertainty this introduces for the risk reported for sorafenib (and thus also for placebo), we **downgrade our certainty from Moderate (as presented in the section on sorafenib vs. placebo) to Low, with the additional downgrade being due to inconsistency.** **For selective internal radiotherapy**, although there are no quantitative estimates to provide, we consider it reasonable to adopt some of the language from Vilgrain 2017's discussion, namely that **treatment can occur in one or two sessions and is generally tolerated well, so stopping treatment due to side effects may be less of a consideration or concern.**
≥1 treatment-related serious adverse event	***sorafenib vs. placebo*** Only Cheng 2009 reports on this outcome, so we use their data for inference, giving: Risk in sorafenib group: 8.7% Risk in placebo group: 1.3% ARD 7.4% (0.6% to 13.1%) We consider this outcome to have **Moderate certainty with a downgrade for imprecision**. For the bolded synthesis statement, we incorporate this imprecision explicitly by using the risk in the placebo group as the baseline and adding the lower and upper bounds of the 95% CI for the ARD to express the risk in the sorafenib group, **giving an estimate of 1.3% for the placebo group and 1.9% to 14.4% for the sorafenib group**. **However, presentation of a placebo group risk for this outcome within the option of best supportive care would not make sense considering no placebo is given in the real world when someone opts to pursue best supportive care alone. Accordingly, for ease of understanding, we simply subtract the placebo group risk from the sorafenib group risk for the simplified synthesis statement, giving an estimate of 0.6% to 13.1% of people in the sorafenib group (note: this amounts to simply using the lower and upper bounds of the 95% CI for the ARD and is in effect the 'attributable risk' for this outcome for sorafenib).**

	selective internal radiotherapy vs. sorafenib Of note, the estimated rate in the sorafenib group presented in the above section on sorafenib vs. placebo (whether considering the point estimate or the range formed by the 95% CI bounds) is consistent with the estimate of treatment-related serious adverse events seen in Chow 2018 (9.3%), which compared selective internal radiotherapy vs. sorafenib. Although we carried out several sensitivity analyses of the data from Vilgrain 2017 given the problems surrounding how Vilgrain 2017 measured and reported this outcome, we do not consider the data from Vilgrain 2017 to be reliable enough to formally incorporate into quantitative synthesis. Given the RR for the comparison in Chow 2018 is 0.5 (95% CI 0.2 to 1.25), the data have considerable imprecision. There are (at least) two options here: (1) simply conclude no clear difference between selective internal radiotherapy and sorafenib based on the wide CI in Chow 2018 (and set the risk with selective internal radiotherapy equal to the risk with sorafenib), or (2) explicitly incorporate the imprecision in the estimate of risk for the selective internal radiotherapy group. We favor the latter. Based on the estimated risk in the sorafenib group of 0.6% to 13.1% (as above) and the lower and upper bounds of the 95% CI for the RR from Chow 2018 (0.2 to 1.25), one finds an **estimate of risk for the selective internal radiotherapy group ranging from 0.12%** (0.2*0.6%) **to 16.4%** (1.25*13.1%). **To acknowledge Vilgrain 2017 in synthesis, we also note one study may have found a higher risk of serious adverse events for both radioembolization and sorafenib, but that it may be less reliable**. We consider this outcome to have **Moderate certainty with a downgrade for imprecision**.

Lastly, for the option of best supportive care, although it would not be sensible to present placebo-group risks here given the nature of what is being measured in that context (vs. what occurs when one truly pursues best supportive care alone), we provide a simple truth for best supportive care in the bolded synthesis statement below that does not require empirical support.

Evidence Synthesis Results

Radioembolization

In general, people often find this treatment to be tolerable. Part of this may be because it is often given in 1 or 2 sessions rather than something they do every day. Because of this, stopping treatment due to side effects may be less of a concern compared with sorafenib.

(Certainty not rated; descriptive response here based on adaptation of Vilgrain 2017)

Research gives different estimates about how often serious side effects occur. They may happen in less than 1 out of 100 people (<1%) or in as many as 16 out of 100 people (16%). One study suggests this number may be higher, but it is not clear how much of that number was really because of radioembolization.

(Moderate certainty; based on 3 trials [Cheng 2009, Chow 2018, Vilgrain 2017])

Sorafenib

Out of 100 people who take sorafenib, about 2 to 11 (2% to 11%) may stop taking it because of side effects. One study suggests this number may be higher, but it is not clear how much of that number was really because of sorafenib.

(Low certainty; based on 2 trials [Llovet 2008, Vilgrain 2017])

Research gives different estimates about how often serious side effects occur. They may happen in 1 out of 100 people (1%) or in as many as 13 out of 100 people (13%). One study suggests this number may be higher, but it is not clear how much of that number was really because of sorafenib.

(Moderate certainty; based on 3 trials [Cheng 2009, Chow 2018, Vilgrain 2017])

Best supportive care

Your cancer may spread more easily without treatment that tries to kill or slow it. Because of this, some people might initially want best supportive care alone but later decide they want specific treatment. If this happens, you can discuss your options with your clinician at that time.

(Certainty not rated**)**

Named adverse events (for contextualization of discontinuation due to adverse events and serious adverse events)

Sorafenib vs. placebo

Evidence Review

Both trials (Cheng 2009, Llovet 2008) explicitly reported that named adverse events were treatment-related. Data for these two trials are presented in the table below.

Drug-related specific adverse events of any grade in safety population*

Outcome	Source	Sorafenib, % (n/N)	Placebo, % (n/N)	P value†	Comparative statistics (95% CI)†
Constitutional symptoms					
Fatigue	Llovet 2008	7.4% (22/297)	5.3% (16/302)	0.07	RR 1.4 (0.75 to 2.61) ARD 2.1% (-1.9% to 6.2%)
	Cheng 2009	20.1% (30/149)	8.0% (6/75)	NR	RR 2.52 (1.1 to 5.78) ARD 12.1% (2% to 20.5%)
Weight loss	Llovet 2008	3.0% (9/297)	0.3% (1/302)	<0.001	RR 9.15 (1.17 to 71.79) ARD 2.7% (0.6% to 5.3%)
Dermatologic events					
Alopecia	Llovet 2008	4.7% (14/297)	0.7% (2/302)	<0.001	RR 7.12 (1.63 to 31.05) ARD 4.1% (1.5% to 7.1%)
	Cheng 2009	24.8% (37/149)	1.3% (1/75)	NR	RR 18.62 (2.61 to 133.12) ARD 23.5% (14.9% to 31.1%)
Dry skin	Llovet 2008	2.7% (8/297)	1.3% (4/302)	0.04	RR 2.03 (0.62 to 6.68) ARD 1.4% (-1.1% to 4%)
Hand–foot skin reaction	Llovet 2008	7.1% (21/297)	1.0% (3/302)	<0.001	RR 7.12 (2.15 to 23.61) ARD 6.1% (3% to 9.6%)
	Cheng 2009	45.0% (67/149)	2.7% (2/75)	NR	RR 16.86 (4.25 to 66.95) ARD 42.3% (32.1% to 50.5%)
Rash or desquamation	Llovet 2008	5.4% (16/297)	3.6% (11/302)	0.12	RR 1.48 (0.7 to 3.13) ARD 1.7% (-1.7% to 5.3%)
	Cheng 2009	20.1% (30/149)	6.7% (5/75)	NR	RR 3.02 (1.22 to 7.47) ARD 13.5% (3.7% to 21.6%)
Gastrointestinal events					
Anorexia	Llovet 2008	4.7% (14/297)	1.0% (3/302)	<0.001	RR 4.75 (1.38 to 16.34)

					ARD 3.7% (1.1% to 6.8%)
	Cheng 2009	12.8% (19/149)	2.7% (2/75)	NR	RR 4.78 (1.14 to 19.99) ARD 10.1% (2.2% to 16.7%)
Diarrhea	Llovet 2008	13.1% (39/297)	3.6% (11/302)	<0.001	RR 3.61 (1.88 to 6.9) ARD 9.5% (5.1% to 14.1%)
	Cheng 2009	25.5% (38/149)	5.3% (4/75)	NR	RR 4.78 (1.77 to 12.9) ARD 20.2% (10.3% to 28.4%)
Abdominal pain NOS	Llovet 2008	2.7% (8/297)	1.0% (3/302)	0.007	RR 2.71 (0.73 to 10.12) ARD 1.7% (-0.6% to 4.3%)
Voice changes	Llovet 2008	2.0% (6/297)	0.3% (1/302)	<0.001	RR 6.1 (0.74 to 50.37) ARD 1.7% (-0.2% to 4%)
Hypertension	Llovet 2008	1.7% (5/297)	0.7% (2/302)	0.05	RR 2.54 (0.5 to 13) ARD 1% (-0.9% to 3.3%)
	Cheng 2009	18.8% (28/149)	1.3% (1/75)	NR	RR 14.09 (1.96 to 101.6) ARD 17.5% (9.5% to 24.6%)

NOS, not otherwise specified; NR, not reported

*The safety population refers to patients who received ≥ 1 dose of a study drug. Authors provided data for drug-related specific adverse events that occurred in ≥ 5% (Llovet 2008) or ≥ 10% (Cheng 2009) of patients in either study group. We extracted the outcomes that showed a significant between-group difference in Llovet 2008, and we took a comparable approach to Cheng 2009; although no comparative statistics were provided for Cheng 2009, the magnitude of the difference suggested to us that the differences would likely meet conventional statistical significance other than for nausea, which we have not included in the above table (rates were quite comparable). We also included rash or desquamation for both trials (despite not reaching conventional statistical significance in Llovet 2008, there was signal of a possible difference in other trials), and we also include fatigue from Llovet 2008 despite not technically reaching conventional statistical significance. Adverse events were graded in severity in both trials, but we focus here on 'any' grade of adverse events. Critical Appraisal contains additional detail limited to 'grade 3 or 4' adverse events.

†All p values are reported by authors (rather than calculated by EBSCO Information Services). We note that some of the p values do not appear concordant with calculation of 95% CIs for the relative risk and absolute risk difference. We also note the authors report using Fisher's exact test for comparison of adverse events, but this appears inaccurate based on spot checks of estimates where the p values appeared discordant with the 95% CIs. Some of these errors seem obvious (e.g., Llovet 2008 reporting a p value of 0.007 for their outcome of abdominal pain not otherwise specified and <0.001 for voice changes despite the event rates being what they are). We place more emphasis on the 95% CIs than the p values, so we did not recalculate all the p values but still include them in their author-reported form for completeness and transparency.

Local therapy vs. sorafenib

Evidence Review

Of the 2 trials considered for specific adverse events (Chow 2018, Vilgrain 2017; we do not consider Yoon 2018 for reasons stated earlier in this FAQ), Vilgrain 2017 reports these were treatment-related Chow 2018 does not specify whether they were treatment-emergent or treatment-related, but we assumed the latter based on the proportions they report. Chow 2018 reports p values for between-group comparisons (but no other comparative

statistics), whereas Vilgrain 2017 did not report any comparative statistics for specific adverse events. Data for these two trials are presented in the table below.

Proportion of patients having specific adverse events of any grade* in the as-treated or per-protocol population†‡

System organ class, disorder	Source	Radioembolization	Sorafenib	Comparative statistics (95% CI)
Gastrointestinal disorders				
constipation	Chow 2018	0	5.6 % (9/162)	RR not estimated due to 0 cell ARD -5.6% (-10.2% to -1.7%)
diarrhea	Chow 2018	1.5% (2/130)	29.6% (48/162)	RR 0.05 (0.01 to 0.21) ARD -28.1% (-35.6% to -20.5%)
	Vilgrain 2017	9.8% (17/174)	78.2% (161/206)	RR 0.13 (0.08 to 0.2) ARD -68.4% (-74.6% to -60.3%)
General disorders				
fatigue	Chow 2018	3.8% (5/130)	15.4% (25/162)	RR 0.25 (0.1 to 0.63) ARD -11.6% (-18.3% to -4.8%)
	Vilgrain 2017	50% (87/174)	76.2% (157/206)	RR 0.66 (0.56 to 0.78) ARD -26.2% (-35.3% to -16.6%)
Skin and subcutaneous tissue disorders				
alopecia	Chow 2018	0	9.9% (16/162)	RR not estimated due to 0 cell ARD -9.9% (-15.4% to -5.2%)
	Vilgrain 2017	0	16.5% (34/206)	RR not estimated due to 0 cell ARD -16.5% (-22.2% to -11.6%)
palmo-plantar erythro-dysesthesia	Chow 2018	0.8% (1/130)	54.9% (89/162)	RR 0.01 (0 to 0.1) ARD -54.2% (-61.7% to -45.7%)
	Vilgrain 2017	0	23.3% (48/206)	RR not estimated due to 0 cell ARD -23.3% (-29.5% to -17.6%)
rash or desquamation	Chow 2018	0	11.1% (18/162)	RR not estimated due to 0 cell ARD -11.1% (-16.9% to -6.2%)
	Vilgrain 2017	1.1% (2/174)	9.7% (20/206)	RR 0.12 (0.03 to 0.5) ARD -8.6% (-13.4% to -4.1%)
Vascular disorders				
hypertension	Chow 2018	0	14.8% (24/162)	RR not estimated due to 0 cell ARD -14.8% (-21.1% to -9.3%)
	Vilgrain 2017	2.9% (5/174)	15.5% (32/206)	RR 0.18 (0.07 to 0.46) ARD -12.7% (-18.5% to -7%)
Adverse events typically associated with radioembolization				
gastric ulcer	Chow 2018	0.8% (1/130)	0	RR not estimated due to 0 cell ARD 0.8% (-1.6% to 4.2%)
	Vilgrain 2017	2.9% (5/174)	0.5% (1/206)	RR 5.92 (0.7 to 50.19)

				ARD 2.4% (-0.4% to 6.1%)
upper gastrointestinal hemorrhage	Chow 2018	1.5% (2/130)	1.9% (3/162)	RR 0.83 (0.14 to 4.9) ARD -0.3% (-3.9% to 3.8%)
	Vilgrain 2017	2.3% (4/174)	6.3% (13/206)	RR 0.36 (0.12 to 1.1) ARD -4% (-8.4% to 0.3%)
jaundice	Chow 2018	1.5% (2/130)	1.9% (3/162)	RR 0.83 (0.14 to 4.9) ARD -0.3% (-3.9% to 3.8%)
hepatic cirrhosis	Chow 2018	0	1.2% (2/162)	RR not estimated due to 0 cell ARD -1.2% (-4.4% to 1.8%)
portal hypertension	Chow 2018	0	0.6% (1/162)	RR not estimated due to 0 cell ARD -0.6% (-3.4% to 2.3%)
radiation hepatitis	Chow 2018	1.5% (2/130)	0	RR not estimated due to 0 cell ARD 1.5% (-1% to 5.4%)
	Vilgrain 2017	0	0	NA

*In Chow 2018, it was not clearly reported whether the specific adverse events were treatment-emergent or treatment-related; however, because the lower event rates seem to match more closely with treatment-related adverse events, we have assumed they are treatment-related. Chow 2018 used the National Cancer Institute Common Terminology Criteria for Adverse Events, version 4.02 for adverse events. Vilgrain 2017 explicitly noted adverse events were treatment-related, and they used Medical Dictionary for Regulatory Activities (MeDRA) coding, version 19.0 for adverse events.

