Evidence Review Group Report commissioned by the NHS R&D HTA Programme on behalf of NICE

Sorafenib for advanced hepatocellular carcinoma

Produced by: West Midlands Health Technology Assessment Collaboration
Department of Public Health, Epidemiology & Biostatistics
University of Birmingham
Edgbaston
Birmingham B15 2TT

Authors: Martin Connock[1]
Jeff Round[2]
Sue Bayliss[1]
Sandy Tubeuf[2]
Wendy Greenheld[1]
David Moore[1]

[1]Department of Public Health, Epidemiology & Biostatistics, University of Birmingham.
[2]Institute of Health Sciences, University of Leeds

Acknowledgements: Daniel Palmer for clinical advice, Mohammed Mohammed and Lucinda Billingham for statistical advice, and Janet Farren and Karen Biddle for administrative assistance with this project.

Correspondence to: David Moore
Department of Public Health and Epidemiology,
University of Birmingham
Edgbaston
Birmingham B15 2TT

Date Completed: March 2009

This report was commissioned by the NIHR HTA Programme as project number 08/75/01

Declaration of competing interests of the authors: NONE

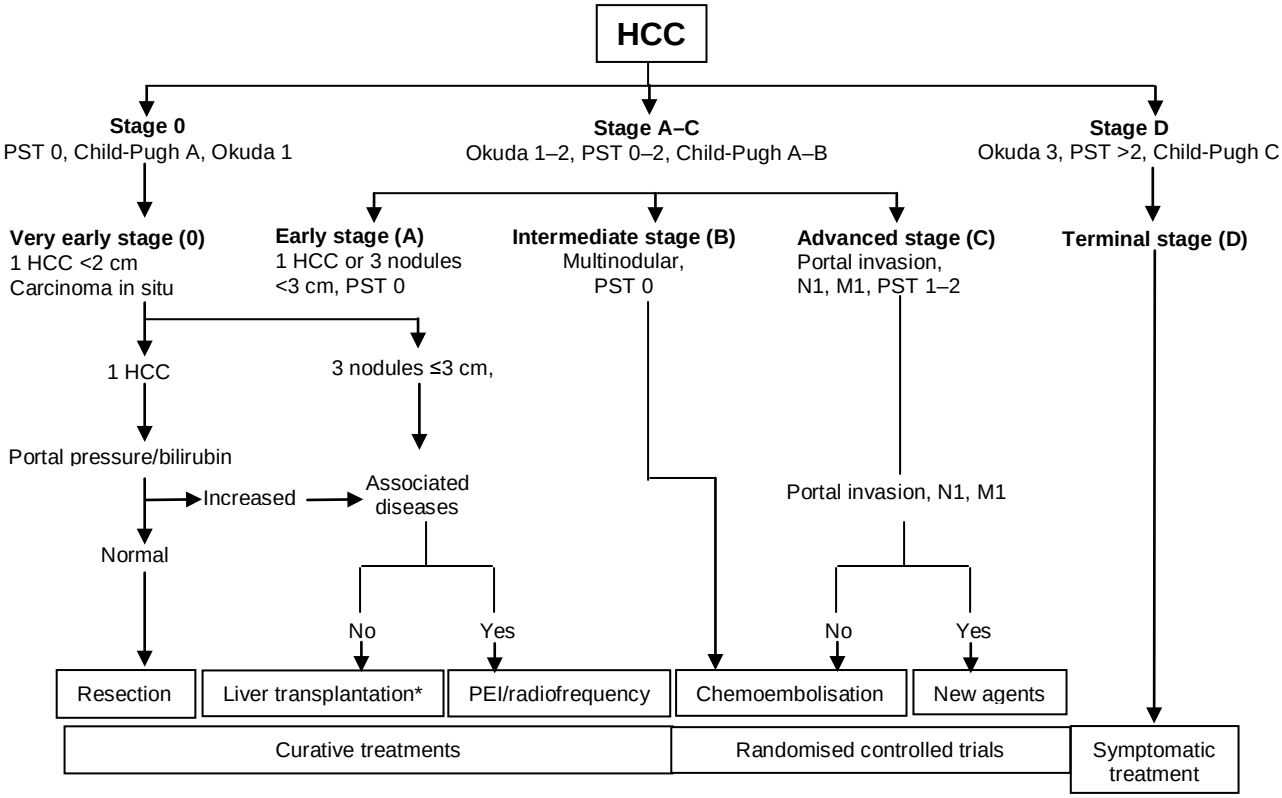
Contents:

1 SUMMARY

1.1 Scope of the submission

The submission considers the effectiveness and cost-effectiveness of sorafenib (Nexavar) in the treatment of advanced hepatocellular carcinoma (HCC) when surgical or loco-regional therapies have failed or are unsuitable.

The treatment pathway based on the Barcelona Clinic Liver Cancer (BCLC) staging system and proposed by Llovet, et al in 2003[1] is shown below. This is consistent with UK guidelines published in 2003[2].



A more recent version of the above pathway has sorafenib as the therapy for advanced HCC.[3] The as yet unpublished draft update of UK guidelines[4] consider sorafenib "*is the standard of care"* for patients with advanced HCC for whom no potential curative option is available and that systemic chemotherapy with standard agents *"can be offered where no alternative therapy is available"*.

Both the NICE scope and the manufacturer's submission state the aim of the appraisal to be the assessment of sorafenib for advanced HCC but neither provide a definition of advanced disease.

Several staging systems have been developed for HCC, each categorising the disease based on severity. None are universally employed. Those based on clinical items include:

- the Okuda system,
- the Cancer of Liver Italian Program (CLIP) system,
- the Barcelona-Clinic Liver Cancer (BCLC) system (Appendix 1),
- the Tumour Node Metastasis (TNM) system,
- the GRoupe d'Etude et de Traitement du Carcinoma Hepatocellulaire (GRETCH) system,
- the VIenna SUrvival model (VISUM),
- the Chinese University Prognostic Index (CUPI)
- the Japanese Integrated System (JIS)

According to a recent publication by Camma et al 2008[5] the TNM system has not been widely adopted in hepatology and the systems most widely used are the BCLC, GRETCH, and CLIP. In a recent study of UK study HCC patients were categorised according to Okuda, BCLC and CLIP classifications[6].

The European Medicines Agency (EMEA) specified the following wording for the indication of sorafenib[7]:
"*Nexavar is indicated for the treatment of hepatocellular carcinoma (see section 5.1)".*

Section 5.1 of the Summary of Product Characteristics (SPC) categorises the investigated HCC population according to the TNM[8] and BCLC[9] staging systems.

The TNM system of The American Joint Committee on Cancer (AJCC) has four stages (I to IV) of increasing severity (stage III having three subcategories A, B and C). This system does not designate an "advanced disease" category.

The Barcelona-Clinic Liver Cancer (BCLC) system has four stages termed A, B, C and D.

- Stage A called “early HCC” has four subcategories (A1 to A4)
- Stage B is termed “intermediate HCC”
- stage C “advanced HCC”
- stage D “end-stage HCC”

The SPC (section 5.1) describes the patient population investigated to establish clinical effectiveness (SHARP trial [10]) according to TNM stage as follows:

stage I	< 1%;
stage II	9.3%
stage III	40.7%
stage IV	48.8%

According to the Barcelona-Clinic Liver Cancer (BCLC) system the patients were staged as follows:

stage B	17.5 %;
stage C	82.4%
stage D	< 1%

Thus clinical effectiveness for licensing was determined in a patient population with more than 80% having “advanced” HCC according to the BCLC system.

The aetiology of HCC is associated with hepatitis virus infection, alcoholism, insult from agents such as aflatoxin and rare genetic conditions such as haemochromatosis all of which give rise to liver pathology. According to draft UK guidelines most HCC patients in the UK have chronic liver disease at the stage of cirrhosis.[4] The Child-Pugh grade (Appendix 2) is a widely used method of assessing patient risk according to liver function impairment and classifies patients into three grades (A, B, C) of increasing severity. Poor Child-Pugh grade influences the choice of therapies and correlates with poor prognosis. Patients with advanced HCC eligible for sorafenib would be a mixed population with various Child-Pugh grades of liver function. The SHARP trial which provided the evidence base for European licensing of

sorafenib recruited almost exclusively Child-Pugh A patients, i.e. those with less severe liver function impairment.

1.2 Summary of submitted clinical effectiveness evidence

The submitted clinical evidence consisted of two RCTs (the SHARP[10] and Asia-Pacific[11] studies) that estimated the effectiveness of sorafenib plus best supportive care versus placebo plus best supportive care in the treatment of advanced HCC, and an uncontrolled open label study (Abou-Alfa et al[12]) of sorafenib therapy for advanced HCC. The SHARP and Abou-Alfa studies recruited participants of predominantly European ethnicity while the Asia-Pacific study enrolled patients of predominantly non-European ethnicity. SHARP and Asia-Pacific studies randomised 602 and 226 patients respectively of whom >95% in both trials were classified as Child-Pugh grade A. In the Abou-Alfa study there were 98 Child-Pugh grade A and 38 grade B patients. The submission made the assumption that the results from the SHARP study were those most likely to reflect effectiveness in UK patients and consequently used the Asia-Pacific and Abou-Alfa studies for supporting evidence.

The RCTs demonstrated that sorafenib significantly extended median overall survival (by 11.9 weeks in SHARP and 10 weeks in the Asia-Pacific study) and that sorafenib significantly extended the median time to radiologically determined disease progression (by 11.7 weeks according to independent assessment or [redacted] weeks according to investigator assessment in SHARP, and by 6.1 weeks in the Asia-Pacific study). The RCT results indicate that sorafenib therapy has little or no effect on health related quality of life of HCC patients (as estimated with the FACT-Hep questionnaire) or upon the time to symptomatic progression and induces complete or partial tumour responses in less than [redacted] of patients. Adverse events were common in both arms of the RCTs with an excess of hypertension and of dermatological and gastrointestinal adverse events in the sorafenib groups. However, withdrawal from treatment due to adverse events was almost the same in sorafenib and placebo arms. In SHARP dose reductions due to adverse events were more common in the sorafenib group than the placebo group (32% vs 13%).

1.3 Commentary on the robustness of submitted clinical effectiveness evidence

1.3.1 Strengths

The strength of the submitted clinical effectiveness is due to the following factors:

- The estimates of clinical effectiveness are based on two RCTs of sufficient power to demonstrate that sorafenib induces significant improvements in two outcomes of major interest, namely overall survival and time to radiologic disease progression.

- It is highly unlikely any relevant RCT evidence investigating sorafenib versus placebo has been missed.

- Overall the submission was of an acceptable standard.

1.3.2 Weaknesses

- The effectiveness evidence in the submission pertained almost exclusively to patients with relatively good liver function (Child-Pugh grade A), but these patients represent a subgroup, although the major one, of those addressed by the decision problem.

- The results from SHARP included analyses using both independent assessment and investigator assessment of radiographs to determine major outcomes such as time to disease progression and disease response. There were clear discrepancies between these analyses particularly for sorafenib treated patients, but this was not remarked upon in the submission and potential explanations were not explored.