†This table has been adapted from Chow 2018 Table 3 and Vilgrain Table S9. Chow 2018 Table 3 reports adverse events in the as-treated population that were: experienced by >5% of treated patients in either group or known to be associated with selective internal radiation therapy and had an onset date on or after study treatment start date. We adapted this to include only adverse events that were (1) significantly different between groups or (2) typically associated with radioembolization). Vilgrain 2017 Table S9 reports treatment-related adverse events of interest in their per-protocol population. Vilgrain 2017 report a similar table in their main report (Table 4) that was specific to what they refer to as their safety population. However, the safety population also includes patients who did not get the treatment to which they were assigned. For example, in the radioembolization group, the safety population included 42 patients who did not receive radioembolization. Because it makes more sense to use an operational definition of treatment-related adverse events that minimizes inclusion of patients who did not receive the assigned treatment, we extracted data from Vilgrain 2017's per-protocol population (Table S9) instead. However, although Table S9 reported 29 individual treatment-related adverse events, it did not report between-group comparative statistics, thus introducing inferential impediments. Given the data available in Chow 2018, we considered the adverse event data from Vilgrain 2017 to be most useful as a comparison against the results reported in Chow 2018 (of note, Chow 2018 defines its safety population in a manner similar to what Vilgrain 2017 uses to define its per-protocol population). Given these considerations, we selected for extraction among the 29 individual treatment-related adverse events in Vilgrain 2017 only those that met our criteria for extraction from Chow 2018. Adverse events meeting these criteria and reported in Chow 2018 that were not mentioned in Vilgrain 2017 Table S9 are the following: constipation and 3 adverse events listed under "AEs typically associated with radioembolization" (namely, jaundice, hepatic cirrhosis, portal hypertension). Finally, in adapting Vilgrain 2017 Table S9 for our summary table, we also collapsed their grade categories 3, 4, and 5 into a single category of grade >3 to be more consistent with Chow 2018, and we used consistent disorder naming as used in Chow 2018 for easy comparisons (e.g.,, "skin and subcutaneous tissue disorders" in Chow 2018 vs. "Dermatologic events" in Vilgrain 2017).

‡As part of aggregation in preparation for synthesis, we present here only adverse events of any grade of severity; the original publications report adverse events in additional stratification by level of severity (see Critical Appraisal for details). Of note, where comparative statistics were reported by the authors, they were limited to p values and were not reported for any grade of adverse events (i.e., they were only reported for certain subgroups of severity). Although these tests for statistical significance informed our data extraction, we did not calculate p values for the group of any grade of adverse event when calculating the comparative statistics we report, because we consider the comparative statistics we present to be much more informative than a p value.

Evidence Synthesis

Evidence Synthesis Process

Reconciling sorafenib vs. placebo evidence and local therapy vs. sorafenib evidence

In the preceding sections, we introduced two aggregate data tables for the evidence concerning sorafenib vs. placebo and local therapy vs. sorafenib, respectively. Here, we consolidate these tables further to ease comprehension and synthesis by presenting one table presenting the evidence for both comparisons in one table. The comparison being made by the respective study (as indicated by the row with a citation in 'Author Year' format) is indicated by which intervention columns (namely, "Radioembolization", "Sorafenib", or "Placebo") have data in them. We do not repeat the detailed footnotes in this simplified table, and we give this table the simple title "Proportion of patients having specific adverse events".

Proportion of patients having specific adverse events

System organ class, disorder	Source	Radioembolization	Sorafenib	Placebo	Comparative statistics (95% CI)
Gastrointestinal disorders					
anorexia	Llovet 2008	----------	4.7% (14/297)	1.0% (3/302)	RR 4.75 (1.38 to 16.34) ARD 3.7% (1.1% to 6.8%)
	Cheng 2009	----------	12.8% (19/149)	2.7% (2/75)	RR 4.78 (1.14 to 19.99) ARD 10.1% (2.2% to 16.7%)
constipation	Chow 2018	0	5.6 % (9/162)	----------	RR not estimated due to 0 cell ARD -5.6% (-10.2% to -1.7%)
diarrhea	Llovet 2008	----------	13.1% (39/297)	3.6% (11/302)	RR 3.61 (1.88 to 6.9) ARD 9.5% (5.1% to 14.1%)
	Cheng 2009	----------	25.5% (38/149)	5.3% (4/75)	RR 4.78 (1.77 to 12.9) ARD 20.2% (10.3% to 28.4%)
	Chow 2018	1.5% (2/130)	29.6% (48/162)	----------	RR 0.05 (0.01 to 0.21) ARD -28.1% (-35.6% to -20.5%)
	Vilgrain 2017	9.8% (17/174)	78.2% (161/206)	----------	RR 0.13 (0.08 to 0.2) ARD -68.4% (-74.6% to -60.3%)
abdominal pain not otherwise specified	Llovet 2008	----------	2.7% (8/297)	1.0% (3/302)	RR 2.71 (0.73 to 10.12) ARD 1.7% (-0.6% to 4.3%)
General disorders/constitutional symptoms					
weight loss	Llovet 2008	----------	3.0% (9/297)	0.3% (1/302)	RR 9.15 (1.17 to 71.79) ARD 2.7% (0.6% to 5.3%)
fatigue	Llovet 2008	----------	7.4% (22/297)	5.3% (16/302)	RR 1.4 (0.75 to 2.61) ARD 2.1% (-1.9% to 6.2%)
	Cheng 2009	----------	20.1% (30/149)	8.0% (6/75)	RR 2.52 (1.1 to 5.78) ARD 12.1% (2% to 20.5%)
	Chow 2018	3.8% (5/130)	15.4% (25/162)	----------	RR 0.25 (0.1 to 0.63)

					ARD -11.6% (-18.3% to -4.8%)
	Vilgrain 2017	50% (87/174)	76.2% (157/206)	----------	RR 0.66 (0.56 to 0.78) ARD -26.2% (-35.3% to -16.6%)
voice changes	Llovet 2008	----------	2.0% (6/297)	0.3% (1/302)	RR 6.1 (0.74 to 50.37) ARD 1.7% (-0.2% to 4%)
Skin and subcutaneous tissue disorders					
dry skin	Llovet 2008	----------	2.7% (8/297)	1.3% (4/302)	RR 2.03 (0.62 to 6.68) ARD 1.4% (-1.1% to 4%)
alopecia	Llovet 2008	----------	4.7% (14/297)	0.7% (2/302)	RR 7.12 (1.63 to 31.05) ARD 4.1% (1.5% to 7.1%)
	Cheng 2009	----------	24.8% (37/149)	1.3% (1/75)	RR 18.62 (2.61 to 133.12) ARD 23.5% (14.9% to 31.1%)
	Chow 2018	0	9.9% (16/162)	----------	RR not estimated due to 0 cell ARD -9.9% (-15.4% to -5.2%)
	Vilgrain 2017	0	16.5% (34/206)	----------	RR not estimated due to 0 cell ARD -16.5% (-22.2% to -11.6%)
palmo-plantar erythro-dysesthesia (hand-foot skin reaction)	Llovet 2008	----------	7.1% (21/297)	1.0% (3/302)	RR 7.12 (2.15 to 23.61) ARD 6.1% (3% to 9.6%)
	Cheng 2009	----------	45.0% (67/149)	2.7% (2/75)	RR 16.86 (4.25 to 66.95) ARD 42.3% (32.1% to 50.5%)
	Chow 2018	0.8% (1/130)	54.9% (89/162)	----------	RR 0.01 (0 to 0.1) ARD -54.2% (-61.7% to -45.7%)
	Vilgrain 2017	0	23.3% (48/206)	----------	RR not estimated due to 0 cell ARD -23.3% (-29.5% to -17.6%)
rash or desquamation	Llovet 2008	----------	5.4% (16/297)	3.6% (11/302)	RR 1.48 (0.7 to 3.13) ARD 1.7% (-1.7% to 5.3%)
	Cheng 2009	----------	20.1% (30/149)	6.7% (5/75)	RR 3.02 (1.22 to 7.47) ARD 13.5% (3.7% to 21.6%)
	Chow 2018	0	11.1% (18/162)	----------	RR not estimated due to 0 cell ARD -11.1% (-16.9% to -6.2%)
	Vilgrain 2017	1.1% (2/174)	9.7% (20/206)	----------	RR 0.12 (0.03 to 0.5) ARD -8.6% (-13.4% to -4.1%)
Vascular disorders					
hypertension	Llovet 2008	----------	1.7% (5/297)	0.7% (2/302)	RR 2.54 (0.5 to 13) ARD 1% (-0.9% to 3.3%)
	Cheng 2009	----------	18.8% (28/149)	1.3% (1/75)	RR 14.09 (1.96 to 101.6) ARD 17.5% (9.5% to 24.6%)
	Chow 2018	0	14.8% (24/162)	----------	RR not estimated due to 0 cell ARD -14.8% (-21.1% to -9.3%)

	Vilgrain 2017	2.9% (5/174)	15.5% (32/206)	----------	RR 0.18 (0.07 to 0.46) ARD -12.7% (-18.5% to -7%)
Adverse events typically associated with radioembolization					
gastric ulcer	Chow 2018	0.8% (1/130)	0	----------	RR not estimated due to 0 cell ARD 0.8% (-1.6% to 4.2%)
	Vilgrain 2017	2.9% (5/174)	0.5% (1/206)	----------	RR 5.92 (0.7 to 50.19) ARD 2.4% (-0.4% to 6.1%)
upper gastrointestinal hemorrhage	Chow 2018	1.5% (2/130)	1.9% (3/162)	----------	RR 0.83 (0.14 to 4.9) ARD -0.3% (-3.9% to 3.8%)
	Vilgrain 2017	2.3% (4/174)	6.3% (13/206)	----------	RR 0.36 (0.12 to 1.1) ARD -4% (-8.4% to 0.3%)
jaundice	Chow 2018	1.5% (2/130)	1.9% (3/162)	----------	RR 0.83 (0.14 to 4.9) ARD -0.3% (-3.9% to 3.8%)
hepatic cirrhosis	Chow 2018	0	1.2% (2/162)	----------	RR not estimated due to 0 cell ARD -1.2% (-4.4% to 1.8%)
portal hypertension	Chow 2018	0	0.6% (1/162)	----------	RR not estimated due to 0 cell ARD -0.6% (-3.4% to 2.3%)
radiation hepatitis	Chow 2018	1.5% (2/130)	0	----------	RR not estimated due to 0 cell ARD 1.5% (-1% to 5.4%)
	Vilgrain 2017	0	0	----------	NA

There are some obvious limitations in the above data, including marked imprecision in many estimates any several adverse events not having data for all 3 options. For instance, the lower and upper 95% CI bounds for the RRs for some comparisons vary by more than an order of magnitude, whereas others may not be quite as extreme but still rather imprecise. Although precision is ultimately best assessed at the absolute level (rather than the relative level), and although the corresponding ARDs often seem less imprecise, we still have very serious concerns about imprecision for many estimates. Furthermore, we note ARDs are not considered to be as 'portable' across varying baseline risks compared to RRs; this may be mathematically apparent simply because the ARD is – by definition – inextricably linked to a particular baseline risk. For instance, consider two studies evaluating the effect of a given intervention vs. placebo on the risk of outcome X. Consider further that the absolute risks in the intervention and placebo groups are 3% vs. 2% in the first study and 13.5% vs. 9% in the second study. The corresponding ARDs are 1% and 4.5%, which might suggest a 'different' effect of the intervention in these studies; in fact, however, the effect of the intervention is the same, namely that it increases a person's risk by a relative 50% (RR 1.5), and the only reason for the different ARD between the two studies is the different absolute risk in the placebo group.

Considering the above along with: (1) the imprecision in the RRs (which we would want to expressly incorporate into any quantified estimate), (2) the heterogeneity in the absolute risks against which the RRs would be applied (e.g., we note particular heterogeneity for the absolute risks in the sorafenib group), and (3) the inability to compare across the three options in a direct manner for some outcomes, we consider quantitative summarization of these results to be impractical/unfeasible outside of conducting a network meta-analysis (NMA) for each specific adverse event. Conducting an NMA for each of the adverse events is beyond the scope of this evidence report; furthermore, we note we would have concerns about

conducting or making inferences from an NMA based on these data due to concerns in the domains of transitivity (a fundamental principle underlying the validity of NMA) and coherence of the network estimates. We consider it inappropriate to present a simple range for each option, largely – but not solely – due to (1) the marked heterogeneity in the absolute risks for sorafenib and (2) the intent to present all 3 options in a side-by-side manner. For instance, consider diarrhea, where the absolute risks for sorafenib range from 13.1% to 78.2% (based on the values 13.1%, 25.5%, 29.6%, and 78.2%). The absolute risks for the placebo group range from 3.6% to 5.3% (based on those 2 values alone), but simply presenting both ranges in a side-by-side manner would arguably be misleading, because the upper bound of sorafenib's range does not have a corresponding placebo value. In other words, given the placebo group range, the corresponding range for sorafenib would be 13.1% to 25.5%. An attempt to use the RRs for sorafenib vs. placebo to 'impute' the 'missing' (hypothetically missing) placebo group estimates from the data on radioembolization vs. sorafenib would still suffer from marked imprecision: The range of RRs for sorafenib vs. placebo for diarrhea (based on the lower of the 2 lower 95% CI bounds and the higher of the 2 upper 95% CI bounds) is 1.14 to 19.99. Lastly, simply presenting the range of 3.6% and 5.3% for placebo and 13.1% to 25.5% for sorafenib would arguably be inappropriate, because this would entirely discard the additional data from the comparisons of radioembolization vs. sorafenib.

Considering all these factors, and because the specific adverse events were included in Scoping as a means of offering some contextualization of the primary outcomes of serious (treatment-related) adverse events and (treatment-related) adverse events leading to discontinuation, we synthesize the above findings without quantitation.

Under the above approach of a qualitative synthesis but still aiming to give patients and clinicians some idea of commonality (e.g., relative frequency of gastric ulcer vs. anorexia), we consider it reasonable to present the specific adverse events in 2 broad categories of general adverse events and potentially serious, but not common adverse events. In reporting general adverse events, we set an (admittedly arbitrary) threshold for inclusion in the simplified summary statement of ≥1 estimate having a lower 95% CI bound for the ARD of ≥2%.

The above yields anorexia, diarrhea, fatigue, alopecia, palmoplantar erythrodysesthesia (hand-foot skin reaction), rash or desquamation, and hypertension as adverse events of consideration for sorafenib (though one could debate the appropriateness of including hypertension given it is not a patient-relevant outcome by itself). Because some symptoms can be experienced due to the disease itself (in particular, anorexia and fatigue can both be seen), we mention these in the radioembolization option and best supportive care option, but we make it clear these are more commonly experienced with sorafenib. For radioembolization, we note the standard risks of bleeding or infection with any procedure or surgery, and we also note the uncommon but potentially serious risks of gastric ulcer and radiation hepatitis (the latter two being based on Chow 2018 and Vilgrain 2017). We acknowledge neither gastric ulcer nor radiation hepatitis met conventional statistical significance (let alone the threshold set for general adverse events), but these risks have a strong prior probability of being real risks and are even categorized as being typically associated with radioembolization by some literature (including some literature addressed in this report). We also take into consideration the fact that the studies may have been underpowered to detect these findings statistically (i.e., at $\alpha = 0.05$). Considering all this, we felt it reasonable to include these specific risks in our bolded synthesis statement. We did not consider other adverse events to have sufficient signal for inclusion here, especially considering that some (e.g., portal hypertension, hepatic cirrhosis, gastrointestinal bleed as a specific type of bleeding) could be attributable to disease progression rather than the intervention (particularly in the absence of a corresponding placebo group risk).

Lastly, because best supportive care is (or at least should be) occurring 'in the background' when someone elects to pursue specific management via radioembolization or sorafenib, we do not address any side effects that might be related to best supportive care (e.g., side effects that might occur from opiates being used for pain control).

Evidence Synthesis Results

Radioembolization

As with any treatment like this, there is a risk of infection and bleeding, but these things are not common. Other risks that can be serious but are not common include stomach ulcer and inflammation of the liver from the treatment. Some people can also have symptoms like fatigue or loss of appetite, but these things may be due to the illness itself and are probably less likely to happen compared with someone who takes sorafenib. You can talk more about these and other possible risks or side effects with your clinician.

Sorafenib

Sorafenib may cause loss of appetite, diarrhea, fatigue, hair thinning or hair loss, increased blood pressure, and different kinds of skin rashes or reactions. Sometimes the skin rashes or reactions can be serious. Although some people may have fatigue or loss of appetite from the illness itself, these symptoms probably occur more often in people taking sorafenib. You can talk more about these and other possible risks or side effects with your clinician.

Best supportive care

People may have loss of appetite or fatigue from the illness itself, but these things are probably less likely to happen compared with taking sorafenib.

(Certainty not rated; based on Cheng 2009, Chow 2018, Llovet 2008, and Vilgrain 2017)

FAQ 6: If I have other health issues, do they change how the treatment works?

Radioembolization vs. sorafenib vs. placebo

Evidence Review and Synthesis

We specified in Scoping that we would assess this outcome via assessment of whether comorbidities yielded convincing evidence for heterogeneity of treatment effect within the randomized, controlled trials (or systematic reviews thereof) that assessed these options.

As noted in Critical Appraisal, where subgroup analyses were conducted, they routinely showed consistent findings for most subgroups examined, and we note many of the subgroups were not focused on comorbidities, but rather, basic demographics (e.g., sex), baseline markers (or potential markers) of disease severity (e.g., ECOG score, Child-Pugh classification, α-fetoprotein level), or tumor characteristics (e.g., presence of vascular invasion, nodular vs. infiltrative tumor type).

As such, we do not forward any firm conclusions about whether comorbidities change how the treatment works, and we note that changing how the treatment works is a different question from assessing whether comorbidities might make an option more or less desirable (e.g., if a person already has difficult-to-manage diarrhea secondary to inflammatory bowel disease (IBD), this might enter into decision-making enough that the clinician and patient shy away from sorafenib in favor of best supportive care alone or radioembolization). Nevertheless, we include this concept as a bolded synthesis statement given patients may still wonder about how, for instance, something like the above example of IBD could factor into deciding on a treatment even if it does not have empirical evidence of altering the treatment's efficacy or harms *per se*.

Radioembolization vs. sorafenib vs. placebo

If you have other health issues, you can talk with your clinician about whether one option might be better or worse for you based on things like the side effects or possible problems with other medications you might take.

(Certainty not rated; based on absence of evidence for heterogeneity for treatment effect based on comorbidities from 4 trials [Cheng 2009, Chow 2018, Llovet 2008, Vilgrain 2017] and basic clinical practice paradigm of taking side effects and medication interactions into account when helping a patient decide on the best treatment option for him/her)

Note: Because the sample decision aid was already lengthy given the complexity of this topic and because this bolded synthesis statement did not seem to add substantively to the sample decision aid beyond what a clinician could discuss at the bedside in general, we did not progress this outcome to the sample decision aid.

Evidence report for local therapy vs. systemic therapy in patients with hepatocellular carcinoma and cirrhosis who are ineligible for resection or transplantation

Search and Selection

EBSCO Information Services

June 12, 2020
Searches updated July 6, 2020

Prepared by EBSCO Health Innovations and EBM Development Department (in alphabetic order):

Brian S. Alper, MD, MSPH, FAAFP, FAMIA, Chief Medical Knowledge Officer

Eric Manheimer, PhD, EBM Editor

Martin Mayer, DMSc, MS, PA-C, Senior Clinical/Medical Editor

Methods

First-round searching will consist of searching DynaMed Plus summaries pertinent to the topic. The search starts with DynaMed Plus because DynaMed Plus content is based on a thorough, systematic search and active literature surveillance system.

> Systematic literature surveillance (SLS) has been a cornerstone of creating DynaMed content since its inception. Currently, the SLS team works closely with members of McMaster University's Department of Health Research Methods, Evidence, and Impact to define and refine optimal search strategies, utilizing search terms for methodology and for clinical concepts (especially diagnosis and treatment) for each journal. These filters for PubMed searches are derived from PubMed Clinical Queries, McMaster University Hedges system, and DynaMed search filters and were found to capture >95% of relevant valid articles for use in DynaMed.[1]
> [1]Alper BS, Iorio A. Optimising search filters for active literature surveillance: a concordance study. Global Evidence Summit. Cape Town, South Africa; September 2017.
>
> The SLS team, in collaboration with McMaster University, continually evaluates and improves these filters to keep abreast of the ever-changing medical literature landscape. PubMed searches are performed daily. Each article is assessed by an external screener for clinical relevance, then reviewed by internal staff for both clinical relevance and validity. Any disagreements are settled by the Deputy Editor of the SLS team.
> In addition, the SLS team continually monitors the Journal source list against the list of journals catalogued in MEDLINE and existing DynaMed content to determine the set of journals that provide the highest yield for valid, clinically relevant content. As of November 2018, 507 Journals are systematically searched (see http://dynamed.com/home/content/content-sources). The journal selection threshold leads to routine adjustments to the number of journals selected.

Second-round searching will involve PubMed/MEDLINE® searches for intervention-specific systematic reviews or guidelines that involve a systematic review that are more recent than those identified in first-round searching.

Third-round searching will involve searches for original evidence reports. Such searching will be limited to areas considered insufficiently addressed by earlier searches, such as focused concepts published after search dates in systematic reviews.

Fourth-round searching will involve citation tracing from sources found in the preceding rounds of searching. This may be done to identify additional evidence sources if the results from the preceding rounds of searching are insufficient.

Additional selective searching, such as tracing citations from articles found or seeking related articles, may be done for additional concepts discovered.

Additional evidence updates may also be added continuously from feedback from any source.

FAQ Summary

Results

First-round searching – DynaMed

DynaMed subsection	Number of article citations	Number of unique references included
Management of Hepatocellular Carcinoma in Adults > Locoregional Therapy > Ablation > Efficacy of Ablation (June 11, 2020; last checked July 6, 2020)	14	0
Management of Hepatocellular Carcinoma in Adults > Locoregional Therapy > Arterially Directed Therapy > Transarterial embolization (TAE) and Transarterial Chemoembolization (TACE) > Efficacy of TAE and TACE (June 11, 2020; last checked July 6, 2020)	4	0
Management of Hepatocellular Carcinoma in Adults > Locoregional Therapy > Arterially Directed Therapy > Transarterial Chemoembolization with Drug-eluting Beads (DEB-TACE) > Efficacy of DEB-TACE (June 11, 2020; last checked July 6, 2020)	4	0
Management of Hepatocellular Carcinoma in Adults > Locoregional Therapy > Arterially Directed Therapy > Transarterial Radioembolization (TARE) > Efficacy of TARE (June 11, 2020; last checked July 6, 2020)	10	2 (Chow 2018, Vilgrain 2017)
Management of Hepatocellular Carcinoma in Adults > Locoregional Therapy > External Beam Radiation Therapy (EBRT) > Efficacy of EBRT (June 11, 2020; last checked July 6, 2020)	5	0
Management of Hepatocellular Carcinoma in Adults > Systemic Therapy > Primary Systemic Therapy > First-line Systemic Therapy (June 11, 2020; last checked July 6, 2020)	17	4 (Finn 2020, Shao 2015, Llovet 2008, Cheng 2009)*†
Management of Hepatocellular Carcinoma in Adults > Systemic Therapy > Primary Systemic Therapy > Second-line Therapies (June 11, 2020; last checked July 6, 2020)	19	0

*Shao 2015 does not appear in Critical Appraisal, as it only served to (1) identify placebo-controlled trials of sorafenib (namely Llovet 2008 and Cheng 2009), and (2) serve as a starting point for forward searching for more current systematic reviews (see second-round searching). Because Llovet 2008 and Cheng 2009 are summarized within DynaMed as part of the summary for Shao 2015, these articles are captured as separate references in the above tallies; accordingly, we do not include Shao 2015 in fourth-round searching (citation tracing).

†Finn 2020 is included as an ancillary reference in accordance with Scoping

Second-round searching – searches for intervention-specific systematic reviews

Database	Search string	Number of search results	Number of unique references included
PubMed June 11, 2020; last checked July 6, 2020	sorafenib AND hepatocellular carcinoma AND (cirrhosis OR Child-Pugh A OR Child-Pugh B OR Child Pugh A OR Child Pugh B) AND (systematic[sb] OR meta-analysis[ti]) AND (2015/01/01[PDAT]:3000/12/31[PDAT])	8 (on June 11, 2020) 8 (on July 6, 2020)	2 (Finn 2018, Roccarina 2017)*
PubMed June 11, 2020; last checked July 6, 2020	hepatocellular carcinoma AND (((radiofrequency OR microwave) AND ablation) OR (transarterial AND (chemoembolization OR radioembolization)) OR embolization OR selective internal radiation therapy OR selective internal radiotherapy OR external beam radiotherapy OR stereotactic body radiation therapy OR stereotactic body radiotherapy OR linear accelerator OR LINAC OR CyberKnife OR Varian) AND (systemic therapy OR sorafenib OR regorafenib OR lenvatinib OR cabozantinib OR ramucirumab OR (atezolizumab AND bevacizumab)) AND systematic[sb]	43 (on June 11, 2020) 43 (on July 6, 2020)	1 (Abdel-Rahman 2020)†

*McNamara 2018 is a review of questionable reliability (for instance, it says its search terms were "sorafenib, HCC and Child-Pugh B"), allows inclusion of observational and randomized trial data, and seems to miss important trials. It was therefore judged not useful for this evidence report. Finn 2018 and Roccarina 2017 are addressed in Scoping (they are not included in Critical Appraisal), yielding the trials Llovet 2008 and Cheng 2009 (for which we prepared full summaries; see third-round searching as well).

†The following articles were not progressed to Critical Appraisal: Abdel-Rahman 2017, Chen 2019, Kalogeridi 2015, Katsanos 2017, Zou 2019. Abdel-Rahman 2017 is a systematic review with all included trials comparing different approaches to local therapy (specifically external beam radiotherapy plus chemoembolization vs. chemoembolization alone *or* radioembolization alone vs. chemoembolization alone). Chen 2019 is a systematic review including non-scoped interventions (including traditional Chinese medicine) and presents results for a network meta-analysis. Because of how network meta-analysis works, this means any estimates for the scoped interventions will be – to some extent – influenced by the data for non-scoped interventions (via how indirect evidence can inform the network). They also only identify 2 of 3 trials we identified for local therapy vs. systemic therapy. We judged this systematic review not to add anything meaningful to what could be ascertained from other sources. Kalogeridi 2015 is described as a systematic review in its title, but it provides no description of its search methods and appears better described as a narrative review. We did not consider this article further. Katsanos 2017 is a systematic review and network meta-analysis of different types of transarterial embolization therapies alone or in combination with local ablative therapies or adjuvant systemic treatment. Zou 2019 is a systematic review that included randomized trials and observational studies and appears only to report outcomes separately for certain subgroup analyses (though we note the observational studies contribute a relative minority of weight to the meta-analysis for overall survival, the addition of observational studies is unhelpful and undesirable in this context). They also only identify 2 of 3 trials we identified for local therapy vs. systemic therapy. We judged this systematic review not to add anything meaningful to what could be ascertained from other sources.

PARSE Mapping

PARSE Name	Search string	Number of search results	Number of new references selected for inclusion
HCC second round string #1	sorafenib AND hepatocellular carcinoma AND (cirrhosis OR Child-Pugh A OR Child-Pugh B OR Child Pugh A OR Child Pugh B) AND (systematic[sb] OR meta-analysis[ti]) AND (2015/01/01[PDAT]:3000/12/31[PDAT])	8 (on June 12, 2020) 8 (on July 6, 2020)	0
HCC second round string #2	hepatocellular carcinoma AND (((radiofrequency OR microwave) AND ablation) OR (transarterial AND (chemoembolization OR radioembolization)) OR embolization OR selective internal radiation therapy OR selective internal radiotherapy OR external beam radiotherapy OR stereotactic body radiation therapy OR stereotactic body radiotherapy OR linear accelerator OR LINAC OR CyberKnife OR Varian) AND (systemic therapy OR sorafenib OR regorafenib OR lenvatinib OR cabozantinib OR ramucirumab OR (atezolizumab AND bevacizumab)) AND systematic[sb]	41* (on June 11, 2020) 41* (on July 6, 2020)	0

*The discrepancy from the count in the above search is due to the cutover to the new PubMed/MEDLINE® (PARSE still uses the previous version of PubMed/MEDLINE®).