- The submission included a journal reference to the latest results from the uncontrolled open label study of Abou-Alfa et al[13] but did not present this data. This abstract presented some information about the effectiveness of sorafenib in patients with Child-Pugh grade B liver function unavailable elsewhere. This information was germane to the

decision problem and implied that Child-Pugh grade B patients may respond less well to sorafenib than Child-Pugh grade A patients.

- The submission included three systematic reviews the separate elements of which were difficult to disentangle.

- The methodology for the main review (i.e. that in the submission itself) although generally systematic was marred by poorly defined inclusion/exclusion criteria.

1.3.3 Areas of uncertainty

Uncertainties about effectiveness of sorafenib versus placebo are:

- By how much does sorafenib extend the time to radiological disease progression (i.e. which analysis, independent or investigator, is the more reliable?)

- What is the effectiveness of sorafenib in Child-Pugh B patients and are they legitimate candidates for sorafenib therapy in the absence of good evidence of effectiveness for this group?

- What proportion of advanced HCC patients eligible for sorafenib treatment would be classified as Child-Pugh grade A and Child-Pugh grade B, and therefore what is the effectiveness of sorafenib in the overall population defined by the decision problem?

- What is an agreed operational definition of advanced HCC?

1.4 Summary of submitted cost-effectiveness evidence

Given the evidence presented in the submission, sorafenib does not appear to represent a cost-effective use of NHS resources given the commonly

accepted upper threshold of willingness to pay of £30,000/QALY. In the base-case analysis sorafenib generates an additional 0.51 life years and 0.36 QALYs than best supportive care, at an incremental cost of £23,232. The resulting ICERs are £45,502/LY and £64,754/QALY. Sensitivity analysis included in the submission suggests that sorafenib approaches the upper threshold limit in certain sub-groups, although for no subgroup does it fall below the threshold. In a "best case" scenario analysis the best performing subgroup ([redacted]) appears to be cost-effective, [redacted].

1.5 Commentary on the robustness of submitted cost-effectiveness evidence

1.5.1 Strengths

- The overall quality of the economic evaluation is good.
- The manufacturer has provided a reliable, internally valid model, based primarily on robust clinical data from a randomised controlled trial. The model structure is appropriate to the decision problem.
- The NICE reference case for economic evaluations was closely followed, although some deviations have been noted in this report. It is the view of the ERG that were the reference case strictly followed, this would not materially effect the results presented by the manufacturer.

1.5.2 Weaknesses and areas of uncertainty

- The choice to focus on best-supportive care as the only comparator in the economic model is not unequivocally supported by the evidence that doxorubicin can and maybe used in clinical practice in the UK.

- The economic evaluation relies heavily on the use of expert opinion for estimating resource use for the treatments in the model. As a result, significant uncertainties remain over the cost of treatments that are not adequately expressed in the model.

- The economic evaluation relies on the use of mapped estimates of health related quality of life. Although this is acceptable according to the NICE methods guidance, it is a second best option and primary data collection of health related quality of life information within the clinical trial using a validated preference based instrument would be more appropriate.

- The submission does not include cost-effectiveness estimates of the independent assessment of TTP. This is an important omission, and the ERG have undertaken some sensitivity analysis to address this.

1.6 Key issues

The appraisal appears to hinge on the following issues:

- To what extent does the clinical effectiveness observed in the SHARP RCT validly apply for the NICE scope and licensed UK populations?
- Is the manufacturer's cost-effectiveness strategy reasonable in view of a possible mismatch between the SHARP population and the UK advanced HCC population?
- Did the selection of best supportive care as the sole comparator represent a reasonable approach?
- Has the health related quality of life for sorafenib and best supportive care patients been appropriately ascertained in the submission?
- Does an average extension in survival of 2.8 months (83 days) at an incremental life time cost of £23,232 and an incremental cost-effectiveness ratio of £45,502/life year gained represent good value for money?
- Was a convincing case made that sorafenib should be considered within the "*Treatments for End Life*" policy defined by NICE?

2 BACKGROUND

2.1 Critique of manufacturer's description of underlying health problem

Section 4.1 of the submission includes the following:

> Unfortunately, there are often no specific symptoms, and less than 30% of patients are diagnosed in the early stages where liver tumours are considered more amenable to curative resection or transplantation. Some patients may be suitable for "loco-regional" treatments: ablation (radiofrequency ablation (RFA); percutaneous ethanol injection (PEI) or cryosurgery);(chemo) embolisation, and radiotherapy. For patients where surgical or loco-regional treatments have failed or are unsuitable (approximately 25-35% of HCC patients (2), systemic therapy is the only active treatment option. Prior to sorafenib, no drug or regimen could be defined as the standard systemic treatment in advanced HCC as no treatment had ever been shown to demonstrably improve overall survival (OS) in a randomised controlled trial (RCT) (6,27) .
>
> Doxorubicin is used in a minority of patients, but low overall response rates (10-15%) and the risks associated with its use often outweigh any short-term benefits, and clinicians usually opt for a best supportive care (BSC) approach instead. Therefore, within the present therapeutic landscape, the prognosis for patients with advanced HCC is bleak, with 5-year survival rates of <5%(18) . Consequently, there is a compelling clinical need for effective treatments in order to improve the outlook for these patients.
>
> With the introduction of sorafenib, there has been a noticeable shift in opinion as to the standard systemic treatment in advanced inoperable HCC. Due to sorafenib being shown to prolong survival in this patient group, several guidelines and review papers (20) , including the revised UK guidelines (21) (as yet unpublished) now include sorafenib as the standard of care systemic therapy for patients with advanced HCC for whom no potential curative option is available.

Comment:

- The estimate that for approximately 25 to 35% of UK HCC patients systemic therapy represents the only active treatment option was based on opinion of a "UK advisory board" convened by Bayer (reference 2 in the submission). The remit of this advisory board, its constitution and the manner of its selection are not described.
- The evidence review group (ERG) requested clarification regarding the nature of the UK advisory board (submission ref 2) that estimated the

proportion of HCC patients that would be eligible for sorafenib. The following response was received:

The meeting was attended by 6 clinical advisors, specialising in HCC in the UK. The names have not been disclosed for confidentiality reasons.

The relevant section of these minutes can be summarised as follows:

Agenda Item: HCC Treatment Paradigm

The vast majority of HCC cases (80-90%) are reviewed for loco regional therapy (RFA, TACE, TEA, internal radiation) and around half of these cases are found to be unsuitable for loco regional treatment and are either given palliative care, no treatment or (rarely) doxorubicin.

The group agreed that an estimated 25-35% of HCC patients may be candidates for Nexavar.

2.2 Critique of manufacturer's overview of current service provision

From the submission:

A systematic review of the literature, prior to the introduction of sorafenib, suggested that no anti-cancer treatment had been clearly identified as the treatment of choice in this advanced, inoperable patient group (22) . Best supportive care (BSC) is the most common patient management strategy. Hence placebo / BSC is justified as being a relevant comparator arm in studies evaluating novel agents such as sorafenib for the treatment of HCC (7) .

Studies involving doxorubicin, placebo or BSC, identified during the systematic review, confirmed the lack of clarity on standard treatment and the heterogeneity in terms of dosage and treatment regimens, study population characteristics and outcome measures. Although doxorubicin may be used in a small number of patients, it's use is not supported by current guidelines and data identified in the systematic review (22) was insufficient to support even an indirect comparison. The doxorubicin trials are small, with methodological flaws (e.g. lack of intention to treat analysis) and the heterogeneity of the patient groups makes the true effects of doxorubicin difficult to determine. The uncertainty about best practice and treatment options for patients with inoperable advanced HCC is clearly highlighted by the lack of direction regarding specific therapy recommendations in guidelines produced prior to the introduction of sorafenib (see section 4.6).

Since sorafenib approval, there has also been a noticeable shift in opinion as to the standard systemic treatment in advanced inoperable HCC. Due to sorafenib being shown to prolong survival in this patient group, several guidelines and review papers (20) , including the revised UK guidelines (21) (as yet unpublished) now include sorafenib as the standard of care systemic therapy for patients with advanced HCC for whom no potential curative option is available.

Comment:

- The dismissal of any role for doxorubicin in HCC diverts attention from the fact that there may be continuing genuine debate about its place in the management of HCC. This interest is exemplified in a Bayer-sponsored phase II randomised trial of sorafenib plus doxorubicin versus doxorubicin ("final results" published in abstract, Abou-Alfa et al 2008[13]); the single treatment arm with doxorubicin was presumably submitted for ethical approval by the sponsors and thus would have been viewed as a viable therapy.
- With respect to sorafenib, the National Comprehensive Cancer Network (NCCN) guidelines (version 2, 2008) [14] (submission ref 20) caution that there is "*limited safety data available for Child-Pugh grade B patients*" and that the drug has been shown effective for "*selected patients*". Thus uncertainty about best practice and treatment options remains despite the implication in the submission that uncertainty only existed prior to the introduction of sorafenib.

Overall the overview is reasonable. Arguably there is too great an emphasis on the limitations of alternative systemic therapies but this does reflect the three sets of guidelines that are quoted and referenced in the submission. Further minor comments can be found in Appendix 3.

3 Critique of manufacturer's definition of decision problem

The manufacturer's definition of the decision problem taken from the submission is shown in Appendix 4 .

Except in the definition of suitable comparator the submission statement appears consistent with the remit issued by NICE which was: "*To appraise the clinical and cost effectiveness of sorafenib, within its licensed indication, for the first line systemic treatment of advanced hepatocellular carcinoma*". However, the effectiveness and cost-effectiveness analyses presented in the submission address a narrower population than indicated above.

The following section outlines a variety of issues:

3.1 Population

The NICE final scope indicates: "*Adults with advanced hepatocellular carcinoma whose disease is unsuitable for local or loco-regional curative therapy or has progressed after those types of therapy*."

The manufacturer's submission states:

> The decision problem addressed in the submission is the clinical benefit and cost-effectiveness of sorafenib as a treatment in those patients with advanced stage hepatocellular carcinoma disease who have failed or are unsuitable for surgical or locoregional therapies.

Comment:

- The manufacturer's submission takes its clinical evidence from the SHARP RCT that recruited a narrower population than that defined in the decision problem with respect to liver function (Child-Pugh grade A) and a performance status at the upper end of the Eastern Cooperative Oncology Group (ECOG) scale (i.e. 0 to 2). It was this same clinical evidence that fed the reference case for the economic model.
- There is no definition of advanced HCC.

3.2 Intervention

The NICE final scope indicates this to be sorafenib.

The manufacturer's submission amplifies this as follows:

> Sorafenib is administered orally in the form of 200 mg film coated tablets. The recommended dose of sorafenib in adults is 400 mg twice daily (bd; equivalent to a total daily dose of 800 mg). Treatment should continue as long as clinical benefit is observed or until unacceptable toxicity occurs.