Third-round searching – searches for original evidence reports

Database	Search string	Number of search results	Number of unique references included
PubMed June 12, 2020; last checked July 6, 2020	sorafenib AND hepatocellular carcinoma AND (cirrhosis OR Child-Pugh A OR Child-Pugh B) AND randomized controlled trial NOT (systematic[sb] OR meta-analysis[ti]) AND (2016/01/01[EDAT]:3000/12/31[EDAT])	27 (on June 12, 2020) 28 (on July 6, 2020)	0
PubMed June 12, 2020; last checked July 6, 2020	hepatocellular carcinoma AND (((radiofrequency OR microwave) AND ablation) OR (transarterial AND (chemoembolization OR radioembolization)) OR embolization OR selective internal radiation therapy OR selective internal radiotherapy OR external beam radiotherapy OR stereotactic body radiation therapy OR stereotactic body radiotherapy OR linear accelerator OR LINAC OR CyberKnife OR Varian) AND (systemic therapy OR sorafenib OR regorafenib OR lenvatinib OR cabozantinib OR ramucirumab OR (atezolizumab AND bevacizumab)) AND randomized controlled trial NOT (systematic[sb] OR meta-analysis[ti])	72 (on June 12, 2020) 74 (on July 6, 2020)	1 (Yoon 2018)

PARSE Mapping

PARSE Name	Search string	Number of search results	Number of new references selected for inclusion
HCC third round string #1	sorafenib AND hepatocellular carcinoma AND (cirrhosis OR Child-Pugh A OR Child-Pugh B) AND randomized controlled trial NOT (systematic[sb] OR meta-analysis[ti]) AND (2016/01/01[EDAT]:3000/12/31[EDAT])	27 (on June 12, 2020) 27 (on July 6, 2020)	0
HCC third round string #2	hepatocellular carcinoma AND (((radiofrequency OR microwave) AND ablation) OR (transarterial AND (chemoembolization OR radioembolization)) OR embolization OR selective internal radiation therapy OR selective internal	69* (on June 12, 2020) 70* (on July 6, 2020)	0

	radiotherapy OR external beam radiotherapy OR stereotactic body radiation therapy OR stereotactic body radiotherapy OR linear accelerator OR LINAC OR CyberKnife OR Varian) AND (systemic therapy OR sorafenib OR regorafenib OR lenvatinib OR cabozantinib OR ramucirumab OR (atezolizumab AND bevacizumab)) AND randomized controlled trial NOT (systematic[sb] OR meta-analysis[ti])		

*The discrepancy from the count in the above search is due to the cutover to the new PubMed/MEDLINE® (PARSE still uses the previous version of PubMed/MEDLINE®).

Fourth-round searching – citation tracing

Source	Tracing	Number of unique references included
Roccarina 2017	Tracing for relevant trials not already retrieved via the above methods	0*
Finn 2018	Tracing for relevant trials not already retrieved via the above methods	0†

*Captures Llovet 2008, but no mention of Cheng 2009 (including in their list of excluded studies)
†Captures Llovet 2008 and Cheng 2009

Search summary:

Number of articles screened and selected

	Screened	Selected*
1st-round (DynaMed)	73	6
2nd-round (Systematic review searches)	51	3
3rd-round (Original article searches)	102	1
4th-round (Citation tracing)	6	0
Total	*232*	*10†*

*Number shown is after deduplication

†Only 7 of the 10 appear in Critical Appraisal and Evidence Synthesis, as the utility of 3 (Finn 2018, Roccarina 2017, and Shao 2015) were strictly for citation tracing and/or serving as a timepoint for forward searching for more current systematic reviews and/or randomized trials

	Articles selected for evaluation (Author Year)
Systematic reviews	4 (Abdel-Rahman 2020, Finn 2018*, Roccarina 2017*, Shao 2015*)
Randomized, controlled trials	6 (Cheng 2009, Chow 2018, Finn 2020, Llovet 2008, Vilgrain 2017, Yoon 2018)
Other article types	0

*This reference only appears in Search and Selection, as its only purpose for this evidence report was for citation tracing and/or serving as a timepoint for forward searching for more current systematic reviews and/or randomized trials

Evidence report for local therapy vs. systemic therapy in patients with hepatocellular carcinoma and cirrhosis who are ineligible for resection or transplantation

Critical Appraisal

EBSCO Information Services

July 16, 2020

Prepared by EBSCO Health Innovations and EBM Development Department (in alphabetic order):

Brian S. Alper, MD, MSPH, FAAFP, FAMIA, Chief Medical Knowledge Officer

Eric Manheimer, PhD, EBM Editor

Martin Mayer, DMSc, MS, PA-C, Senior Clinical/Medical Editor

Methods

For each study selected for critical appraisal, we assessed for criteria needed for Level 1 [likely reliable] evidence ratings in DynaMed's Levels of Evidence criteria in which evidence quality is labeled in 1 of 3 levels:

- **(level 1 [likely reliable] evidence)** for studies with clinical outcomes and minimal risk of bias;
- **(level 2 [mid-level] evidence)** for studies with clinical outcomes and significant methodological or statistical limitations; and
- **(level 3 [lacking direct] evidence)** for reports that do not include scientific analysis of clinical outcomes or for study types that do not include comparisons between groups.

The specific concepts we critically assessed are fully mapped to cover all the criteria used in GRADE. We summarized key concerns identified that reduce the certainty of evidence in categories used for GRADE methods (i.e., risks of bias, indirectness, imprecision, inconsistency, and other considerations if relevant). In the study-level summaries inconsistency assessment was applied to consistency of within-study results – this applies to consistency across related outcomes for both original studies and for systematic reviews, and also applies to consistency across studies when applied to systematic reviews.

FAQ Summary

FAQ 1: What does the treatment involve?

FAQ 2: How long will I live?

FAQ 3: How long will my cancer stay controlled?

FAQ 4: Will the treatment change my quality of life?

FAQ 5: What are the risks or side effects?

Results

General notes:

- We use the terms “radioembolization”, “selective internal radiation therapy”, “selective internal radiotherapy”, and “SIRT” interchangeably in this evidence report.
- We specified in Scoping that we would assess whether comorbidities may lead to heterogeneity of treatment effect as assessed within randomized, controlled trials (or systematic reviews thereof).
    - Where subgroup analyses were conducted, they routinely showed consistent findings for most subgroups examined, and we note many of the subgroups were not focused on comorbidities, but rather, basic demographics (e.g., sex), baseline markers (or potential markers) of disease severity (e.g., ECOG score, Child-Pugh classification, α-fetoprotein level), or tumor characteristics (e.g., presence of vascular invasion, nodular vs. infiltrative tumor type).
    - As also noted in Scoping, because subgroup analyses sacrifice the virtues of randomization (barring a design that stratifies randomization on the subgroup variable of interest) and can often lead to spurious ‘findings’, they are to be viewed as hypothesis generating in the absence of rigorous verification (Oxman 1992, Sun 2012, Sun 2014, Wallach 2017). We thus approached any subgroup analysis with commensurate scrutiny, and given the overall pattern of the evidence, we kept our appraisal documentation for subgroup analyses very brief but still included documentation for each applicable study.
    - We do acknowledge one potential exception, namely that Cheng 2009 and Llovet 2008 both assessed sorafenib vs. placebo and both provide some support for the hypothesis that there may be heterogeneity for treatment effect based on the presence of extrahepatic spread. However, this inference becomes very complicated to contextualize appropriately, because Chow 2018 and Vilgrain 2017 (which assessed radioembolization vs. sorafenib) and Yoon 2018 (which studied transarterial chemoembolization plus external beam radiotherapy vs. sorafenib) all excluded patients with extrahepatic spread; accordingly, there is no way to assess this finding in these studies. The key question then seems to become: Even if one assumes the findings from Cheng 2009 and Llovet 2008 are real, what do they mean for decision-making? Although we acknowledge there may be attenuation of efficacy in patients with extrahepatic spread, it would be overly simplistic to say these studies show that sorafenib is ‘ineffective’ in these patients simply because the 95% CI for that subgroup crosses the null; however, giving different quantitative estimates of efficacy based on this subgroup could also run the risk of false precision or even being misleading. In sum, although we acknowledge these data in Cheng 2009 and Llovet 2008 (including acknowledging the consistency of the finding in two separate trials), we do not progress the finding for further consideration in Evidence Synthesis for FAQ 6 (“If I have other health issues, do they change how the treatment works?”) given (1) this consideration can be had qualitatively at the bedside if appropriate/relevant, and (2) this consideration falls outside of scoped intent for assessing comorbidities (rather than levels of severity of the hepatocellular carcinoma *per se*). Nevertheless, because of the way we synthesize the outcomes involving median survival time (namely, by explicitly specifying some people may have longer or shorter survival times), we present an opportunity for the clinician to have this discussion with the patient if desired/applicable (further details are in Evidence Synthesis).

Characteristics and Results of Critically Appraised Studies/Sources

ABDEL-RAHMAN 2020
Abdel-Rahman O, Elsayed Z. Yttrium-90 microsphere radioembolisation for unresectable hepatocellular carcinoma. Cochrane Database Syst Rev. 2020;1(1):CD011313. PubMed

This Cochrane review included a comparison of radioembolization vs. sorafenib for unresectable hepatocellular carcinoma in patients with locally advanced hepatocellular carcinoma, which included two trials (Vilgrain 2017, Chow 2018) we have summarized in full below. The Cochrane review provided pooled estimates for overall survival expressed as a hazard ratio (1.12, 95% CI 0.97 to 1.3; favors radioembolization). However, this statistic is not especially helpful by itself (e.g., it would require an absolute measure of some kind to impart more meaning for clinicians and patients). We also specified in Scoping that we would prefer survival expressed as median survival if available, which we have obtained from the trial reports. Abdel-Rahman 2020 also provide pooled estimates for some named adverse events, specifically: fatigue (RR 0.43, 95% CI 0.25 to 0.73), hand-foot skin reaction (RR 0.02, 95% CI 0.00 to 0.06), skin rash (RR 0.11, 95% CI 0.04 to 0.34), diarrhea (RR 0.11, 95% CI 0.04 to 0.34), all favoring radioembolization. However, we have obtained more granular adverse event capture via our direct summarization of the trials and therefore do not consider these pooled outcomes further in this evidence report. Hypertension (RR 0.10, 95% CI 0.01 to 0.88) and hypoalbuminemia (RR 1.08, 95% CI 0.84 to 1.38) are also included, but these are not adverse events in and of themselves (i.e., they are not patient-relevant outcomes in and of themselves). Finally, we note Abdel-Rahman 2020 included quality of life as a primary outcome and found no follow-up publications tied to Vilgrain 2017 (as assessed by reviewing their references section); the relevance of this observation is contextualized in the summary for Vilgrain 2017, specifically the section on quality of life.

CHENG 2009
Cheng AL, Kang YK, Chen Z, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. Lancet Oncol. 2009;10(1):25-34. PubMed

Relevant to FAQ2, FAQ3, FAQ4, FAQ5, FAQ6

- **Study design:** randomized trial
  - phase 3
  - randomization stratified by
    - geographical region (China, Taiwan, or South Korea)
    - ECOG performance status (score of 0, 1, or 2)
    - presence or absence of macroscopic vascular invasion or extrahepatic spread (or both)
  - setting was 23 centers in China, South Korea, and Taiwan
- **Population:** adults with advanced (unresectable or metastatic) hepatocellular carcinoma
  - no previous systemic therapy
  - Child–Pugh liver function class A
  - ECOG performance status of 0, 1, or 2
  - life expectancy of ≥12 weeks
  - if previous local therapy to target lesion, target lesion must have increased ≥25% and local therapy must have ceased ≥4 weeks prior to enrollment
  - no known HIV infection, substance use disorder, or cardiovascular disease

- **Intervention:** sorafenib
  - 800 mg/day
    - 400 mg twice daily on a continuous basis
- **Comparison:** placebo

- **Study inclusion criteria:** no additional

- **Number of participants:** 227
  - sorafenib, 150
  - placebo, 76
- **Duration of follow-up:**

- **Results:** not explicitly stated, but median overall survival for each group (noted below) offers some insight into duration of follow-up (cannot follow-up beyond death)

overall survival (months median, 95% CI) and HR (95% CI) in ITT* analysis:

- 6.5 months (5.56 to 7.56) in sorafenib group
- 4.2 months (3.75 to 5.46) in placebo group
- HR 0.68 (0.50 to 0.93)
- consistent effects across subgroups defined by following variables
  - age ≥ 65 years (Y/N)
  - macroscopic vascular invasion (Y/N)
  - extrahepatic spread (Y/N)
  - macroscopic vascular invasion or extrahepatic spread (Y/N)
  - ECOG performance status (Y/N)
  - hepatitis B infection (Y/N)

*ITT population included all randomized patients

overall survival (%) at 6 months in ITT analysis:

- 53.3% in sorafenib group
- 36.7% in placebo group
- no CI or p value reported, but associated with same analysis yielding the above HR (0.68, 95% CI 0.50 to 0.93)

time to symptomatic progression* (months median, 95% CI) and HR (95% CI) in ITT analysis:

- 3.5 months (2.80 to 4.24) in sorafenib group
- 3.4 months (2.40 to 4.08) in placebo group
- HR 0.90 (0.67 to 1.22)

*Defined as deterioration to ECOG performance status 4 or a change from baseline score on the 8-item, symptom focused Functional Assessment of Cancer Therapy–Hepatobiliary Symptom Index (FHSI-8) questionnaire, associated with a deterioration of symptoms

time to radiologic progression* (months median, 95% CI) and HR (95% CI) in ITT analysis:

- 2.8 months (2.63 to 3.58) in sorafenib group
- 1.4 months (1.35 to 1.55) in placebo group
- HR 0.57 (0.42 to 0.79)

*As defined by RECIST criteria

quality of life
measured with the Functional Assessment of Cancer Therapy–Hepatobiliary (FACT–HP) questionnaire; authors reported no difference between the two groups (data was not shown)

adverse events

Proportion of patients having adverse events in the safety population*

Outcome	Sorafenib % (n/N)	Placebo % (n/N)
≥1 treatment-emergent adverse event	98.0% (146/149)	94.7% (71/75)
≥1 drug-related adverse event	81.9% (122/149)	38.7% (29/75)
treatment discontinuation due to adverse events	19.5% (29/149)	13.3% (10/75)
≥1 treatment-emergent serious adverse† event	47.7% (71/149)	45.3% (34/75)
≥1 drug-related serious adverse event	8.7% (13/149)	1.3% (1/75)

*Safety analysis included all patients who received at least one dose of study drug; no comparative statistics were reported
†Defined as those that were life-threatening, resulted in death, required patient hospitalization or prolongation of hospitalization, or resulted in a persistent or significant disability or incapacity

Drug-related specific adverse events*

Outcome	Sorafenib group % (n/N)		Placebo group % (n/N)	
	all	grade 3/4	all	grade 3/4
hand-foot skin reaction	45.0% (67/149)	10.7% (16/149)	2.7% (2/75)	0% (0/75)
diarrhea	25.5% (38/149)	6.0% (9/149)	5.3% (4/75)	0% (0/75)
alopecia	24.8% (37/149)	..	1.3% (1/75)	..
fatigue	20.1% (30/149)	3.4% (5/149)	8.0% (6/75)	1.3% (1/75)
rash/desquamation	20.1% (30/149)	0.7% (1/149)	6.7% (5/75)	0% (0/75)
hypertension	18.8% (28/149)	2.0% (3/149)	1.3% (1/75)	0% (0/75)
anorexia	12.8% (19/149)	0% (0/149)	2.7% (2/75)	0% (0/75)
nausea	11.4% (17/149)	0.7% (1/149)	10.7% (8/75)	1.3% (1/75)

*Authors provided data for drug-related specific adverse events that occurred in ≥ 10% of patients in any study group; no comparative statistics were reported

- **Certainty:**
  - efficacy outcomes
    - overall survival (months median)
      - Moderate certainty
        - downgrade due to the combination of the certainty statement above being for a single trial rather than a body of evidence (thus precluding assessment of consistency/replicability in findings) and the finding being imprecise (HR 0.68 with 95% CI 0.50 to 0.93 and a median difference of 2.3 months)
    - time to symptomatic progression (months median)
      - Moderate certainty
        - imprecision downgrade because wide CIs for differences between groups compatible with either treatment being better

            - concern regarding this being a single trial (see overall survival outcome) also applies here, but was not enough to warrant a separate downgrade
    - time to radiologic progression
        - Low certainty
            - indirectness downgrade because surrogate outcome measure
            - imprecision downgrade (HR 0.57 with 95% CI 0.42 to 0.79 and a median difference of 1.4 months)
            - concern regarding this being a single trial (see overall survival outcome) also applies here, but was not enough to warrant a separate downgrade
- adverse events
    - Moderate certainty
        - ≥1 drug-related serious adverse event
        - downgraded due to the combination of the certainty statement above being for a single trial rather than a body of evidence (thus precluding assessment of consistency/replicability in findings) and the authors not providing any estimate of precision
    - Low certainty
        - ≥1 treatment-emergent serious adverse event
        - treatment discontinuation due to adverse events
        - downgraded due to the combination of the certainty statement above being for a single trial rather than a body of evidence (thus precluding assessment of consistency/replicability in findings), the authors not providing any estimate of precision, and the indirectness of the outcome (not clear if serious adverse events are attributable to the treatment)
    - certainty not rated for out-of-scope adverse event outcomes such as ≥1 treatment-emergent or drug-related adverse event (which we also note are both too broad to be meaningful)
    - certainty not rated for specific/named adverse events, as these were scoped to inform the primary outcomes of interest (namely serious adverse events and treatment discontinuation due to adverse events)

CHOW 2018

Chow PKH, Gandhi M, Tan SB, et al; Asia-Pacific Hepatocellular Carcinoma Trials Group. SIRveNIB: Selective Internal Radiation Therapy Versus Sorafenib in Asia-Pacific Patients With Hepatocellular Carcinoma. J Clin Oncol. 2018 Jul 1;36(19):1913-1921. PubMed

Relevant to FAQ2, FAQ3, FAQ4, FAQ5

- **Study design:** randomized trial
  - phase 3
  - randomization stratified by center and presence of portal vein thrombosis
  - setting was 11 countries in Asia-Pacific region
- **Population:** adults with unresectable locally advanced HCC
  - BCLC stage B or C (and 1 patient with stage A)
    - proportion with BCLC C
      - 48.4% in selective internal radiation therapy group
      - 44.9% in sorafenib group
    - Table 1 indicates that 1 patient (in sorafenib group) was BCLC stage A
  - no extrahepatic disease
  - Child-Pugh liver function class A or B
    - Table 1 indicates a small minority of patients were not Child-Pugh A or B (~1.4% of intention-to-treat population and as-treated population)
  - ECOG performance status of 0 or 1
  - not amenable to curative treatment modalities
  - had not received > 2 previous administrations of hepatic artery–directed therapy, hepatic artery–directed treatment within 4 weeks, or prior treatment with radiotherapy or sorafenib or vascular endothelial growth factor inhibitors
- **Intervention:** selective internal radiation therapy with yttrium-90 (^{90}Y) resin microspheres
  - allowed single delivery only
- **Comparison:** sorafenib
  - 800 mg/d
    - 400 mg twice daily on a continuous basis

- **Study inclusion criteria:** no additional
- **Number of participants:** 360
  - selective internal radiation therapy, 182
  - sorafenib, 178
- **Duration of follow-up:** not explicitly stated (authors note a planned duration but not the actual duration), but median overall survival for each group (noted below) offers some insight into duration of follow-up (cannot follow-up beyond death)

- **Results:**

overall survival (months median, 95% CI) and HR (95% CI) in ITT analysis:

- 8.8 months (7.5 to 10.8) in selective internal radiation therapy group (135 events)
- 10.0 months (8.6 to 13.8) in sorafenib group (131 events)
- HR 1.12 (0.9 to 1.4)
  - no statistically significant difference in overall survival between treatment groups in preplanned subgroups

Overall survival rates at 6, 12, and 18 months in the intention-to-treat population

Time point	Selective internal radiation therapy survival rate, % (95% CI)	Sorafenib survival rate, % (95% CI)	Difference in survival rates, % (95% CI)
6 months	67.7 (60.6 to 74.7)	67.5 (60.4 to 74.6)	0.2 (-9.7 to 10.2)
12 months	37.3 (29.8 to 44.7)	47.0 (39.3 to 54.7)	-9.8 (-20.5 to 1.0)
18 months	26.9 (19.9 to 33.9)	31.1 (23.6 to 38.5)	-4.2 (-14.4 to 6.1)

progression-free survival* (months median, 95% CI) and HR (95% CI) in ITT analysis:

- 5.8 (3.7 to 6.3) in selective internal radiation therapy group (146 events)
- 5.1 (3.9 to 5.6) in sorafenib group (152 events)
- HR 0.89 (0.7 to 1.1)**

*Defined according to RECIST

**Difference was significant in as-treated analysis (6.3 months vs. 5.2 months; HR 0.73 [95% CI 0.6 to 0.9])

quality of life

- no significant differences in the EuroQol-5 Dimension (EQ-5D) questionnaire comparing selective internal radiation therapy vs. sorafenib groups throughout the study (EQ-5D collected at each visit) in either the ITT or as-treated populations

Quality of life as assessed by the EuroQol-5 Dimension*

Population and timing of assessment	Selective internal radiation therapy (mean [95% CI])	Sorafenib (mean [95% CI])
Intention-to-treat	*N = 182*	*N = 178*
Baseline	83 (79 to 87)	85 (82 to 88)
9 months	72 (61 to 83)	69 (58 to 80)
As-treated	*N = 130*	*N = 162*
Baseline	86 (82 to 90)	86 (83 to 89)
9 months	71 (59 to 83)	69 (57 to 81)

*As estimated from Fig A5 in data supplement for Chow 2018. The EuroQol-5 Dimension (EQ-5D) assesses 5 dimensions of health status: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Depending on the subtype of EQ-5D a person uses, these may be assessed slightly differently (e.g., on a 3-level scale vs. a 5-level scale). Higher scores are better. The authors present the overall EQ-5D scores on a scale from 0 to 1. We have multiplied the data by 100 for easier comprehension. Extraction at 9 months informed by median overall survival estimates and the 9-month timepoint also having 95% CIs (next closest timepoint with 95% CIs was 12 months).

adverse events

Proportion of patients having adverse events in the as-treated* population

Outcome†	Sorafenib (n = 162), %	Selective internal radiation therapy (n = 130), %	P value comparing groups
≥1 adverse event	84.6	60.0	p < 0.001
≥1 treatment-related adverse event	74.7	31.5	p < 0.001
≥1 grade ≥3 adverse event	50.6	27.7	p < 0.001
≥1 grade ≥3 treatment-related adverse event	37.7	13.1	p < 0.001
≥1 serious adverse event‡	35.2	20.8	p = 0.0091
≥1 treatment-related serious adverse event	9.3	4.6	p = 0.17

*As-treated population refers to patients who received the allocated intervention. It was not clear if there was a minimum duration for receiving sorafenib required for inclusion in treated population.
†We noted above when the text specified the outcome as "treatment-related". For outcomes not specified as "treatment-related", we assume that the outcome was "treatment emergent" (ie, occurred during treatment but was not attributed to the treatment).
‡Per the protocol (text that follows is largely verbatim, with very minimal revision for clarity as a footnote), defined as adverse events that led to: (1) death; (2) a serious deterioration in health that (2a) resulted in a life-threatening illness or injury, (2b) resulted in a permanent impairment of a body structure or a body function, (2c) required in-patient hospitalization or prolongation of existing hospitalization, or (2d) resulted in medical or surgical intervention to prevent permanent impairment to body structure or a body function; or (3) led to fetal distress or death or a congenital abnormality or birth defect

Adverse events* in the as-treated population†

System organ class, disorder‡	Radioembolization (n = 130)		Sorafenib (n = 162)		P value	
	grade 1 or 2, n (%)	grade ≥ 3, n (%)	grade 1 or 2, n (%)	grade ≥ 3, n (%)	grade 1 or 2	grade ≥ 3
Gastrointestinal disorders						
constipation	0	0	9 (5.6)	0	0.0051	-
diarrhea	2 (1.5)	0	42 (25.9)	6 (3.7)	< 0.001	0.035
General disorders and administration site conditions						
fatigue	5 (3.8)	0	19 (11.7)	6 (3.7)	0.0175	0.035
skin and subcutaneous tissue disorders						
alopecia	0	0	16 (9.9)	0	p < 0.001	-
palmar-plantar erythrodysesthesia syndrome	1 (0.8)	0	62 (38.3)	27 (16.7)	p < 0.001	p < 0.001
rash	0	0	18 (11.1)	0	p < 0.001	-
vascular disorders						
hypertension	0	0	22 (13.6)	2 (1.2)	p < 0.001	0.50
Adverse events typically associated						

with radioembolization						
gastric ulcer	0	1 (0.8)	0	0	-	0.45
upper gastrointestinal hemorrhage	1 (0.8)	1 (0.8)	0	3 (1.9)	0.45	0.63
jaundice	1 (0.8)	1 (0.8)	1 (0.6)	2 (1.2)	> 0.95	> 0.95
hepatic cirrhosis	0	0	1 (0.6)	1 (0.6)	> 0.95	> 0.95
portal hypertension	0	0	0	1 (0.6)	-	> 0.95
radiation hepatitis	0	2 (1.5)	0	0	-	0.20

*It was not clearly reported whether these are treatment-emergent adverse events or treatment-related adverse events; however, because the lower event rates seem to match more closely with treatment-related adverse events, we have assumed they are treatment-related.

†This table has been adapted from Chow 2018 Table 3, which reported adverse events in the as-treated population that were: experienced by >5% of treated patients in either group or known to be associated with selective internal radiation therapy and had an onset date on or after study treatment start date. Our adapted table restricts to the adverse events experienced by > 5% of treated patients that also were significantly different between groups (and still includes the adverse events typically associated with radioembolization). Adverse events typically associated with sorafenib include fatigue, diarrhea, and skin disorders; these are already captured in the table because they show significantly increased risk in sorafenib group.

‡Adverse events were graded with National Cancer Institute Common Terminology Criteria for Adverse Events, version 4.02.

- **Certainty:**
  - Moderate certainty for survival; Low certainty for progression-free survival
    - Downgrade for both outcomes: imprecision because wide CIs for differences between groups compatible with potentially-meaningful differences
      - although more patients randomized to SIRT than to sorafenib did not receive the assigned treatment (28.6% and 9%, respectively) due to some patients assigned to SIRT being found ineligible after work-up and assessment, we did not assign a risk of bias downgrade for this because we did not expect that these greater proportion of exclusions in SIRT group would explain why the analysis failed to demonstrate favorable effect of SIRT vs. sorafenib, particularly for survival outcomes in which the point estimates were superior in sorafenib group (though non-significantly) for median survival and survival at 12 and 18 months
        - additionally:
          - the as-treated analyses may offer at least some indirect support of the notion that the ITT analysis is not at substantial risk of bias due to differential receipt of assigned treatment
          - even if a meaningful difference really did exist between the options in this trial, the differential receipt of assigned treatment in the ITT analysis would arguably have contributed to the imprecision of the results; thus, downgrading for both imprecision and risk of bias would be unduly punitive
    - Downgrade for progression free survival: indirectness (surrogate measure)
  - Low certainty for quality of life
    - Downgrades*:*

        - imprecision downgrade described above also applies to quality of life outcome, though for quality of life, this is based on the estimates and 95% CIs for each respective group, as no estimates were provided for mean differences (or CIs thereof)
        - indirectness
            - EQ-5D is a health-related quality of life instrument not specific for liver disease, so it may not adequately capture the higher toxicity profile of sorafenib and the impact that has on liver-disease health-related quality of life
    - Moderate certainty for the following adverse event-related outcomes
        - ≥1 treatment-related adverse event
        - ≥1 grade ≥3 adverse event
        - ≥1 grade ≥3 treatment-related adverse event
        - ≥1 serious adverse event
        - downgrade for inability to assess consistency/replicability of findings given this certainty statement pertains to a single trial
            - this concern also applies to other outcomes from this trial, but the concern is not enough to warrant a separate downgrade
        - certainty not rated for out-of-scope adverse event outcomes such as ≥1 adverse event (which we also note are both too broad to be meaningful)
            - although technically outside of scope, we did rate certainty for grade ≥3 adverse events, because in the absence of more specific data, this might give insight into 'serious' adverse events
        - certainty not rated for specific/named adverse events, as these were scoped to inform the primary outcomes of interest (namely serious adverse events and treatment discontinuation due to adverse events)

FINN 2020

Finn RS, Qin S, Ikeda M, et al. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. N Engl J Med. 2020;382(20):1894-1905. PubMed