3.3 Comparators

The comparator in the NICE final scope is stated as:

"*Standard care which may include doxorubicin, cisplatin or biological agents, depending on performance status and disease severity*"

The submission states:

> Sorafenib will be compared to best supportive care.
>
> Due to the underlying liver disease and lack of effective treatments, patients diagnosed with advanced HCC have a bleak prognosis. Sorafenib is the only treatment to have demonstrated a survival benefit in advanced HCC for over 30 years (3) . No systemic agent has shown survival benefit versus placebo in HCC in more than 75 randomised controlled trials (4) and, in most cases, such treatments are associated with a high rate of side effects. As a result, there are no treatments, other than sorafenib, with FDA and/or EMEA approval for advanced HCC. Furthermore, because of the advanced nature of the disease, surgery is not a treatment option.
>
> Guidelines (BSG 2003) (5) recommend that systemic chemotherapy with standard agents have a poor response rate and should only be offered in the context of clinical trials of novel agents. Best supportive care is the most appropriate comparator for these patients. This is supported by various reviews (6,7) , meta-analyses (8) and systematic reviews (4,8-11) published over the past decade which conclude that no anti-cancer treatment has clearly been identified as either a 'gold standard' or to demonstrably improve overall survival.

Comment.

- A mismatch between NICE's standard care comparator and the manufacturer's best supportive care (BSC) is evident.
- If no systemic agent other than Sorafenib has EMEA approval for advanced HCC then, assuming NICE's "within license" stipulation applies to both intervention and comparator, it is reasonable that BSC should be the appropriate comparator for this appraisal. The ERG requested clarification

regarding the relevant licensing status of doxorubicin, the manufacturer's response is given below:

> As per the information available on The Electronic Medicines Compendium (EMC), there are four companies (Hospira UK Ltd, Hameln pharmaceutical ltd, Pharmacia Limited and medac GmbH) which market doxorubicin in UK. However to the best of Bayer Schering Pharma's knowledge, HCC as a specific indication in not included in Therapeutic indications section 4.1 of the SmPC for any of the above companies. However "use in solid tumours" is mentioned in a couple of the aforementioned SmPCs

- As HCC is a solid tumour it can be considered that doxorubicin is licensed to treat it when solid tumour is specified in the Summary of Product Characteristics (SmPC).

Further minor comments can be found in Appendix 5 .

3.4 Outcomes

The NICE scope indicates the outcome measures considered should include:

- *overall survival*
- *progression-free survival*
- *time to symptomatic progression*
- *tumour response*
- *health-related quality of life*
- *adverse effects of treatment*

The manufacturer's submission states that the outcomes listed will be presented in the submission; with regard to health related quality of life (QoL) it goes on to state:

> Advanced HCC is a unique condition which poses methodological issues when evaluating the impact of new treatments on health related quality of life.
>
> Patients with hepatocellular carcinoma are heterogeneous, with a diverse range of underlying causes of cirrhosis, including hepatitis B, hepatitis C, alcoholism and haemochromatosis. In some patients, typically younger women, HCC may develop where cirrhosis is not present. Due to this diverse liver disease, it is particularly difficult to disentangle the effect of the advanced HCC, underlying liver disease and interventions on quality of life. More specifically, quality of life is likely to be affected by the symptoms of the underlying liver disease, including liver failure, irrespective of whether the tumour has stabilised or regressed. As a result, it is not possible to demonstrate the impact of treatments in advanced HCC on quality of life, and no robust and reliable utility data is available that separates out the effect of the primary liver cancer from the underlying liver disease causes.

Comment:

- Although heterogeneity of HCC may differ from many other diseases for which QoL has been estimated; heterogeneity itself does not preclude estimation of QoL.
- Disentangling the treatment effect upon the primary liver cancer from that on underlying liver disease is not material to a QoL estimation since the point of interest here is the whole patient and not just "the tumour".

3.5 Time frame

The NICE scope suggests that the time horizon for the economic evaluation "*should be sufficiently long so as to incorporate all the important costs and benefits related to this condition*".

The submission states:

> Due to the advanced nature of the disease, the model will be a lifetime model, consisting of three health states; non-progressive advanced disease, progressive disease, and death.

Comment.

- The manufacturer's model had a 14 year time horizon that was based on the time for patient survival to reduce to 1% of that at start of treatment. Shorter time horizons were used in sensitivity analyses, an approach that seems reasonable.

3.6 Other relevant factors

As other relevant factors the NICE final scope indicates:

"*If evidence permits, the appraisal will seek to identify subgroups of individuals for whom sorafenib may be particularly clinically and cost effective, for example by age, performance status or degree of underlying cirrhosis. Guidance will only be issued in accordance with the marketing authorisation*"

The submission observes:

> Applying a single estimate of cost-effectiveness to the overall advanced HCC group of patients is unreliable because of the unique large variation in underlying disease (e.g. liver cirrhosis), rarely seen in other cancers, it is therefore of utmost importance to base decisions on patient sub-groups where the health and economic outcomes are most likely to vary considerably from the overall mean.
>
> It is acknowledged there is a high degree of variability around the point estimate of cost effectiveness due to the heterogeneous nature of the disease and the difficulty disentangling the underlying liver disease and treatment effects. For these reasons it would be appropriate to collect further evidence as recommended under the end of life scheme.

Comment:

- The submission statement seems consistent with the final scope but fails to identify criteria that might be employed to define patient subgroups. The main issue is that the reference case submitted by the manufacturer already only addresses a subgroup of the population defined in the NICE scope, namely those advanced HCC patients that start treatment with Child-Pugh grade A liver function.
- The table below summarises the major points of concordance and discordance between the NICE scope definition of the decision problem and that actually addressed by the evidence presented on the submission.

	Decision problem defined in the NICE scope	Decision problem addressed by the evidence submitted
Population	Adults with advanced hepatocellular carcinoma whose disease is unsuitable for local or loco-regional curative therapy or has progressed after those types of therapy	In practice the submitted evidence related to a subpopulation of advanced HCC patients who had relatively mild impairment of liver function (Child-Pugh A) and relatively good ECOG performance status (> 53% status = 0)
Intervention	Sorafenib (Nexavar)	Sorafenib (Nexavar)
Comparator(s)	Standard care which may include doxorubicin, cisplatin or biological agents, depending on performance status and severity	The evidence presented only related to best supportive care. Other potential comparators were considered ineffective and were not considered.
Outcomes	The outcome measures to be considered include: Overall survival Progression free survival Time to symptomatic progression Tumour response Health related quality of life Adverse effects of treatment	The outcome measures in submission were: Overall survival Progression free survival Time to symptomatic progression Tumour response Health related quality of life Adverse effects of treatment
Economic Analysis	The reference case should be expressed in terms of incremental cost per quality adjusted life year. The time horizon should be sufficiently long so as to incorporate all the important costs & benefits related to the condition. Where the evidence allows, any likely dose adjustment during the treatment should be taken account of. Costs considered from an NHS and Personal and Social Services Perspective	Cost effectiveness was expressed in incremental £/QALY and incremental £/LYG. The time horizon was 14 years. Dose adjustments were taken into account. Costs were considered from the NHS and PSS perspective.
Subgroups to be considered	If the evidence permits, the appraisal will seek to identify subgroups of individuals for whom sorafenib may be particularly clinically and cost effective, for example by age, performance status or degree of underling cirrhosis. Guidance will only be issued in accordance with the marketing authorisation.	The subgroups addressed in the submission were only those encompassed with the Child-Pugh grade A patient population.

4 CLINICAL EFFECTIVENESS

The submission aimed at reviewing evidence on the effectiveness of sorafenib versus best supportive care (BSC) for advanced HCC. BSC was interpreted as no active systemic therapy. An additional review (cited as reference 22 in the submission) was presented as a separate document and had the stated objective: "*To gather evidence pertaining to systemic anti-cancer therapies in advanced hepatocellular carcinoma (HCC) for input into the evaluation of the clinical and pharmacoeconomic benefits of sorafenib (Nexavar) in HCC when compared with current UK clinical practice."*

It was difficult to delineate which parts of the submission referred specifically to which review.

4.1 Critique of manufacturer's approach

4.1.1 Description of manufacturers search strategy and comment on whether the search strategy was appropriate.

A summary of the manufacturer's search strategy is shown in Appendix 6

Comment:

- The full details of the strategies and databases searched for the effectiveness review were clearly documented in the submission. The submission searches were kept intentionally broad. Searches were restricted to English language, which increases the risk of publication bias. No date limits were used.

- Although the choice of terms and combination of MeSH and controlled vocabulary is appropriate for construction of a broad strategy, the terms used to describe the population are more restrictive than may be appropriate. Specifying terms to capture systemic therapies means that relevant studies which do not use these terms may be missed. Similarly the

use of the Boolean NOT operator in line 10 seeks to exclude studies on other types of cancer. However, this strategy means that studies which focus on liver cancer but also mention any of these other listed cancers will not be located. A simplified strategy for the population (as far as line 6) would ensure a more inclusive search. The ERG tested a broader strategy (see Appendix 7) which resulted in 63 hits on MEDLINE and 419 on EMBASE. Upon examination no additional relevant fully published studies were found.

- Ongoing studies identified in the submission are shown in Appendix 8 together with comments. The study BAY 43-9006 listed in the submission as ongoing is a completed study of doxorubicin plus sorafenib versus doxorubicin, the "final results" of which are available as a presentation downloadable from the internet (Abou-Alfa et al[13]).

4.1.2 Statement of the inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate.

From the submission, the inclusion and exclusion criteria were:

Included: Randomised, controlled trials (RCTs) comparing sorafenib as a single agent with other therapies (including placebo), involving patients aged 18 with a diagnosis of advanced inoperable HCC. Patients were to have had no prior *systemic* therapy (as this was one of the inclusion criteria for the phase III SHARP trial). Excluded: Phase I studies, open-label studies, dose-ranging studies, non-English language references, trials involving intra-arterial agents or Transarterial embolisation (TAE) and Transarterial Chemo-embolisation (TACE) studies were excluded. See 10.2.6 for list of full inclusion and exclusion criteria for the overall search.

Comment:

- The above description of the inclusion and exclusion criteria is confusing and unclear:
    - The requirement that patients are 18 years old is an obvious error

- The criteria appear to have been guided by the inclusion criteria for recruitment to the SHARP trial; their objectivity could therefore be questioned
- The inclusion/exclusion status of Phase II studies is unclear
- No explicit procedure is described for dealing with abstracts
- Directing the reader to section 10.2.6 (in a separate document) "*for list of full inclusion and exclusion criteria for the overall search*" lacks requisite clarity by mixing two separate systematic reviews; the criteria in 10.2.6 differ from those in the main submission and are for a review with a different stated objective to that in the main submission document.

- The restriction of studies to only those using "*sorafenib as a single agent*" precluded the inclusion of potentially informative indirect evidence. In view of the probable scarcity of direct evidence this could be considered a limitation. The exclusion of non-English language studies could be viewed as a weakness opening the review to potential publication bias.

4.1.3 Table of identified studies. What studies were included in the submission and what were excluded?

The submission for this section is shown in Appendix 9.

Comment:

- There is no explicit list of the studies that were included. However, it was abundantly clear which three studies were actually used for the evidence base of clinical effectiveness. These were:
  - the SHARP study[10], a placebo-controlled multicentre RCT with sorafenib in 602 mostly European patients with advanced HCC.

- a multicentre RCT of sorafenib versus placebo conducted in 226 patients from a population with endemic hepatitis B (the Asia-Pacific study[11])

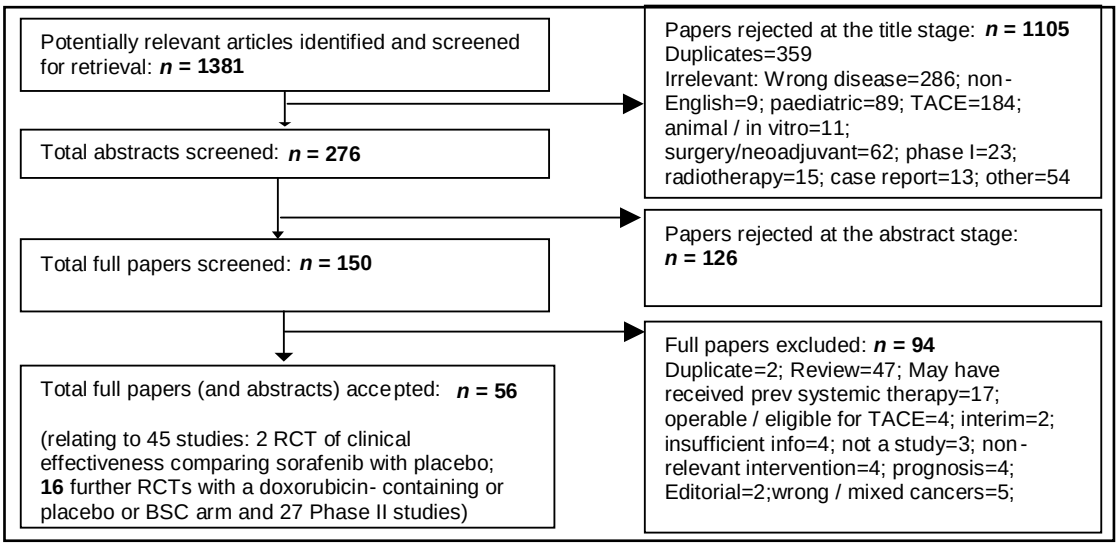
- an uncontrolled open label study (Abou-Alfa 2006[12] and Abou-Alfa 2008[15]) with 137 predominantly European patients.

- The submission stated that the SHARP RCT would "*provide the evidence for the clinical effectiveness of sorafenib in HCC in this submission*". The other two studies "*will be provided as supporting data*" . One of these two studies was an open label study and thus satisfied the exclusion criteria.

- The uncontrolled open label study of Abou-Alfa was given two citations in the submission (i.e. references 24 and 37). Reference 37 was published in the May 2008 supplement of the Journal of Oncology; this supplement contains several other abstracts about sorafenib in HCC and raises an issue concerning whether these should be included or excluded (see next section).

The manufacturer's flow chart for identification of included studies is shown below.

Figure 1 Flow chart of the clinical evidence screening process for sorafenib in inoperable advanced HCC



Comment:

- This figure does not describe the process leading to the identification of the 3 included / relevant sorafenib studies.
- The 56 papers (and abstracts) of 45 studies (in the final box) refer to the additional systematic review presented in a separate document from the main submission.
- A list of excluded studies for the main submission was not found.
- A consistent method for dealing with abstracts has not been implemented.

4.1.4 Details of any relevant studies that were not included in the submission?

Additional searches by the ERG did not identify any full papers describing studies that would fit the submission's inclusion criteria.

Since the submission inclusion criteria were ill defined the ERG also applied its own criteria as follows:

Population	patients with advanced HCC unsuitable for surgery and loco-regional interventions or in whom such interventions had failed
Intervention	sorafenib
Comparator	any
Outcomes	survival, time to progression, quality of life
Study design	RCTs or other controlled studies.
Publication	only fully published studies accepted (no abstracts)

Using these criteria the ERG failed to identify any further studies.

A reference for the included study by Abou-Alfa (reference 37 in the submission) was an abstract from the May 20 2008 supplement of J Oncology vol 26. Another abstract from this supplement entitled *"Efficacy and tolerability of single agent sorafenib in poor risk advanced hepatocellular carcinoma patients"* was neither listed as an included or excluded study; if reference 37 was included then this other abstract should also have been.

Limiting studies to only those with sorafenib as sole systemic agent precluded a consideration of indirect / mixed comparison evidence such as might be derived from the Phase II RCT of doxorubicin + sorafenib versus doxorubicin available as a presentation in abstract.[13] The submission states "data identified in the systematic review (ref 22) was insufficient to support even an indirect comparison". This statement refers to the additional systematic review attached as an appendix to the main submission, an aim of this additional review was "*to allow for any later decisions to do indirect comparisons between sorafenib and other relevant treatments to the UK*". The inclusion criterion for study type for this additional review was: "*Studies with sorafenib, placebo, doxorubicin or best supportive care as a treatment arm.*" According to the submission best supportive care was interpreted as no active systemic treatment and consequently this inclusion strategy would fail to select all studies that could potentially provide data allowing an indirect/mixed treatment comparison approach. Also it would not capture studies investigating other potential comparators to sorafenib defined in the decision problem by NICE.

4.1.5 Description and critique of manufacturers approach to validity assessment

Sections from the submission describing the critical appraisal of the SHARP trial and of the uncontrolled open label study are provided in Appendix 10.

Comment:

- The submission's appraisal of SHARP included consideration of: allocation concealment, randomisation procedure, justification of sample size, adequacy of follow up, blinding, baseline comparability, and appropriateness of statistical analysis (including intention to treat). This approach is reasonable. No particular validation instrument was identified and it is unclear how many reviewers undertook this appraisal. The appraisal is fair except that it omits to mention that the published account of the SHARP trial failed to include the QoL outcome measured using the FACT-Hep instrument potentially opening it to the charge of outcome selection bias. This outcome was partially reported in the submission itself and was designated "commercial in confidence" (CIC).

- The validity of the supporting RCT (Asia-Pacific study) was not appraised in the submission.

- The validity of the supporting uncontrolled study (Abou-Alfa 2006[12]) was not considered in the submission, as although section 6.8.3 was headed "*Critical appraisal of relevant non-RCTs*" no actual appraisal was presented.

4.1.6 Description and critique of manufacturers outcome selection

From the submission:

> The **primary endpoints** in SHARP were:
>
> 1. Overall survival (OS)
> 2. Time to symptomatic progression (TTSP)
>
> The primary endpoints were assessed independently. If the analysis were positive for either endpoint, the efficacy of sorafenib in HCC was to be considered established.
> Secondary endpoints were:
>
> 1. Time to progression (TTP)
> 2. Overall Disease Control Rate (DCR)
> 3. Quality of Life : Functional Assessment of Cancer Therapy – Hepatobiliary (FACT-Hep) response rate
>
> [redacted] Other endpoints included safety, population pharmacokinetics, [redacted] [Of the 'other' endpoints, only safety results are reported in this submission.]
> Due to the difficulty in distinguishing whether clinical deterioration or death in patients with HCC is as a result of HCC progression or deterioration of liver function and complications of underlying cirrhosis, TTP (based only on radiologically-documented tumour progression) was included as a secondary endpoint rather than progression-free survival (PFS).
> [redacted]

Comment:

- Although the above list of outcomes are those defined for the SHARP trial rather for the submission's effectiveness review they correspond to those identified by NICE as appropriate for the decision problem.
- QoL assessment with the FACT-Hep instrument was not reported as an outcome in the published account of the SHARP trial.[10]
- It is not clear how [redacted] of TTP.
- It is not explicit that TTP was assessed separately by trial investigators and by independent assessors for over half of progressions observed.

A fuller description of the QoL outcome given in the submission follows:

Quality of Life : FACT-Hep response rate (see Appendix 8) (41,42)	The FACT-Hep was completed at baseline and at week 12, and at the 'end of treatment' visit for patients discontinued before week 12. The FACT-Hep response rate was based on the proportion of patients who achieved the 8-point Minimal Important Difference (MID) in baseline total score to FACT-Hep total score at week 12 (or end of treatment). The FACT-Hep response rate analysis was based on the sum of the scores from patient responses to 45 items in the questionnaire (see Appendix 8); FACT-Hep total score ranges from 0 to 180. Higher scores on all scales of the FACT-Hep reflect better quality of life or fewer symptoms. (42)

Comment:

- The sentence "*Higher scores on all scales of the FACT-Hep reflect better quality of life or fewer symptoms*" causes some confusion because elsewhere (the submission appendix 8) higher scores on the Physical wellbeing scale define poorer QoL while higher scores on the Functional wellbeing scale define better QoL.

- Further information from the submission about measurement of overall disease control rates and response rates makes it clear these were measured using RECIST criteria in SHARP and by WHO criteria in the uncontrolled open label study.[12] RECIST and WHO criteria are listed in Appendix 11.

4.1.7 Describe and critique the statistical approach used

From the submission:

6.3.5 Statistical analysis and definition of study groups

The primary population for efficacy analysis was the intent-to-treat (ITT) population, which was defined as all randomised patients. ….The main analysis was measured by log rank test (see Table 6). ….the study was stopped at an interim analysis, …. analysed using data cut-off 17th October 2006. The efficacy of sorafenib was to be considered established if either analyses based on the co-primary efficacy endpoints were positive.

The null hypotheses are:

H_0: The overall survival function of placebo is the same or better than that of Nexavar

H_A: The overall survival function of Nexavar is better than that of placebo

H_0: The TTSP function of placebo is the same or better than that of Nexavar

H_A: The TTSP function of Nexavar is better than that of placebo

The efficacy of sorafenib is considered established if either of the null hypotheses for Overall Survival or TTSP are rejected.

Table 6: Primary efficacy variables with primary and secondary statistical methods (3,28)

PRIMARY EFFICACY VARIABLE	PRIMARY STATISTICAL METHOD	SECONDARY STATISTICAL METHOD
Overall Survival (OS)	1-sided Log rank test (overall α = 0.02 stratified as per randomisation i.e. by region, ECOG PS and tumour burden).	Cox Regression Model Kaplan-Meier(KM) estimates and survival curves for each treatment group. The differences of KM estimates at some time points e.g. 6 months, 12 months, and corresponding 95% confidence intervals (CIs) were also calculated between the sorafenib and placebo groups. For each treatment group, FHSI-8 scores were summarised by visit for observed values and changes from baseline using descriptive statistics. Graphs of average score changes were generated to see if a time trend existed.
Time to Symptomatic Progression (TTSP)	1-sided Log rank test (overall α = 0.005 stratified as per randomisation i.e. by region, ECOG PS and tumour burden).	