Relevant to FAQ2, FAQ3, FAQ4, FAQ5, FAQ6

- **Study design:** randomized trial
    - phase 3
    - randomization stratified by
        - geographic region (Asia excluding Japan vs. the rest of the world)
        - macrovascular invasion or extrahepatic spread of disease (presence vs absence)
        - baseline alpha-fetoprotein level (< 400 vs ≥ 400 ng per milliliter)
        - ECOG performance status (0 vs. 1)
- **Population:** patients with unresectable hepatocellular carcinoma
    - no previous systemic therapy
    - not amenable to curative or locoregional therapies or that had progressed thereafter
    - Child–Pugh liver function class A
    - ECOG performance status of 0 or 1
    - no coinfection with hepatitis B or C
- **Intervention:** atezolizumab 1200 mg plus bevacizumab 15 mg/kg IV every 3 weeks

- **Comparison:** sorafenib 400 mg orally twice a day

- **Study inclusion criteria:** no additional

- **Number of participants:** 501
  - atezolizumab plus bevacizumab, 336
  - sorafenib, 165
- **Duration of follow-up:** 8.6 months median
  - 8.9 months in the atezolizumab–bevacizumab group
  - 8.1 months in the sorafenib group

- **Results:**

Overall survival

	% (no. of events/no. of patients)	**Median overall survival, months (95% CI)**	**Overall survival at 6 months, % (95% CI)**
atezolizumab plus bevacizumab	28.6% (96/336)	NE	84.8% (80.9% to 88.7%)
sorafenib	39.4% (65/165)	13.2 (10.4 to NE)	72.2% (65.1% to 79.4%)

NE = could not be evaluated

stratified HR for death, 0.58 (95% CI 0.42 to 0.79)

- generally consistent effects across subgroups defined by baseline demographic and disease characteristics

Progression-free survival*

	% (no. of events/no. of patients)	**Median progression-free survival, months (95% CI)†**	**Progression-free survival at 6 months (%)**
atezolizumab plus bevacizumab	58.6% (197/336)	6.8 (5.7 to 8.3)	54.5%
sorafenib	66.1% (109/165)	4.3 (4.0 to 5.6)	37.2%

*Defined as the time from randomization to disease progression according to RECIST 1.1, as assessed at an independent review facility, or death from any cause, whichever occurred first

†Results refer to median progression-free survival (95% CI) as assessed at an independent review facility (from Fig 1B of publication); results for this outcome were also assessed by investigators (supplementary Table S4) and showed consistent results (atezolizumab plus bevacizumab 7.1 months [95% CI 5.7 to 8.4 months], sorafenib 2.9 months [2.8 to 4.2 months]).

stratified HR for progression or death, 0.59 (95% CI 0.47 to 0.76)

- generally consistent effects across subgroups defined by baseline demographic and disease characteristics

Quality of life

	Quality of life* – median time to deterioration (95% CI) months
atezolizumab plus bevacizumab	11.2 (6.0 to NE)
sorafenib	3.6 (3.0 to 7.0)

NE = could not be evaluated

*Evaluated with the use of the European Organization for Research and Treatment of Cancer (EORTC) quality-of-life questionnaire for cancer (EORTC QLQ–C30). This 30-item quality of life questionnaire assesses 5 function domains (physical, emotional, role, cognitive, and social), 8 symptoms (fatigue, nausea and vomiting, diarrhea, constipation, pain, dyspnea, insomnia, and anorexia), financial difficulties, and global health/quality of life. Higher scores are better for function and global health/quality of life; higher scores are worse for other elements. Time to deterioration was defined dichotomously as "the time from randomization to first deterioration (decrease from baseline of ≥10 points), maintained for at least two consecutive timepoints or one timepoint followed by death from any cause within 3 weeks in the following EORTC QLQ-C30 subscales: HRQoL/GHS [health-related quality of life/global health status], physical functioning, role functioning".

HR 0.63 (95% CI 0.46 to 0.85)

adverse events in the safety population (defined as patients who received ≥1 dose of the study drug)

Proportion of patients having treatment-emergent adverse events in the safety population

Outcome	**Atezolizumab plus bevacizumab (n = 329), n (%)**	**Sorafenib (n = 156), n (%)**
≥ 1 adverse event	323 (98.2)	154 (98.7)
≥ 1 grade 3 or 4 event	186 (56.5)	86 (55.1)
≥ 1 grade 5 event	15 (4.6)	9 (5.8)
≥ 1 serious adverse event	125 (38.0)	48 (30.8)
adverse event leading to withdrawal from any trial drug	51 (15.5)	16 (10.3)

Proportion of patients in the safety population with treatment-related adverse events*

Outcome	**Atezolizumab plus bevacizumab (n = 329) n (%)**		**Sorafenib (n = 156) n (%)**	
	Any grade	**Grade 3 or 4**	**Any grade**	**Grade 3 or 4**
Non-surrogate†				
Fatigue	50 (15.2)	5 (1.5)	24 (15.4)	5 (3.2)
Pruritus	43 (13.1)	0	13 (8.3)	0
Infusion-related reaction	36 (10.9)	7 (2.1)	0	0
Diarrhea	34 (10.3)	1 (0.3)	67 (42.9)	6 (3.8)
Decreased appetite	33 (10.0)	2 (0.6)	31 (19.9)	6 (3.8)
Rash	29 (8.8)	0	26 (16.7)	4 (2.6)
Nausea	21 (6.4)	0	20 (12.8)	0
Asthenia	11 (3.3)	0	16 (10.3)	3 (1.9)
Alopecia	3 (0.9)	0	21 (13.5)	0
Palmar-plantar erythrodysesthesia syndrome	2 (0.6)	0	75 (48.1)	13 (8.3)
Surrogate‡				

Hypertension	78 (23.7)	34 (10.3)	31 (19.9)	14 (9.0)
Proteinuria	62 (18.8)	9 (2.7)	7 (4.5)	1 (0.6)
Aspartate aminotransferase increase	46 (14.0)	14 (4.3)	11 (7.1)	4 (2.6)
Alanine aminotransferase increase	34 (10.3)	7 (2.1)	4 (2.6)	0
Platelet count decrease	27 (8.2)	8 (2.4)	15 (9.6)	1 (0.6)
Blood bilirubin increase	27 (8.2)	2 (0.6)	9 (5.8)	4 (2.6)
Hypophosphatemia	3 (0.9)	1 (0.3)	7 (4.5)	5 (3.2)

*This table is based on Table S6. Adverse events within each major grouping (non-surrogate or surrogate) are listed in descending order of frequency of occurrence of any grade within the atezolizumab plus bevacizumab group. Given the ancillary role that Finn 2020 serves in this evidence report (i.e., per Scoping, not being incorporated into Evidence Synthesis other than noting it may impact the standard of care for systemic therapy), we have listed all the adverse events in Table S6 rather than only those meeting criteria specified in Scoping.
†Defined by EBSCO Information Services as those outcomes which are directly patient-relevant (i.e., causing hardship, symptoms, etc. *per se* rather than being an intermediary that might lead to hardship, symptoms, etc.)
‡Defined by EBSCO Information Services as those outcomes which are not directly patient-relevant *per se*; of note, some may debate our inclusion of hypertension within this category, but we emphasize hypertension by itself is not patient relevant (e.g., patients only care about hypertension because of what it potentially signifies/portends).

- **Certainty:**
  - efficacy outcomes
    - overall survival (months median)
      - Moderate certainty
        - downgrade due to the certainty statement above pertaining to a single trial, thus precluding assessment of consistency/replicability of findings
    - progression-free survival (months median)
      - Very low certainty
        - downgrade due to indirectness because surrogate outcome measure
        - downgrade due to the certainty statement above pertaining to a single trial, thus precluding assessment of consistency/replicability of findings
        - downgrade due to concerns about imprecision (e.g., note difference in medians, and associated with a HR that would be consistent with smaller differences)
    - quality of life
      - Low certainty
        - downgrade due to the certainty statement above pertaining to a single trial, thus precluding assessment of consistency/replicability of findings
        - downgrade due to potential indirectness given the way in which the outcome was assessed (time to deterioration, with deterioration defined in a dichotomous manner rather than simply measuring quality of life in its native form)
  - adverse events
    - Low certainty
      - ≥1 serious adverse event
      - adverse event leading to withdrawal from any trial drug
      - ≥1 grade 3 or 4 adverse event

        - ≥1 grade 5 adverse event
        - downgrades
            - certainty statement above pertaining to a single trial, thus precluding assessment of consistency/replicability of findings
            - authors provide no comparative statistics, including not providing any measurement of precision
            - indirectness of the above-noted outcomes given it is not clear whether the adverse events were attributable to the intervention
    - certainty not assessed for non-scoped adverse events such as ≥1 adverse event (which we also note is too broad to be meaningful)
        - although technically outside of scope, we did rate certainty for grade 3 or 4 adverse events and grade 5 adverse events, because in the absence of more specific data, this might give insight into 'serious' adverse events
    - certainty not rated for specific/named adverse events, as these were scoped to inform the primary outcomes of interest (namely serious adverse events and treatment discontinuation due to adverse events); moreover, Finn 2020 serves as ancillary role in this evidence report (see above and Scoping)

- **Other considerations:**
    - Per Scoping, we include Finn 2020 given its potential to impact the standard of care for systemic therapy in this population. However, as noted in Scoping, we do not progress quantitative outcomes to Evidence Synthesis or attempt to make any indirect comparisons in Evidence Synthesis based on any of the above data. Per Scoping, we only note it has "potential to impact the standard of care for systemic therapy"

LLOVET 2008
Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med. 2008;359(4):378-390. PubMed

Relevant to FAQ2, FAQ3, FAQ5, FAQ6

- **Study design:** randomized trial
    - phase 3
    - setting was 121 centers in 21 countries in Europe, North America, South America, and Australasia
    - randomization stratified by
        - region
        - ECOG performance status (score of 0 vs. 1 or 2)
        - presence or absence of macroscopic vascular invasion (portal vein or branches) or extrahepatic spread
- **Population:** patients with advanced-stage hepatocellular carcinoma; advanced-stage disease defined as not eligible for or had disease progression after surgical or locoregional therapies
    - no previous systemic therapy
    - Child–Pugh liver function class A

  - ECOG performance status of 0, 1, or 2
  - life expectancy of ≥ 12 weeks

- **Intervention:** sorafenib
  - 800 mg/day
    - 400 mg twice daily on a continuous basis
- **Comparison:** placebo

- **Study inclusion criteria:** no additional

- **Number of participants:** 602
  - sorafenib, 299
  - placebo, 303
- **Duration of follow-up:**
  - enrollment began March 2005 and study was stopped February 2007
  - median duration of treatment
    - 5.3 months (range 0.2 to 16.1) in sorafenib group
    - 4.3 months (range 0.1 to 16.6) in placebo group
- **Results:**

overall survival (months median, 95% CI) and HR (95% CI) in ITT* analysis:

- 10.7 (9.4 to 13.3) in sorafenib group
- 7.9 (6.8 to 9.1) in placebo group
- HR 0.69 (0.55 to 0.87)
- consistent effects across most subgroups defined by randomization stratification variables
- consistent effect in analysis adjusting for prognostic factors (HR 0.73; 0.58 to 0.92)

*ITT population included all randomized patients

overall survival (%) at 12 months in ITT analysis:

- 44% in sorafenib group
- 33% in placebo group
- "31% relative reduction in risk of death"; p value = 0.009

time to symptomatic progression* (months median, 95% CI) and HR (95% CI) in ITT analysis:**

- 4.1 (3.5 to 4.8) in sorafenib group
- 4.9 (4.2 to 6.3) in placebo group
- HR 1.08 (0.88 to 1.31)

*Defined as either decrease of ≥ 4 points from baseline score on Functional Assessment of Cancer Therapy–Hepatobiliary Symptom Index 8 questionnaire or an ECOG status of 4 or death, whichever occurred first
**Reported only as median (not at specific time points)

time to radiologic progression* (months median, 95% CI) and HR (95% CI) in ITT analysis:

- 5.5 (4.1 to 6.9) in sorafenib group
- 2.8 (2.7 to 3.9) in placebo group
- HR 0.58 (0.45 to 0.74)

*As defined by RECIST, on basis of an independent review of radiologic data

proportion with progression-free survival* at 4 months in ITT analysis:

- 62% in sorafenib group
- 42% in placebo group
- no comparative statistics reported

*On basis of an independent review of radiologic data

adverse events

Proportion of patients having adverse events in the safety* population

outcome	sorafenib % (n = 297)	placebo % (n = 302)	Additional notes
≥ 1 treatment-related adverse event	80%	52%	Authors do not provide any comparative statistics
≥ 1 serious adverse event†	52%	54%	Authors note the rates were "similar"
discontinuations due to adverse events‡	29%	30%	Authors note the rates were "similar"
discontinuations due to adverse events considered to be drug-related§	11%	5%	Authors do not provide any comparative statistics

*Safety population refers to patients who received ≥ 1 dose of a study drug.

†Table C2 shows no clear difference between groups for any specific serious adverse event.

‡The authors report "The rate of discontinuation of the study drug due to adverse events was similar in the two study groups (38% vs. 37%)." However, as can be ascertained from Figure 1, the reported proportions are based on a denominator of those who discontinued treatment. If instead using the number of participants in the safety population and the reported number who discontinued therapy due to adverse events in Figure 1 (86/297 and 90/302, respectively), one arrives at the estimates shown in the table (rounded to nearest whole number for consistency with the other estimates in the table).

§Percentages are again presented as whole numbers in line with the way they are reported in the publication. These figures are based on proportions of 34/297 and 15/302, respectively.