Table 7: Primary and secondary statistical methods for secondary, tertiary and other endpoints

STUDY ENDPOINT	PRIMARY STATISTICAL METHOD	SECONDARY STATISTICAL METHOD
Time to Progression (TTP)	1-sided Log rank test (overall α = 0.025 stratified as per randomisation i.e. by region, ECOG PS and tumour burden) Kaplan-Meier(KM) estimates and plots presented for each treatment group. Based on independent radiological assessment (using data up to cut-off date for 1st interim analysis of OS, 12th May 2006 i.e. after approximately 227 radiological progression events had occurred) [NB This analysis was delayed to the end of study]	Based on investigator radiological assessment (using data up to cut-off date for 2nd interim analysis of OS, 17th October 2006) [redacted]

As of data cut-off of 17th October 2006, a total of 468 patients had discontinued double-blind treatment: 242 (80.1%) placebo patients and 226 (76.1%) sorafenib patients (see Figure 2). Overall, 132 (n=61 placebo; n=71 sorafenib) patients were still receiving double-blind study treatment. After discontinuing study treatment, patients were to enter post-treatment follow-up. As of 17th October 2006, 36 (11.9%) placebo patients and 47 (15.7%) sorafenib patients were still in follow-up.

6.5 Meta-analysis

Not applicable. Evidence from only one RCT was fully available for analysis and relevant to the decision problem (SHARP study)(3) . The Asia-Pacific trial (36) corroborates the findings from the SHARP study, however patients had different baseline and demographic characteristics making it inappropriate to perform a meta-analysis.

Comment:

- The statistical analyses listed appear appropriate to the SHARP trial. There was no explicit summary of methods to be used in the systematic review.
- The decision not to conduct a meta-analysis is defendable. The differences in demographic and baseline characteristics referred to were identified elsewhere in the submission; for convenience the ERG have tabulated these as follows:

POPULATION	ASIA-PACIFIC RCT China, Taiwan, Korea		SHARP RCT Europe, N & S America, Australia	
	Sorafenib (n=150)	**Placebo (n=76)**	**Sorafenib (n = 299)**	**placebo (n=303)**
Median age, years (range or SD)	51 (23-86)	52 (25-79)	64.9±11.2	66.3±10.2
Male	84.7%	86.8%	87%	87%
ECOG PS (%)				
0	25.3%	27.6%	54%	54%
1	69.3%	67.1%	38%	39%
2	5.3%	5.3%	8%	7%
Extrahepatic sites				
Lung	52.0%	44.7%	30%	21%
Lymph node	30.7%	34.2%	22%	19%
BCLC stage C (%)	95.3%	96.1%	82%	83%
Child-Pugh grade				
A	97.3%	97.4%	95%	98%
B	2.7%	2.6%	5%	2%
Cause of disease				
HBV infection	70.7%	77.6%	16%	19%
HCV infection	10.7%	3.9%	29%	27%
Alcohol	NR	NR	26%	26%
Unknown	NR	NR	19%	18%
Other	NR	NR	9%	10%
Number of tumour sites				
1	13.3%	6.6%	NR	NR
2	34.7%	35.5%	NR	NR
3	20.0%	18.4%	NR	NR
>4	32.0%	39.5%	NR	NR

4.1.8 Summary statement on the manufacturer's approach

The submission is complete firstly in that it is unlikely to have excluded any relevant RCT evidence of sorafenib used as a single agent and secondly in considering appropriate outcomes to judge clinical effectiveness.

The most appropriate comparator for the decision problem is a moot point. As stated in the submission, the literature appears to lack any study of sorafenib (as a single agent) versus any other systemic intervention. The submission took the view that doxorubicin was not a valid comparator stating that *"the doxorubicin trials are small, with methodological flaws…and the heterogeneity of the patient groups makes the true effects of doxorubicin difficult to determine".* According to UK expert clinical opinion[4] the use of doxorubicin or standard systemic agents other than sorafenib for this population should be within confines of clinical trials. The EMEA in their scientific discussion document on sorafenib considered a phase III RCT of nolatrexed versus doxorubicin[16] in advanced HCC (N = 445) and concluded on the basis of the observed 2.3 month median survival advantage for doxorubicin that, on balance, doxorubicin was likely an effective intervention. The EMEA scientific discussion document[7] states: *"theoretically this could be due to nolatrexed being worse than placebo, especially as no difference in PFS was demonstrated. Nolatrexed, however, belongs to a well known class of cytotoxic compounds (thymidylate synthase inhibitor) and the adverse event profile appears as expected and seemingly not worse than doxorubicin 60 mg/m^2 every three weeks. Thus the most likely explanation to the observed difference is that doxorubicin therapy also provides a survival benefit to patients with advanced HCC."*

4.2 Summary of submitted evidence

The submitted evidence for effectiveness was based on:

- the SHARP trial[10], a multicentre RCT that randomised 602 patients with advanced HCC to receive sorafenib (plus best supportive care) or placebo (plus best supportive care).
- Two other studies were drawn upon for supportive evidence only; these were
    - a multicentre RCT of sorafenib versus placebo conducted in 226 patients from a population with endemic hepatitis B (the Asia-Pacific study[11])
    - an uncontrolled open label uncontrolled study (Phase II trial, Abou-Alfa 2006[12]) with 137 predominantly European patients.

The diagram below summarises the time lines for the SHARP study.

4.2.1 Summary of results

The clinical effectiveness results in the submission were arranged as follows:

- effectiveness results from the SHARP trial for primary endpoints,
- results for secondary endpoints from SHARP (other than safety),
- subgroup analyses from SHARP,
- supporting evidence from the Asia-Pacific RCT,
- safety results,
- non-RCT evidence.

Each of these is considered in turn below with safety (adverse events) considered last.

Results from the submission about overall survival:

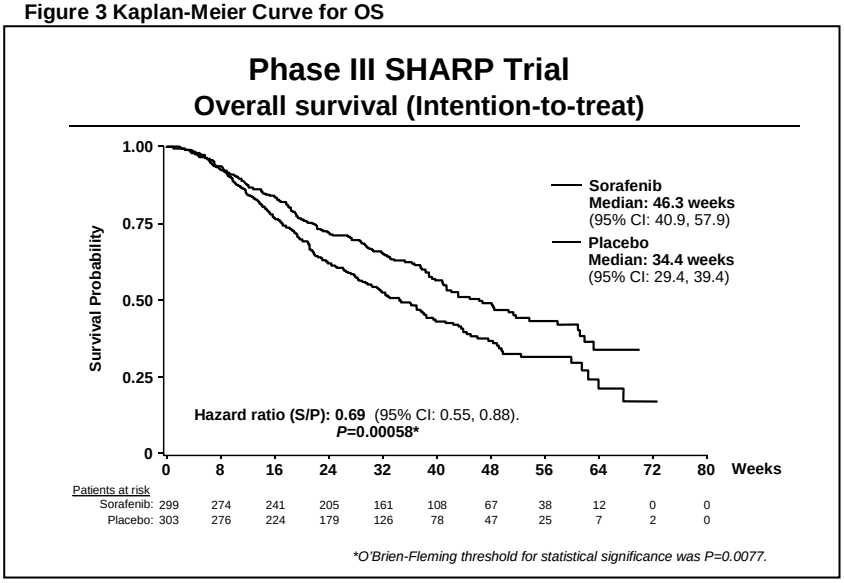
6.4 Primary endpoints – Overall Survival (OS),

The second interim analysis of efficacy data based on 321 survival events (178 events in the placebo arm, and 143 events in the sorafenib arm), demonstrated that sorafenib significantly prolonged overall survival compared with placebo. This led to early cessation of the trial.

Median overall survival was 34.4 weeks [95%CI 29.4, 39.4] in patients randomised to placebo and 46.3 weeks [95% CI 40.9, 57.9] in patients randomised to sorafenib (see figure 3). The stratified log-rank test had a 1-sided nominal p-value of 0.000583 and the estimated hazard ratio for survival (sorafenib over placebo) was 0.69 [95% CI 0.55, 0.87], representing a 30.7% reduction in hazard (risk of death) over placebo (or 44.3% increase in survival time over placebo) (*P* = 0.000583).

This represents a clinically meaningful and statistically significant improvement in overall survival attributable to sorafenib treatment, and also represents the first definitive demonstration of a meaningful survival benefit with any systemic treatment for HCC versus placebo.

Figure 3 Kaplan-Meier Curve for OS



Comment:

- The ERG requested and were granted access to the full SHARP trial report. The ERG checked median OS (sorafenib 46.3 weeks, placebo 34.4 weeks) and hazard ratio (0.69) from the submission against those in the SHARP publication and those in the SHARP trial report, as each used different times scales (days, weeks, months). The results correspond (see Appendix 12).

The ERG requested clarification for the submission statement that there was *"a 44.3% increase in survival time over placebo"*; the manufacturer's response is shown below:

> The percentage increase in survival was calculated using the Hazard Ratio, which takes into account the whole K-M survival curve by averaging the treatment effect across the curves. Formula: HR = hazard of sorafenib / hazard of placebo. Thus the relative improvement of sorafenib = 1/HR, i.e. 1/0.6931 =1.44 (i.e. prolongation in survival by 44%).
>
> (Note: Under the assumption of exponential survival distribution, the ratio of hazards is the inverse of that of the medians. Comparing the medians directly is considered the most intuitive, but less reliable since it only takes one point of the K-M curve).

- The use of hazard ratio (HR) to calculate a % increase in survival time is potentially misleading if the assumption of exponential survival distribution is not supported (see Spruance et al 2004[17]). HR informs on the likelihood that a random patient from one group will reach an end point before a patient selected randomly from the comparator group. When the exponential assumption is not supported HR may inflate (or deflate) the apparent survival benefit.

- The ERG extracted individual patient data for the placebo group and tested the exponential assumption (see Appendix 13). On the basis of this analysis the ERG consider that the assumption is not supported and that a 44% increase in survival benefit probably inflates the apparent benefit. A more reliable indicator in this case is the % increase in median survival, which for overall survival is 34.6%.

Overall survival continued

From the submission:

> The efficacy of sorafenib was also supported by the survival rates at 3, 6 and 12 months. The 3,6 and 12 month survival rates for sorafenib vs placebo are 86% vs 83%, 71% vs 61%, 44% vs 33% respectively (p=0.009).

Comment:

- These survival rates at 3 and 6 months correspond to those in the trial report (below):

[redacted]

- The 1 year survival rates in the SHARP publication corresponded to those in the submission (44% sorafenib, 33% placebo) and in the trial report.

Results from the submission about TTSP follow:

> TTSP, a co-primary outcome, was defined in the SHARP study as time from randomisation to the first documented symptomatic progression, based on patient-reported symptoms (PRO), deterioration to Eastern Cooperative Oncology Group (ECOG) performance status (PS) 4 or death.
>
> The primary analysis of the TTSP demonstrated no statistically significant difference between the sorafenib and placebo arms. Median TTSP was 18 weeks [95%CI 15, 21] for sorafenib-treated patients and 21.1 weeks [95%CI 18.4, 27.4] for placebo. The hazard ratio was 1.08 (0.88, 1.31) for sorafenib over placebo which is not statistically significant (p=0.77). These results, inconsistent with sorafenib's positive impact on overall survival, suggest that the FHSI-8 questionnaire may have been too sensitive to offer reliable information about the impact of treatment on symptomatic tumour progression. The FHSI8 questionnaire is a patient-oriented outcome instrument that may have been influenced by both the toxicity of the drug, as well as the effect of tumour symptom response. The lack of significant differences in FHSI8-TSP might reflect the impact of early reporting of sorafenib toxicities on FHSI8 scores.

Comment:

- These results correspond to those in both the trial report (median TTSP [redacted] days for sorafenib and [redacted] days for placebo patients) and the published account of the SHARP trial (median TTSP of 4.1 months sorafenib and 4.9 months placebo).

- The submission appears to argue that because sorafenib has benefit in terms of overall survival there is also probably an underlying benefit for TTSP but this has been masked by sorafenib toxicities. If such a putative

benefit is easily masked by sorafenib toxicities then it could be argued that it probably has little clinical relevance.

Results from the submission about TTP follow:

Time to Progression (TTP)

Analyses of TTP based on both independent (primary analysis) and investigator assessments demonstrated a statistically significant improvement in patients treated with sorafenib compared with placebo.

By independent assessment, the median TTP was longer for the sorafenib arm 24 weeks [95% CI 18, 30]) than the placebo group 12.3 weeks (95% CI 117, 17.1). The hazard ratio for TTP was 0.58 (95% CI: 0.45, 0.74) representing a 42.4% reduction in risk of progression (or 73.5% improvement in TTP) in patients treated with sorafenib compared with placebo (P=0.000007) (3,38) .

[redacted]

Table 8: Results of analyses of the TTP endpoint

	Independent Assessment (cut-off date 12th May 2006)		**Investigator assessment** (cut-off date 17th October 2006)		[redacted]	
	Sorafenib n=299	Placebo n=303	Sorafenib n=299	Placebo n=303	[redacted]	[redacted]
Number of progressions	107 (35.8%)	156 (51.5%)	181 (60.5%)	222 (73.3%)	[redacted]	[redacted]
Median TTP	24 weeks [95% CI 18, 30]	12.3 weeks [95% CI 11.7, 17.1]	17 weeks [95% CI 13,18]	11.9 weeks [95% CI 11.1, 12.4]	[redacted]	[redacted]
Hazard ratio (Sorafenib/placebo)	0.58 [95% CI 0.45,0.74] *p=0.000007*		0.6889 [95% CI 0.5634, 0.8423] *p=0.000130*		[redacted]	

Sensitivity analyses using scheduled radiological assessment dates rather than actual visit dates also concluded that sorafenib significantly prolongs TTP compared to placebo.

PFS was included in the SHARP study as a sensitivity analysis of TTP to evaluate the impact of deaths before progression. Based on independent tumour assessment and actual visit date, PFS rates at 4 months were 62% for sorafenib compared with 42% for placebo. These results support those reported for TTP.

Comment:

The submission presents three different analyses of TTP. One based on independent assessment of radiographs up to 12 May 2006 (263 progressions) (this analysis was presented in the SHARP publication[10]), and two analyses based on unpublished data referred to as investigator analysis

(403 progressions) and [redacted].

With regard to the latter analyses the trial report states the following:

[redacted]

- The SHARP study was conducted at 121 centres and presumably there were about this number of investigator assessors. The independent assessment was probably centralised and involved a smaller number of assessors. The ERG checked the TTP summary data presented in the submission against that in the SHARP publication and that in the SHARP trial report and found good correspondence (Appendix 14). The ERG were unable to find in the trial report a listing of individual patient TTP by investigator assessment.

- The independently assessed median TTP (published) is more favourable to sorafenib (difference in median TTPs; sorafenib – placebo = 11.1 weeks) than the investigator assessed median TTP (unpublished) (difference in median TTPs = 5.1 weeks).

- [redacted].

- There was a noticeable difference in the HR between independent and investigator analyses (0.58 vs. 0.6889). The ERG requested clarification regarding possible disagreement between the independent and investigator assessments. The manufacturer's response is given below:

> The difference was because of differences in assessments between investigators and the independent review as well as different data cutoff dates. There is no analysis of investigator assessed TTP using May 12, 2006, as the cutoff date.

- For the independent and investigator TTP analyses it is noticeable that although there is good agreement between the two analyses for median TTP for the placebo group (12.3 weeks vs 11.9 weeks) the disagreement for median TTP for the sorafenib group is substantial (24 weeks and 17 weeks).

Since only the published Kaplan-Meier curves for the independent analysis were available the ERG requested access to the Kaplan-Meier plots for the other TTP assessments. These were supplied by the manufacturer.

Below are shown Kaplan-Meier plots comparing the independent with investigator assessment analyses. A substantial separation of the curves for the sorafenib group is evident but this does not apply for the placebo plots. [redacted] (see Appendix 15).

- All other things being equal the more mature data from the investigator analysis would be accepted as the preferred analysis. However the evident asymmetry in disagreement between independent and investigator assessments (i.e. for sorafenib only) is of concern. For economic modelling the submission base case employs the investigator analysis while the independent [redacted] were not used. As TTP was identified as a main driver for the economic model the ERG considered it important that a sensitivity analysis should be undertaken using the independent analysis. In order to obtain lognormal fits for the independent TTP analysis it was necessary to extract individual patient data for the independent TTP

assessment from the SHARP trial report and use STATA software to generate lognormal fits to the resulting Kaplan-Meier plots. The resulting parameters were then used in the economic sensitivity analysis described elsewhere in this report. As a check on the accuracy of this process parameters from the independent analysis and from the trial report were compared and Kaplan-Meier plots superimposed; an apparently exact correspondence was observed (see Appendix 16).

Results from the submission about disease control rate follow:

Disease Control Rate (DCR)
In the SHARP study, DCR was higher in the sorafenib arm (43% [n=130]) than in the patients receiving placebo (32% [n=96]).

Comment:

- Disease control rate (DCR) is the percentage of patients with a response rated better than progressive disease (according to RECIST criteria) lasting at least 28 days from the first manifestation of that rating.
- The trial publication provides a p value of 0.002 for the comparison between groups.

Results from the submission about tumour response follow:

Tumour response

Table 9: Analyses of Tumour Response parameter per independent and investigator assessment

	Independent Assessment *(as of 12th May 2006)*		Investigator assessment	
	Sorafenib N=299 (%)	**Placebo n=303 (%)**	**Sorafenib n=299 (%)**	**Placebo n=303 (%)**
Number evaluated radiologically post-baseline	272	279	276	276
Best Response				
-complete response (CR)	0	0	0	0
-partial response (PR)	7 (2.34)	2 (0.66)	18 (6.02)	8 (2.64)
-stable disease (SD)	211 (70.57)	204 (67.33)	181 (60.54)	167 (55.12)
-progressive disease (PD)	54 (18.06)	73 (24.09)	77 (25.75)	101 (33.33)
-not assessable	27 (9.03)	24 (7.92)	23 (7.69)	27 (8.91)
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*[redacted]

No complete responses (CRs) were observed but there were 7 partial responses (PRs) (2%) in sorafenib-treated patients and 2 PRs (1%) in the placebo group. Stable disease was reported for 211 patients (71%) receiving sorafenib and 204 (67%) placebo-treated patients.

[redacted]

Comment:

- Differences between sorafenib and placebo groups are small for response outcomes with very low levels of complete and partial response in both groups (≤ 7% irrespective assessment by investigators or by independent assessors).
- The “tumour response” is the proportion of patients during treatment or within 30 days of stopping treatment that achieved a best response rated as complete response, partial response, stable disease or progressive disease (RECIST criteria; see Appendix 10).
- For investigator assessment of tumour response the submission and trial report results correspond.

- For independent assessment of tumour response the submission and published SHARP report correspond except the latter did not specify percentage with progressive disease or the percentage not assessable.

Results from the submission about health-related quality of life follow:

Health-related quality of life
Approximately 8% more placebo than sorafenib patients (19.6% versus 11.5%, respectively) achieved the 8-point MID for the FACT-Hep at Cycle 3, Day 1 or end of treatment visit

Comment:

- FACT-Hep is a self administered questionnaire yielding a total score between 0 and 180. The 19.6% and 11.5% results above refer to the proportion of individuals in each trial arm achieving at least an 8-point minimally important difference (MID) improvement in score.
- The SHARP trial report additionally presented p values as follows:

[redacted]

- These results were not presented in the SHARP publication. They tend to indicate better QOL in the placebo group than in the sorafenib group.

Further results from the submission

Health-related quality of life

[redacted]

Comment:

- [redacted].
- [redacted]:

[redacted]

[redacted]

[redacted]

- [redacted].

- [redacted]:

[redacted]

Submission results about overall survival for subgroups follow:

Analysis of overall survival by subgroup, using the patient stratification variables at randomisation, showed a consistent significant trend favouring the sorafenib arm for nearly all subgroups. The subgroup analyses were intended to be descriptive only. The study was not powered to assess differential patient response to treatment in subgroups, and no adjustments were made for multiple comparisons.

An exploratory multivariate analysis with the use of a Cox proportional-hazards model identified eight baseline characteristics that were prognostic indicators for overall survival: ECOG performance status, presence or absence of macroscopic vascular invasion, extent of tumour burden (defined as presence or absence of vascular invasion, extrahepatic spread, or both), Child–Pugh status, and median baseline levels of alpha-fetoprotein, albumin, alkaline phosphatase, and total bilirubin. After adjustment for these prognostic factors, the effect of sorafenib on overall survival remained significant (hazard ratio, 0.73; 95% CI, 0.58 to 0.92; P = 0.004). A prespecified subgroup analysis showed a consistent survival benefit for sorafenib over placebo in most of the subgroups analysed:

Table 10: Subgroup analysis SHARP study

Subgroup	Median OS (months)		Hazard Ratio (95% CI)
	Sorafenib	Placebo	
ECOG PS			
0	13.3	8.8	0.68 (0.50, 0.95)
1-2	8.9	5.6	0.71 (0.52, 0.96)
Macroscopic vascular invasion ***or*** extrahepatic spread or both	8.9	6.7	0.77 (0.60, 0.99)
No tumour burden	14.5	10.2	0.52 (0.32, 0.85)
[redacted]	[redacted]	[redacted]	0.68 (0.49, 0.93)
[redacted]	[redacted]	[redacted]	0.74 (0.54, 1.00)
Alcohol-related HCC	10.32	7.99	0.55 (0.39, 0.77)
Baseline Transaminase levels			0.76 (0.50, 1.16)
Normal ALT/AST (<1.8 x ULN)	13	9	
Mild ALT/AST (≥1.8 to ≤3 x ULN)	11	8	
Moderate ALT/AST (>3 x ULN)	8	5.5	NR
			NR
Hepatitis C	14	7.9	NR
[redacted]	[redacted]	[redacted]	0.58 (0.37, 0.91)
[redacted]			[redacted]
			[redacted]

Comment:

- Hazard ratios for several subgroups were already published and these correspond to the values in the submission. The remaining subgroup data

correspond to that in the trial report (which presents additional exploratory results for several other small subgroups).