Drug-related specific adverse events* in safety† population

Outcome	Sorafenib group (n = 297), % (n)		Placebo group (n = 302), % (n)		P values	
	any grade	grade 3/4	any grade	grade 3/4	any grade	grade 3 or 4
Constitutional symptoms						
Fatigue	7.4% (22)	1.3% (4)	5.3% (16)	<1.3% (<4)	0.07	1.00
Weight loss	3.0% (9)	0.7% (2)	0.3% (1)	0% (0)	<0.001	0.03
Dermatologic events						
Alopecia	4.7% (14)	0% (0)	0.7% (2)	0% (0)	<0.001	NA
Dry skin	2.7% (8)	0% (0)	1.3% (4)	0% (0)	0.04	NA
Hand–foot skin reaction	7.1% (21)	2.7% (8)	1.0% (3)	< 0.3% (<1)	<0.001	<0.001
Rash or desquamation	5.4% (16)	0.3% (1)	3.6% (11)	0% (0)	0.12	0.12
Gastrointestinal events						
Anorexia	4.7% (14)	<0.3% (<1)	1.0% (3)	0.3% (1)	<0.001	1.00
Diarrhea	13.1% (39)	2.7% (8)	3.6% (11)	0.7% (2)	<0.001	<0.001
Voice changes	2.0% (6)	0% (0)	0.3% (1)	0% (0)	<0.001	NA
Hypertension	1.7% (5)	0.7% (2)	0.7% (2)	0.3% (1)	0.05	0.28

Abdominal pain not otherwise specified	2.7% (8)	0.7% (2)	1.0% (3)	0.3% (1)	0.007	0.17

*Authors provided data for drug-related specific adverse events (as defined by the National Cancer Institute Common Terminology Criteria [version 3.0]) that occurred in ≥ 5% of patients in either study group. We extracted outcomes that showed a significant between-group difference; we also extracted rash or desquamation given it was significantly different in other trials

†Safety population refers to patients who received ≥ 1 dose of a study drug

Certainty:

- Efficacy outcomes
  - overall survival (months) median and overall survival (%) at 12 months
    - Moderate certainty
      - Downgrade due to potential imprecision in the results (e.g. median survival of 10.7 months vs. 7.9 months with a HR of 0.69, 95% CI 0.55 to 0.87) and the inability to assess consistency/replicability of findings given this certainty statement pertains to a single trial
  - time to symptomatic progression (months) median
    - Moderate certainty
      - Downgrade due to imprecision because wide CIs for differences between groups compatible with either treatment being better
        - concern regarding this being a single trial (see overall survival outcome) also applies here, but was not enough to warrant a separate downgrade
  - time to radiologic progression
    - Moderate certainty
      - Downgrade due to indirectness due to surrogate outcome measure
        - concern regarding this being a single trial (see overall survival outcome) also applies here, but was not enough to warrant a separate downgrade
- adverse events
  - ≥1 serious adverse event
    - Low certainty, with downgrades for
      - imprecision
      - indirectness, because this outcome does not capture whether the serious adverse event is attributable to the intervention
  - discontinuations due to adverse events
    - Low certainty, with downgrades for
      - imprecision
      - indirectness because this outcome does not capture whether the adverse event triggering discontinuation is attributable to the intervention
  - discontinuations due to adverse events considered to be drug-related
    - Moderate certainty, with downgrade for imprecision
  - concern regarding this being a single trial (see overall survival outcome) also applies to these outcomes, but not enough to warrant a separate downgrade
  - certainty not rated for the outcome of ≥ 1 treatment-related adverse event or specific treatment-related adverse events

- the former is outside scope and too broad to be meaningful
- the latter were scoped to inform the primary outcomes of interest (namely serious adverse events and treatment discontinuation due to adverse events)

VILGRAIN 2017

Vilgrain V, Pereira H, Assenat E, et al; SARAH Trial Group. Efficacy and safety of selective internal radiotherapy with yttrium-90 resin microspheres compared with sorafenib in locally advanced and inoperable hepatocellular carcinoma (SARAH): an open-label randomised controlled phase 3 trial. Lancet Oncol. 2017 Dec;18(12):1624-1636. PubMed

Relevant to FAQ2, FAQ3, FAQ4, FAQ5

- **Study design:** randomized trial
  - phase 3
  - randomization stratified by
    - center
    - ECOG performance status (0 vs. 1)
    - previous transarterial chemoembolization (yes vs. no)
    - presence of macroscopic vascular invasion (yes vs. no)
  - setting was 25 centers specializing in liver diseases in France
- **Population:** adults with locally advanced HCC (BCLC stage C); new HCC not eligible for surgical resection, liver transplantation, or thermal ablation after a previously cured HCC; *or* HCC with 2 unsuccessful rounds of transarterial chemoembolization
  - BCLC stage A, B, or C
    - proportion with BCLC A
      - 4% in selective internal radiation therapy group
      - 5% in sorafenib group
    - proportion with BCLC C
      - 68.4% in selective internal radiation therapy group
      - 67.1% in sorafenib group
  - no extrahepatic metastasis
  - ECOG performance status of 0 or 1
  - Child-Pugh liver function class A or B (score of $\leq$ 7)
  - life expectancy > 3 months
- **Intervention:** selective internal radiation therapy with yttrium-90 (^{90}Y) resin microspheres
  - allowed sequential or repeated delivery
    - 37% of patients had > 1 delivery
- **Comparison:** sorafenib
  - 800 mg/day
    - 400 mg twice daily on a continuous basis

- **Study inclusion criteria:** no additional
- **Number of participants:** 467
  - selective internal radiation therapy, 237
  - sorafenib, 222

- **Duration of follow-up:** 28 months median
  - 27.9 months in selective internal radiation therapy group
  - 28.1 months in sorafenib group

- **Results:**

overall survival (months median, 95% CI) and HR (95% CI) in ITT* analysis:

- 8.0 (6.7 to 9.9) in selective internal radiation therapy group
- 9.9 (8.7 to 11.4) in sorafenib group
- HR 1.15 (95% CI 0.94 to 1.41)
- no convincing evidence of different effects for overall survival according to prespecified and *post hoc* subgroup analyses

*In the per-protocol population, median overall survival was 9·9 months in both groups.

overall survival (%) at 12 months (95% CI) in ITT* analysis:

- 39·5% (33.3% to 45.9%) in selective internal radiation therapy group
- 42.1% (35.6% to 48.7%) in sorafenib group

progression-free survival* (months median, 95% CI) and HR (95% CI) in ITT analysis:**

- 4.1 (3.8 to 4.6) in selective internal radiation therapy group
- 3.7 (3.3 to 5.4) in sorafenib group
- HR 1.03 (0.85 to 1.25)

*Progression-free survival defined as the time from the closest date of radiological examination before first administration of the study treatment to disease progression, according to RECIST version 1.1 criteria, or death.

quality of life

- instruments used to evaluate cancer-specific quality of life were the European Organisation for Research and Treatment of Cancer (EORTC) quality of life questionnaire (QLQ-C30; version 3) and the specific hepatocellular carcinoma module (QLQ-HCC18)
  - only the "global health status subscore" of EORTC QLQ-C30 was reported and it was significantly better in the selective internal radiation therapy group than in the sorafenib group (group effect p = 0.0048; time effect p < 0.0001) and the between-group difference increased with time (group-time interaction p = 0.0447)
  - authors noted in Methods that "...QLQ-HCC18 (and the remaining sub-scores of QLQ-C30) will be reported elsewhere...", but there is no indication of any follow-up publications for Vilgrain 2017 in the Abdel-Rahman 2020 Cochrane review (see subsection "References to studies included in this review" in their "References" section)

Quality of life as assessed by the global health status subscore of the EORTC QLQ-C30*

Timing of assessment	Selective internal radiation therapy		Sorafenib	
	# completed questionnaires (# at risk)	mean (SD)	# completed questionnaires (# at risk)	mean (SD)
Baseline	169 (237)	63.2 (±21.4)	186 (222)	64 (±22.8)

3 months	105 (196)	61.3 (±23.9)	118 (191)	55.5 (±21.3)
6 months	69 (143)	63.6 (±23.5)	85 (153)	53 (±19.5)
9 months	41 (113)	65.3 (±25.6)	46 (124)	57.2 (±26.5)
12 months	26 (90)	64.1 (±24.2)	29 (92)	55.6 (±18.6)

*As estimated from Figure 4 in the intention-to-treat population. SDs are estimated by taking the average of |upper error bar – point estimate| and |lower error bar – point estimate|.

adverse events

The authors defined their per-protocol population as patients who got the treatment to which they were assigned (and had no major protocol deviations). They defined their safety population differently for the SIRT and sorafenib groups. Put simply, to be included in the safety population for the SIRT group, all that was required was for a patient to undergo at least one work-up examinations for SIRT, not actually receiving SIRT. This resulted in 42 patients being included in the SIRT safety group even though they never received SIRT (26 received sorafenib, 16 did not receive any treatment). To be included in the safety group for sorafenib, all that was required was receiving at least 1 dose of sorafenib. Unfortunately, some of the estimates reported by Vilgrain 2017 are only reported for the safety group.

- ≥ 1 treatment-related adverse event in safety population
  - 173 (77%) of 226 (including 92 [41%] grade ≥ 3 adverse events) in the SIRT group
  - 203 (94%) of 216 (including 136 [63%] grade ≥ 3 adverse events) in the sorafenib group
- discontinuations for drug-related toxicity in sorafenib group
  - 108 of 216 (50%) permanently discontinued (139 [64%] discontinued non-permanently)
- ≥ 1 serious adverse event in safety population
  - 174 (77%) of 226 patients in the SIRT group
  - 176 (81%) of 216 patients in sorafenib group
- ≥ 1 treatment-related serious adverse event in safety population
  - 45 (20%) of 226 patients in the SIRT group (36 [16%] related to SIRT and 9 [4%] related to sorafenib)
  - 56 (26%) of 216 patients in the sorafenib group
- "Serious" for the above outcomes was defined as "Any adverse event or reaction that results in the death, or endangers the life, of an individual participating in a research study, requires hospitalisation or prolonged hospitalisation, causes a significant or long-term disability or incapacity, or results in a congenital anomaly or malformation"

Treatment-related adverse events of interest in the per-protocol population*†

System organ class, disorder	**Radioembolization (n = 174)**		**Sorafenib (n = 206)**	
	grade 1 or 2, % (n)	**grade ≥ 3, % (n)**	**grade 1 or 2, % (n)**	**grade ≥ 3, % (n)**
Gastrointestinal disorders				
diarrhea	9.2% (16)	0.6% (1)	63.6% (131)	14.6% (30)
General disorders and administration site conditions				
fatigue	41.4% (72)	8.6% (15)	58.3% (120)	18.0% (37)

skin and subcutaneous tissue disorders				
alopecia	0	0	16.5% (34)	0
palmar-plantar erythrodysesthesia syndrome	0	0	17.5% (36)	5.8% (12)
rash or desquamation	1.1% (2)	0	9.7% (20)	0
vascular disorders				
hypertension	2.9% (5)	0	13.1% (27)	2.4% (5)
Adverse events typically associated with radioembolization				
gastric ulcer	1.1% (2)	1.7% (3)	0	0.5% (1)
upper gastrointestinal hemorrhage	0	2.3% (4)	2.9% (6)	3.4% (7)
radiation hepatitis	0	0	0	0

*According to Medical Dictionary for Regulatory Activities coding, version 19.0

†This table has been adapted from Vilgrain 2017 Table S9, which reported treatment-related adverse events of interest in their per-protocol population. The authors report a similar table in their main report (Table 4) that was specific to what they refer to as their safety population. According to Figure 1, the authors defined their per-protocol population as patients who got the treatment to which they were assigned (and had no major protocol deviations) and their safety population as also including patients who did not get the treatment to which they were assigned. For example, in the SIRT group, the safety population included 42 patients who did not receive SIRT. Because it makes more sense to use an operational definition of treatment-related adverse events that minimizes inclusion of patients who did not receive the assigned treatment, we have extracted data from the per-protocol population (Table S9) instead of the safety population (Table 4) for individual treatment-related adverse events. However, although Table S9 reported 29 individual treatment-related adverse events, it did not report between-group comparative statistics, thus introducing inferential impediments. Given the data available in Chow 2018, we considered the adverse event data from Vilgrain 2017 to be most useful as a comparison against the results reported in Chow 2018 (of note, Chow 2018 defines its safety population in a manner similar to what Vilgrain 2017 uses to define its per-protocol population). Given these considerations, we selected for extraction among the 29 individual treatment-related adverse events only those that met our criteria for extraction from Chow 2018. Adverse events meeting these criteria and reported in Chow 2018 that were not mentioned in Vilgrain 2017 Table S9 are the following: constipation and 3 adverse events listed under "AEs typically associated with radioembolization" (namely, jaundice, hepatic cirrhosis, portal hypertension). Finally, in adapting Vilgrain 2017 Table S9 for our summary table, we also collapsed their grade categories 3, 4, and 5 into a single category of grade >3 to be more consistent with Chow 2018, and we used consistent disorder naming for our Vilgrain 2017 summary table as used in our Chow 2018 table for easy comparisons between the two (e.g.,, "skin and subcutaneous tissue disorders" in Chow 2018 vs. "Dermatologic events", which was used in Vilgrain 2017 Table S9).

- **Certainty:**
  - Moderate certainty for survival; Low certainty for progression-free survival
    - Downgrade for both: imprecision because wide CIs for differences between groups compatible with potentially-meaningful differences
      - although more patients randomized to SIRT than to sorafenib did not receive the assigned treatment (Figure 1) due to some patients assigned to SIRT being found ineligible after work-up and assessment, we did not assign a risk of bias downgrade for this because we would not expect that these greater proportion of exclusions in SIRT group would explain why the analysis failed to demonstrate favorable effect of SIRT vs. sorafenib, particularly for survival outcomes in which the point

estimates were superior in sorafenib group (though non-significantly) for median survival and survival at 12 months
- additionally:
  - the per-protocol analysis for overall survival (not presented for progression-free survival) may be seen as indirectly suggesting the ITT analysis is not at substantial risk of bias due to the differential receipt of assigned treatment
  - even if a meaningful difference really did exist between the options in this trial, differential receipt of assigned treatment in the ITT analysis would arguably have contributed to the imprecision of the results; thus, downgrading for both imprecision and for risk of bias would be unduly punitive

- Downgrade for progression-free survival: indirectness (surrogate measure)

- Very low certainty for quality of life
  - Downgrades*:*
    - extremely serious concern for risk of bias due to selective outcome reporting (very serious concern) and incomplete outcome data (serious concern)
      - authors measured disease-specific quality of life using multiple measures but only reported the results of one and noted that the others "...will be reported elsewhere..." (see section above 'Results, quality of life' for more detail)
      - as seen in the above table, there is considerable concern for increasing missingness over time for quality of life data
    - serious concern for imprecision
      - the differences demonstrated are statistically significant, but we note:
        - although we did not calculate differences in means in the above table, the point estimate for the difference in means is – at its largest – 10.6 (at 6 months)
        - the corresponding mean values in each group are accompanied by large SDs at all time points
        - there is substantial loss of people contributing to the estimates over time (thus a smaller N)
        - the above would be expected to lead to considerable imprecision in the estimates (though CIs were not explicitly calculated for these outcomes)
          - though one might suspect this could be due in part to the loss to follow-up noted above, we also note the large SDs were apparent at every time point (so the loss to follow-up is not considered a sufficient explanation for the large SDs), and moreover, loss to follow-up contributing to imprecision is a separate issue from the potential risk of bias included in the

data for those providing follow-up data at that time point (e.g., response bias)

- adverse events
  - Certainty not evaluated for general or specific adverse events
    - For general adverse event outcomes, we extracted relevant information from Results text; for specific adverse event outcomes, we extracted the data from Supplemental Table S9. Results of between-group comparisons were not reported for either general or specific adverse events outcomes, and for general adverse events outcomes, the population in whom these estimates were made is problematic, as discussed above. Given the degree of detail in Chow 2018, we use the results from Vilgrain 2017 primarily for cross-checking for consistency and do not assign study-level certainty ratings to adverse event outcomes from Vilgrain 2017