- The submission's claim that sorafenib *"showed a consistent survival benefit for … over placebo in most of the subgroups analysed"* is clearly supported.

- The most relevant subgroup for the decision problem, namely patients recruited with advanced disease and Child-Pugh grade B liver function has not been examined because of a lack of sufficient patient numbers in the SHARP trial.

- The submission presented a Kaplan-Meier plot demonstrating improved overall survival with sorafenib for hepatitis virus C positive patients in SHARP; this is shown in Appendix 17.

The supporting RCT data from the Asia-Pacific study

The supporting data from the Asia-Pacific study presented in the submission is shown in Appendix 18. The results presented in the submission corresponded to those in the Asia-Pacific publication. The ERG requested the trial report for the Asia-Pacific study but this was not made available. The data considered below is as found in the 2009 publication.[11]

Differences between the trial populations in SHARP and the Asia-Pacific study included:

- ethnicity of the participants (patients from China, Korea and Taiwan in the Asia-Pacific study but predominantly from Europe in SHARP)

- aetiology of HCC (hepatitis B virus 73% in the Asia-Pacific study and 30% in SHARP)

- prognosis: placebo patients in the Asia-Pacific study had median survival of 18.2 weeks, those in SHARP a median survival of 34 weeks.

This may be partly explained by the poorer average Eastern Cooperative Oncology Group (ECOG) performance status (Appendix 19) and poorer average BCLC stage rating in the Asia-Pacific trial.

For ease of comparison the ERG have tabulated the main results in the SHARP and Asia-Pacific studies as shown below:

		Sorafenib	Placebo	Within-trial difference
POPULATION				
Number	SHARP	303	299	
	Asia-Pacific	150	76	
ECOG performance 0	SHARP	54%	54%	0%
	Asia-Pacific	25%	25%	0%
ECOG performance 1	SHARP	38%	39%	1%
	Asia-Pacific	69%	67%	2%
ECOG performance 2	SHARP	8%	7%	0%
	Asia-Pacific	5%	5%	0%
BCLC stage B*	SHARP	18%	17%	1%
	Asia-Pacific	5%	4%	1%
BCLC stage C	SHARP	82%	83%	1%
	Asia-Pacific	95%	96%	1%
CHILD-PUGH grade A	SHARP	95%	98%	3%
	Asia-Pacific	97%	97%	0%
CHILD-PUGH grade B	SHARP	5%	2%	3%
	Asia-Pacific	3%	3%	0%
OVERALL SURVIVAL				
Median (wks)	SHARP	46.3	34.4	11.9
	Asia-Pacific	28.2	18.2	10.0
At 6 months(%)	SHARP	71%	61%	10%
	Asia-Pacific	53.3%	36.7%	16.6%
TTSP				
Median (wks)	SHARP	18	21.1	-3.1
	Asia-Pacific	15.2	14.8	0.4
TTP**				
Median (wks)	SHARP	17.0	11.9	5.1
	Asia-Pacific	12.2	6.1	6.1
Disease Control Rate %				
	SHARP	43%	32%	11%
	Asia-Pacific	35%	16%	19%
Tumour Response**				
Complete response	SHARP	0%	0%	0%
	Asia-Pacific	0%	0%	0%
Partial response	SHARP	6.0%	2.6%	3.4%
	Asia-Pacific	3.3%	1.3%	2.0%
Stable disease	SHARP	60.5%	55.1%	11%
	Asia-Pacific	54.0%	27.6%	26.4%
Progressive disease	SHARP	25.8%	33.3%	-7.5%
	Asia-Pacific	30.7%	54%	-23.2%
Not assessable	SHARP	7.7%	8.9%	
	Asian Pacific	12.0%	17.1%	
Median duration of treatment (months)				
	SHARP	5.3	4.3	1
	Asia-Pacific	NR	NR	

* For the Asia-Pacific study calculated by difference: 100% - BCLC class C. ** For SHARP the results are for investigator assessment. It was unclear from the published Asia-Pacific study publication if assessment was done by independent assessors or by investigators.

Comment:

- There was good agreement between trials for the outcomes listed.

- The absolute gain in median overall survival and in median TTP by the sorafenib group relative to placebo was very similar in both trials. The increase in median overall survival in the Asia-Pacific study was 10 weeks (a 55% improvement on the 18 weeks median survival in the placebo group), similar to the 11.9 weeks in SHARP. The small number of patients in the placebo group (n = 76) means that the survival analysis for this group is associated with greater uncertainty than in the SHARP study. In the Kaplan-Meier plot for the placebo group [copyright protected] a pronounced kink can be observed that greatly influences the estimate for median survival. The gain calculated using HR (under the assumption of exponential survival distribution) was 47%. As the ERG did not have access to individual patient data in the Asia-Pacific study it was not possible to check the validity of the exponential assumption.

- In the Asia-Pacific trial the median TTP for the sorafenib group was extended by 6.1 weeks relative to placebo (an improvement of 50%), similar to the 5.1 weeks in SHARP. The gain calculated using HR (under the assumption of exponential survival distribution) was 76%. As ERG did not have access to individual patient data it was not possible to check the validity of the exponential assumption.

- With regard to QoL (FACT-HP) in the Asia-Pacific publication the following statement was found *" scores with the FACT-HP questionnaire showed no difference in quality of life between groups (data not shown)"*. No detailed results for QoL were presented in either SHARP or Asia-Pacific publications.

- Neither study included sufficient Child-Pugh grade B patients for a fruitful subgroup analysis of sorafenib benefit for these patients.

The supporting non-RCT evidence in the submission follows:

6.8.4 Results of the relevant non- RCTs
Independent assessment of responses identified no CRs, 3 PRs, 8 MRs and 46 patients with stabilisation of disease. Duration of the 3 PRs ranged from 12 to 14.5 months.

Table 13: Results of primary and secondary endpoints from the phase II uncontrolled study

Endpoint	**ITT analysis (n=137)**	
Response	**Independent assessment**:	**Investigator assessment:**
CR	0	0
PR	3 (2.2%)	8 (5.8%)
MR	8 (5.8%)	6 (4.4%)
SD	46 (33.6%)	50 (36.5%)
Median TTP	5.5 months	4.2 months
Median OS	Not evaluable	9.2 months

Time to response, PFS, and duration of stable disease were not reported in the publication but have been sourced from the study report. Of the subjects who had confirmed PR, time to response ranged from 49 days (approximately 1.6 months) to 296 days (approximately 9.9 months). Median time to response was 191 days. Median PFS (based on investigator assessment) was 123 days (95% CI: 108, 148). Median duration of stable disease was 126 days (95% CI: 112, 168). Results from the phase II study are consistent with those in the phase III study.

Comment:

- These are the results of the Abou-Alfa uncontrolled open label study. Other than those that are commercial in confidence they correspond to those in the 2006 publication[12] of this study.

- The response rates (WHO criteria) refer to a minimum of 16 weeks duration of response.

Further results from this study that provide effectiveness information about Child-Pugh grade B patients were published in abstract in 2008[15] (reference 37 in the submission) but the results were not presented in the submission. The ERG therefore extracted the results from the abstract and has summarised them below together with the sorafenib group results from the SHARP trial in which 97% of patients were Child-Pugh grade A.

	UNCONTROLLED STUDY		SHARP
	CHILD-PUGH B patients (N=38)	**CHILD-PUGH A patients (N=98)**	**97% Child-Pugh A (N=299)**
Median overall survival *(95%CI)*	14 weeks *(11.6 to 25.7)*	41 weeks *(36.6 to 63.6)*	46.3 weeks (*40.9 to 57.9)*
Median time to progression *(95%CI)*	13 weeks* *(9 to 18)*	21 weeks* *(16 to 25)*	24 weeks** *(18 to 30)*
Stable disease (≥ 4 months)	26%	49%	RD
Adverse Events	97%	97%	98%
Serious Adverse Events	68%	52%	51.5%
Fatigue	37%	41%	22%†
Hand Foot Skin Reaction	13%	30%	21%†
Diarrhoea	47%	59%	39%†
Bilirubin Increase	40%	18%	8.8%
Ascites	18%	11%	RD
Encephalopathy	22%	2%	RD
Median length of Therapy	12.9 weeks	24.9weeks	23 weeks††
Dose Reductions	21%	31%	32%

* Unclear if independent or investigator assessment. ** Independent assessment. † treatment related AE. RD = reported differently (e.g. RECIST not WHO criteria). †† The median duration of treatment up to the cut-off date (17 Oct 2006), [18.6 weeks in the placebo group and 23 weeks in the sorafenib group].

Comment:

- The results for Child-Pugh A patients are similar in the uncontrolled open label study and in SHARP.
- The results for overall survival and TTP indicate that the effectiveness of sorafenib for patients with advanced HCC is likely to be better for those with Child-Pugh cirrhosis grade A than for those with cirrhosis level B.
- These results imply that estimates of average sorafenib effectiveness for the population defined in the decision problem are likely to be exaggerated if they are based solely on results from the SHARP study with its predominantly Child-Pugh grade A population.

The ERG requested clarification regarding the effectiveness of sorafenib for Child-Pugh grade B relative to grade A patients. The manufacturer's response is shown in Appendix 20 followed by the ERG comments on the response.

.

4.2.2 Adverse event results

In the submission these were derived mainly from the SHARP trial. Adverse events (AE) were monitored in SHARP using the National Cancer Institute Common Toxicity Criteria (NCI-CTC) version 3. The median duration of treatment was 23 weeks for sorafenib and 18.6 weeks for placebo. The average daily dose was 710.5 mg for the sorafenib and 774.8 mg for placebo.

An overview of AE was presented in the SHARP trial report and this is shown below.

[redacted]

Comment:

- [redacted].

A breakdown of treatment-related AE reported for at least 5% of patients in either arm was tabulated in the submission as shown below. This table also includes data from the uncontrolled open label study (Abou-Alfa 2006); in this study the NCI-CTC version 2 was used for monitoring events.