YOON 2018

Yoon SM, Ryoo BY, Lee SJ, Kim JH, Shin JH, An JH, Lee HC, Lim YS. Efficacy and Safety of Transarterial Chemoembolization Plus External Beam Radiotherapy vs Sorafenib in Hepatocellular Carcinoma With Macroscopic Vascular Invasion: A Randomized Clinical Trial. JAMA Oncol. 2018 May 1;4(5):661-669. PubMed

Relevant to FAQ2, FAQ3, FAQ5

- **Study design:** randomized trial
  - setting was academic tertiary care center in Seoul Korea
  - treatment crossover was allowed after confirming disease progression during the initially assigned treatment
- **Population:** patients with liver-confined hepatocellular carcinoma showing macroscopic vascular invasion
  - could not have had prior treatment
  - all patients had portal vein invasion of HCC
  - no extrahepatic disease
  - ECOG performance status of 0 or 1
  - Child-Pugh liver function class A
- **Intervention:** combined treatment with
  - transarterial chemoembolization (TACE)
    - every 6 weeks for first 4 treatments, then every 6 to 8 weeks thereafter
  - external beam radiotherapy (EBRT)
    - within 3 weeks after the first TACE
    - maximum 45 Gy with fraction size of 2.5 to 3 Gy
- **Comparison:** sorafenib
  - 800 mg/day
    - 400 mg twice daily on a continuous basis

- **Study inclusion criteria:** no additional

- **Number of participants:** 90
- **Duration of follow-up:** 140 weeks maximum

- **Results:**

Median overall survival and progression-free survival*

	Median time (95% CI), weeks			
Outcome	**Sorafenib group (n = 45)**	**TACE+EBRT group (n = 45)**	**Hazard ratio (95% CI)**	**P value**
progression-free survival	11.3 (4.4 to 18.2)	30.0 (22.9 to 37.1)	0.27 (0.17 to 0.44)	<0.001
overall survival†	43.0 (25.9 to 60.1)	55.0 (28.9 to 81.2)	0.61 (0.38 to 0.98)	0.04

*Subgroup analysis not reported.
†Progression was defined as progressive disease by independent radiologic review according to RECIST, version 1.1, criteria or death from any cause. At week 12, the progression-free survival rate was also significantly higher in the TACE+EBRT group than in the sorafenib group (86.7% vs 34.3%; p < 0.001).

Treatment-emergent adverse events in the safety population*

Adverse event category†	**Sorafenib (n = 44), n (%)**	**TACE+EBRT (n = 45), n (%)**	**P value**
≥ 1 adverse event	41 (93.2%)	41 (91.1%)	> 0.99
≥ 1 grade ≥ 3 adverse event	12 (27.3%)	7 (15.6%)	0.18
≥ 1 serious adverse event	5 (11.4%)	5 (11.1%)	0.97
discontinuation due to adverse event	1	0	0.49

TACE+EBRT, transarterial chemoembolization plus external beam radiotherapy
*By modified intention-to-treat analysis, excepting 1 patient in sorafenib group who withdrew consent before treatment
†Patients are counted once for each category

Adverse events* in the safety population†

system organ class, disorder	**Sorafenib (n = 44)**		**TACE+EBRT (n = 45)**	
	grade 1 or 2, n (%)	**grade ≥ 3, n (%)**	**grade 1 or 2, n (%)**	**grade ≥ 3, n (%)**
Gastrointestinal disorders				
constipation	0	0	1 (2.2%)	0
diarrhea	grade 1: 12 (27.3%); grade 2: 4 (9.1%)	2 (4.5%)	grade 1: 2 (4.4%); grade 2: 1 (2.2%)	1 (2.2%)
General disorders and administration site conditions				
fatigue	5 (11.4%)	0	grade 1: 6 (13.3%) grade 2: 3 (6.7%)	0
skin and subcutaneous tissue disorders				
palmar-plantar erythrodysesthesia syndrome	grade 1: 14 (31.8%) grade 2: 9 (20.5%)	2 (4.5%)	0	0
rash	grade 1: 4 (9.1%) grade 2: 4 (9.1%)	0	0	0
vascular disorders				

hypertension	grade 1: 3 (6.8%) grade 2: 7 (15.9%)	4 (9.1%)	0	0

*Yoon 2018 appear to focus on treatment-emergent rather than treatment-related adverse events; for instance, their eTable 6 (on which our table 'Treatment-emergent adverse events in the safety population' is based) explicitly focuses on treatment-emergent adverse events. However, the adverse events presented in the table immediately above are listed by Yoon 2018 (in eTable 7) as "Adverse Event Categories in the Safety Population", thus leaving some uncertainty as to whether they are treatment-emergent or treatment-related. Nevertheless, the context within which they are presented, the comparison of the full data tables (eTables 6 and 7 in Yoon 2018), and consideration of their trial protocol lead us to assume these are in fact treatment-emergent adverse events, not treatment-related adverse events. With respect to the trial protocol, section 9.1 "Definition of Adverse Events" says "All adverse events will be assessed and recorded on the AE CRF page by the investigator. An adverse event (AE) is any untoward medical occurrence in a study patient, regardless of the potential relation with the use of a study drugs." Although a later section (9.6) discusses how causal attribution language should be used for any reported adverse event, the fact such language is not seen anywhere in the publication further strengthens our impression that their data represent treatment-emergent adverse events rather than treatment-related adverse events.

†This table has been adapted from eTable 7 of Yoon 2018, which reports 15 individual treatment-related adverse events but no between-group comparative statistic. This introduces inferential impediments, and we also note Yoon 2018 has a considerably smaller sample size than either Chow 2018 or Vilgrain 2017. We consider Yoon 2018 most useful as a comparison against the results reported in Chow 2018. Chow 2018 also report data for a population defined similarly to the safety population in Yoon 2018, which appeared to be population of patients who received the treatment to which they were assigned and before any discontinuations or crossovers (Figure 1). Given the limited usefulness of Yoon 2018's data and its relevance mainly as a comparison against Chow 2018, we selected for extraction among the 15 individual treatment-related adverse events only those that met our criteria for extraction from Chow 2018. Adverse events meeting these criteria and reported in Chow 2018 that were not mentioned in Yoon 2018 eTable 7 are the following: alopecia or any of the 6 adverse events "typically associated with radioembolization". Finally, in adapting Yoon 2018 eTable 7 for our summary, we also collapsed their grade categories 1, 2, 3, and 4 into two categories of grade 1 or 2 and grade >3 to be more consistent with Chow 2018, and we grouped the individual 15 disorders listed in Yoon 2018 eTable 7 (without broader categories) into the broader categories used in Chow 2018 to facilitate comparisons between the two.

- **Certainty:**
  - Very low certainty for survival and progression-free survival
    - downgrades for both outcomes
      - indirectness (single-center trial and required macroscopic vascular invasion for inclusion in trial, thus constituting a narrower inclusion than specified in Scoping and possibly leading to a group that is sicker on average)
      - imprecision (small trial)
      - inability to assess consistency of findings (single-center, small trial that has – as of yet – not been replicated)
    - survival
      - no downgrades beyond those noted above
    - progression-free survival
      - additional downgrade for indirectness (surrogate outcome)
  - adverse events
    - certainty not assessed for general or specific adverse events given the main purpose of Yoon 2018's adverse event data are for comparison against the larger and more detailed Chow 2018; Yoon 2018 also has limited power relative to Chow 2018

Evidence report for local ablation vs. systemic therapy in patients with hepatocellular carcinoma and cirrhosis who are ineligible for resection or transplantation

Reference list

EBSCO Information Services

July 17, 2020

Prepared by EBSCO Health Innovations and EBM Development Department (in alphabetic order):

Brian S. Alper, MD, MSPH, FAAFP, FAMIA, Chief Medical Knowledge Officer

Eric Manheimer, PhD, EBM Editor

Martin Mayer, DMSc, MS, PA-C, Senior Clinical/Medical Editor

ABDEL-RAHMAN 2020
Abdel-Rahman O, Elsayed Z. Yttrium-90 microsphere radioembolisation for unresectable hepatocellular carcinoma. Cochrane Database Syst Rev. 2020;1(1):CD011313. PubMed

BI 2020
Bi F, Qin S, Gu S, et al. Donafenib versus sorafenib as first-line therapy in advanced hepatocellular carcinoma: An open-label, randomized, multicenter phase II/III trial. JCO. 2020 May 20;38(15_suppl):4506–4506. doi: 10.1200/JCO.2020.38.15_suppl.4506

CANCER RESEARCH UK 2019
Cancer Research UK. Selective internal radiation therapy (SIRT). Last reviewed/updated February 18, 2019. Accessed July 16, 2020. Available at https://www.cancerresearchuk.org/about-cancer/cancer-in-general/treatment/radiotherapy/internal/radioactive-implant-treatment/selective-internal-radiation-therapy-sirt.

CEDARS-SINAI ND
Cedars-Sinai. Selective Internal Radiation Therapy for Liver. Date last reviewed/updated not provided. Accessed July 16, 2020. Available at https://www.cedars-sinai.edu/Patients/Programs-and-Services/Imaging-Center/For-Patients/Exams-by-Procedure/Interventional-Radiology/Selective-Internal-Radiation-Therapy-for-Liver.aspx.

CHENG 2009
Cheng AL, Kang YK, Chen Z, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. Lancet Oncol. 2009;10(1):25-34. PubMed

CHOW 2018
Chow PKH, Gandhi M, Tan SB, et al. Asia-Pacific Hepatocellular Carcinoma Trials Group. SIRveNIB: Selective Internal Radiation Therapy Versus Sorafenib in Asia-Pacific Patients With Hepatocellular Carcinoma. J Clin Oncol. 2018 Jul 1;36(19):1913-1921. PubMed

CLEVELAND CLINIC 2019
Cleveland Clinic. Liver Cancer: Management and Treatment. Last reviewed/updated October 21, 2019. Accessed July 16, 2020. Available at https://my.clevelandclinic.org/health/diseases/9418-liver-cancer/management-and-treatment.

DYNAMED 2020
DynaMed [Internet]. Ipswich (MA): EBSCO Information Services. 1995 - . Sorafenib; [updated 2020 June 1, cited 2020 July 16]. Available from https://www.dynamed.com/drug-monograph/sorafenib.

FINN 2018
Finn RS, Zhu AX, Farah W, et al. Therapies for advanced stage hepatocellular carcinoma with macrovascular invasion or metastatic disease: A systematic review and meta-analysis. Hepatology. 2018;67(1):422-435. doi:10.1002/hep.29486. PubMed

FINN 2020
Finn RS, Qin S, Ikeda M, et al. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. N Engl J Med. 2020;382(20):1894-1905. PubMed

IKEDA 2020
Ikeda M, Inaba Y, Tanaka T, et al. A prospective randomized controlled trial of selective transarterial chemoembolization using drug-eluting beads loaded with epirubicin versus selective conventional transarterial chemoembolization using epirubicin-lipiodol for hepatocellular carcinoma: The JIVROSG-1302 PRESIDENT study. JCO. 2020 May 20;38(15_suppl):4518–4518. doi: 10.1200/JCO.2020.38.15_suppl.4518

KATSANOS 2017
Katsanos K, Kitrou P, Spiliopoulos S, et al. Comparative effectiveness of different transarterial embolization therapies alone or in combination with local ablative or adjuvant systemic treatments for unresectable hepatocellular carcinoma: A network meta-analysis of randomized controlled trials. PLoS One. 2017 Sep 21;12(9):e0184597. doi: 10.1371/journal.pone.0184597. PubMed

KUDO 2018
Kudo M, Finn RS, Qin S, et al. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. Lancet. 2018 Mar 24;391(10126):1163-1173. doi: 10.1016/S0140-6736(18)30207-1. PubMed

LLOVET 2008
Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med. 2008;359(4):378-390. PubMed

MAYO CLINIC 2019
Mayo Clinic. Liver cancer: Diagnosis & treatment. Last reviewed/updated May 4, 2019. Accessed July 16, 2020. Available at https://www.mayoclinic.org/diseases-conditions/liver-cancer/diagnosis-treatment/drc-20353664.

MEMORIAL SLOAN KETTERING CANCER CENTER 2019
Memorial Sloan Kettering Cancer Center. Selective Internal Radiation Therapy (SIRT) Treatment. Last reviewed/updated March 28, 2019. Accessed July 16, 2020. Available at https://www.mskcc.org/cancer-care/patient-education/selective-internal-radiation-treatment-sirt.

OXMAN 1992
Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med. 1992 Jan 1;116(1):78-84. doi: 10.7326/0003-4819-116-1-78. PubMed

ROCCARINA 2017
Roccarina D, Majumdar A, Thorburn D, et al. Management of people with intermediate-stage hepatocellular carcinoma: an attempted network meta-analysis. Cochrane Database Syst Rev. 2017;3(3):CD011649. Published 2017 Mar 10. doi:10.1002/14651858.CD011649.pub2. PubMed

SUN 2012
Sun X, Briel M, Busse JW, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ. 2012 Mar 15;344:e1553. doi: 10.1136/bmj.e1553. PubMed

SUN 2014
Sun X, Ioannidis JP, Agoritsas T, et al. How to use a subgroup analysis: users' guide to the medical literature. JAMA. 2014 Jan 22-29;311(4):405-11. doi: 10.1001/jama.2013.285063. PubMed

WALLACH 2017
Wallach JD, Sullivan PG, Trepanowski JF, et al. Evaluation of Evidence of Statistical Support and Corroboration of Subgroup Claims in Randomized Clinical Trials. JAMA Intern Med. 2017 Apr 1;177(4):554-560. doi: 10.1001/jamainternmed.2016.9125. PubMed

VILGRAIN 2017
Vilgrain V, Pereira H, Assenat E, et al. SARAH Trial Group. Efficacy and safety of selective internal radiotherapy with yttrium-90 resin microspheres compared with sorafenib in locally advanced and inoperable hepatocellular carcinoma (SARAH): an open-label randomised controlled phase 3 trial. Lancet Oncol. 2017 Dec;18(12):1624-1636. PubMed

YOON 2018
Yoon SM, Ryoo BY, Lee SJ, et al. Efficacy and Safety of Transarterial Chemoembolization Plus External Beam Radiotherapy vs Sorafenib in Hepatocellular Carcinoma With Macroscopic Vascular Invasion: A Randomized Clinical Trial. JAMA Oncol. 2018 May 1;4(5):661-669. PubMed