Table 11: **Incidence of treatment-related adverse events reported for at least 5% of patients in either treatment arm in the SHARP study (3,28)**

Adverse Event NCI-CTCAE version 3.0 Category / Term	CTC GRADE	Placebo (n=302) n(%)	Sorafenib (n=297) n(%)	Phase II study* (n=137) n(%)
Any Event	ALL	158 (52%)	236 (80%)	NR
Cardiac General				
Hypertension	3	2 (1%)	5 (2%)	NR
	ALL	6 (2 %)	15 (5%)	
Constitutional Symptoms				
Fatigue	3	10 (3%)	9 (3%)	13 (9.5%)
	4	1 (<1%)	2 (1%)	0 (0%)
	ALL	47 (16%)	64 (22%)	41 (29.9%)
Weight Loss				
	3	0 (0%)	5 (2%)	NR
	ALL	2 (1%)	28 (9%)	
Gastrointestinal				
Anorexia	3	2 (1%)	1 (<1%)	2 (1.5%)
	ALL	10 (3%)	41 (14%)	19 (13.9%)
Diarrhoea	3	5 (2%)	25 (8%)	11 (8%)
	ALL	34 (11%)	116 (39%)	59 (43.1%)
Nausea	3	3 (1.0%)	1 (<1%)	0 (0%)
	ALL	23 (8%)	33 (11%)	22 (16.1%)
Vomiting	3	2 (1%)	3 (1%)	0 (0%)
	ALL	8 (3%)	15 (5%)	14 (10.2%)
Stomatitis	3			1 (0.7%)
	ALL	NR	NR	15 (10.9%)
Pain				
Pain, Abdomen NOS	3	2 (1%)	6 (2%)	NR
	ALL	9 (3%)	24 (8%)	
Pulmonary / Upper Respiratory				
Voice Changes	ALL	2 (1%)	17 (6%)	NR
Dermatology / Skin				
Alopecia	ALL	5 (2%)	41 (14%)	14 (10.2%)
Dry Skin	ALL	12 (4%)	24 (8%)	NR
Hand-Foot Skin Reaction	3	1 (<1%)	23 (8%)	7 (5.1%)
	ALL	8 (3%)	63 (21%)	42 (30.7%)
Dermatology – other (specify)	3	0 (0%)	3 (1%)	
	ALL	2 (1%)	16 (5%)	NR
Pruritus	3	1 (<1%)	0 (0%)	
	ALL	22 (7%)	25 (8%)	NR
Rash / Desquamation	3	0 (0%)	3 (1%)	1 (0.7%)
	ALL	34 (11%)	47 (16%)	23 (16.8%)

*Publication only reports on grade 3 and 4 drug-related adverse events in ≥10% of all 137 patients. NR=not reported

Comment:

- AE were more common in the sorafenib treated patients than in patients who received placebo (80% versus 52%).
- The most common adverse events that occurred at a higher incidence in patients receiving sorafenib than in those receiving placebo were

hypertension and dermatologic and gastrointestinal problems that included: diarrhoea, anorexia, weight loss, hand and foot syndrome, dry skin, and alopecia.

- A meta-analysis [18] of event rates for all-grade hand and foot syndrome in cancer patients treated with sorafenib yielded a pooled incidence estimate of 38% (95% CI 24% to 45%), a somewhat higher rate than in SHARP (21%). The pooled estimate for high grade hand and foot syndrome was 8.9%, similar to that observed in SHARP (8%).

- A meta-analysis[19] of event rates for all-grade hypertension in cancer patients treated with sorafenib yielded a pooled incidence estimate of 23% (95% CI 16% to 39%), a higher rate than in SHARP (5%). The pooled estimate for high grade hypertension was 5.7% again higher than that observed in SHARP (2%).

- No studies directly compared the safety profile of sorafenib with that of a systemic cytotoxic anti-cancer agent (such as doxorubicin). A randomised trial of doxorubicin plus sorafenib versus doxorubicin[13] concluded that side effects of the two drugs appeared to be additive.

A summary of adverse events leading to treatment discontinuation in SHARP was tabulated in the SHARP trial report and major items from this table are shown below.

Information about laboratory abnormalities in the submission is provided below.

Grade 3 or 4 laboratory abnormalities occurred at similar frequencies in the two study groups, placebo group, P<0.001) with the exception of grade 3 hypophosphatemia (11% in the sorafenib group vs. 2% in the placebo group, P<0.001) and grade 3 or 4 thrombocytopenia (4% in the sorafenib group vs. <1% in the placebo group, p = 0.006).

Comment:

- The clinical significance of grade 3 hypophosphatemia for these patients is uncertain.

4.2.3 Critique of submitted evidence syntheses

No meta-analysis and no indirect or mixed comparisons were undertaken. The differences between the SHARP population and the patients in the Asia-Pacific RCT represent a reasonable justification for not pooling results. The inclusion of only studies that used sorafenib as a single agent and the choice of BSC as comparator reduced the possibility of undertaking indirect comparisons. However, the current paucity of fully published studies of sorafenib in HCC means that in practice indirect comparisons would be unlikely to provide useful information. Whether combined systemic therapy (e.g. sorafenib + doxorubicin) is more effective than single agent therapy remains an open and important question in the treatment of advanced HCC.

The submission was based on the premice that "*the phase III SHARP study is the largest and most relevant data source for the decision problem being addressed*". The ERG did not find errors when comparing SHARP results detailed in the submission with those available in the full SHARP trial report or in the SHARP publication. However there was a small amount of relevant QoL information in the trial report which was incompletely represented in the submission and there was some information in the submission relating to investigator assessment of TTP that could not be easily verified from information in the trial report because of a lack of listing of individual patient data.

There was some doubt about the use of HR to quantify survival improvement due to sorafenib. In the submission this was done for overall survival in SHARP and for overall survival and TTP in the Asia-Pacific study and was emphasised in the executive summary. Although this approach is valid under the assumption of an exponential distribution for survival the submission did not attempt to verify this assumption. The ERG tested this assumption for OS in the SHARP placebo group and could find little support. The ERG were unable to pursue this question for the Asia-Pacific study because appropriate data was not available.

The submission used results from the Asia-Pacific RCT to support the evidence base derived from SHARP. This appears to be a reasonable approach with the proviso that outcome selection bias should be avoided. The only instance where such bias may have operated was in the analysis of the FACT-Hep QoL results which were only briefly reported in the Asia-Pacific publication. The ERG requested the trial report for this study but this was not made available.

Similarly the submission used results from the uncontrolled study of Abou-Alfa 2006 to support the evidence base derived from SHARP. This study did not satisfy the submission's study inclusion criteria. The manufacturer's response to a query for clarification on this issue was as follows:

> *"the NICE submission form requires details of relevant non-randomised studies and as such, the phase II study was included because it describes the use of sorafenib in advanced inoperable HCC, in a population of patients not dissimilar to the anticipated UK treatable population, and at the licenced dosage being assessed within the submission".*

The 2008 abstract[15] of this study was included in the list of relevant papers in the submission but data about response to sorafenib in Child-Pugh grade B patients was not presented.

The submission presented three different analyses for TTP in the SHARP RCT; these were an independent analysis, an investigator analysis [redacted]. There were clear differences between these analyses but the submission did not provide an exploration for any underlying cause of such discrepancies. In response to ERG request for clarification the manufacturer's response was as follows:

> *The difference was because of differences in assessments between investigators and the independent review as well as different data cutoff dates. There is no analysis of investigator assessed TTP using May 12, 2006, as the cutoff date.*

The difference between investigator and independent analyses appeared to only affect the sorafenib group and this may be a cause for concern for the exclusive use of the investigator analysis in the economic modelling. Relative to the independent analysis the investigator analysis points to a poorer

effectiveness for sorafenib (less extension in TTP). In contrast, from an economic perspective, the investigator analysis generates a greater proportion of live patients in progressive state who incur low costs and therefore push the incremental cost effectiveness ratio in favour of sorafenib.

Both SHARP and Asia-Pacific studies almost exclusively recruited patients with good liver function status (Child-Pugh grade A). Limited evidence from the uncontrolled open label study[15] suggest that sorafenib is less effective for Child-Pugh grade B patients than for grade A patients. This data was not presented or commented upon in the submission.

Important issues for the decision problem are: what proportion of patients eligible for sorafenib intervention would have Child-Pugh liver function grade worse than grade A and what is the effectiveness of sorafenib for Child-Pugh grade B patients?

4.2.4 Summary

In summary the evidence submitted demonstrates that relative to BSC (placebo) sorafenib:

- reduces risk of death by 31% (HR = 0.69)
- prolongs median overall survival by 83 days (11.9 weeks, a 34% improvement)
- reduces the risk of radiologic disease progression (by 31% or 42% depending on analysis of investigator or independent assessments respectively according to RECIST criteria)
- prolongs median time to radiological progression (by 5.1 weeks or 11.9 weeks depending on analysis of investigator or independent assessments respectively)
- has no effect on risk of symptomatic progression or on the median time to symptomatic progression

- has no effect on quality of life (as determined using the FACT-Hep questionnaire)
- generates less than 7% complete or partial best-tumour-responses
- generates slightly more stable-disease best-responses (28 day duration) than placebo (71% versus 67%%, or 61% versus 55%, according to independent or investigator assessment respectively)
- increases the incidence of gastrointestinal and dermatologic adverse events (diarrhoea, anorexia, weight loss, hand and foot syndrome, rash, alopecia) and of hypertension
- produces few excess adverse events that necessitated withdrawal from treatment (35% placebo versus 32% sorafenib).
- induces an excess of temporary dose reductions (32% in the sorafenib group versus 13% for placebo)

The results summarised above derive from the SHARP RCT which investigated a population with predominantly European ethnicity (broadly comparable to that in the UK) that consisted of 83% of patients with advanced HCC (according to BCLC stage C) and 97% with cirrhosis grade A by Child-Pugh criteria.

Evidence from the Asia-Pacific study undertaken in a population with different ethnicity and HCC aetiology but similar BCLC stage and Child-Pugh grading to that in SHARP confirmed the conclusions from SHARP in respect of improvement in overall survival and extending time to radiological disease progression, and lack of effect on time to symptomatic progression, quality of life and low levels of complete and partial tumour responses.

The uncontrolled open label study provided some evidence that sorafenib is as safe for Child-Pugh grade B patients as it is for Child-Pugh grade A patients. This study indicated that Child-Pugh grade B patients may respond less well to sorafenib than Child-Pugh A patients in terms of stable disease

response (for > 4 months), time to progression, and overall survival, however these results need to be confirmed preferably in a randomised study.

Uncertainties about effectiveness of sorafenib versus placebo that remain are:

- Which of the three analyses of time to radiological progression (independent, investigator [redacted]) is the more reliable?
- What is the effectiveness of sorafenib in Child-Pugh B patients?
- What proportion of advanced HCC patients that, according to clinical judgement would be eligible for sorafenib treatment because of unsuitability or lack of success of loco-regional and surgical therapies, would be classified as Child-Pugh grade A and Child-Pugh grade B, and therefore what is the effectiveness of sorafenib in the overall indicated population?

5 ECONOMIC EVALUATION

5.1 Overview of manufacturer's economic evaluation

The manufacturers provided a Markov model with a lifetime time horizon. This time horizon was assumed to cover up to an additional 14 years of life from a patient population with an average starting age of 67. This time horizon is appropriate to the decision problem and is in keeping with the NICE reference case. Time horizons of two, five and ten years are also used in the sensitivity analyses.

The model included four states; non progressive advanced disease, progressive disease, best-supportive care and death (see figure 5 from the submission, reproduced below). These health states are considered to be relevant and clinically appropriate to the decision problem. Although with few health states in the model some precision may be lost in estimating the costs and benefits of each state, the complex nature of the problem and the paucity of data suggest that any additional health states would have been of little benefit, adding a veneer of accuracy while retaining significant but undisclosed uncertainties. Each model cycle lasts for one month.

The key structural and data assumptions underlying the economic model are as follows (from page 61 of the submission):

- The phase III SHARP study is the largest and most relevant data source for the decision problem being addressed

- Best-supportive care is used as the comparator to sorafenib

- Time to progression (TTP) was based on the trial investigators' assessment, as this was believed to be the best representation of clinical practice

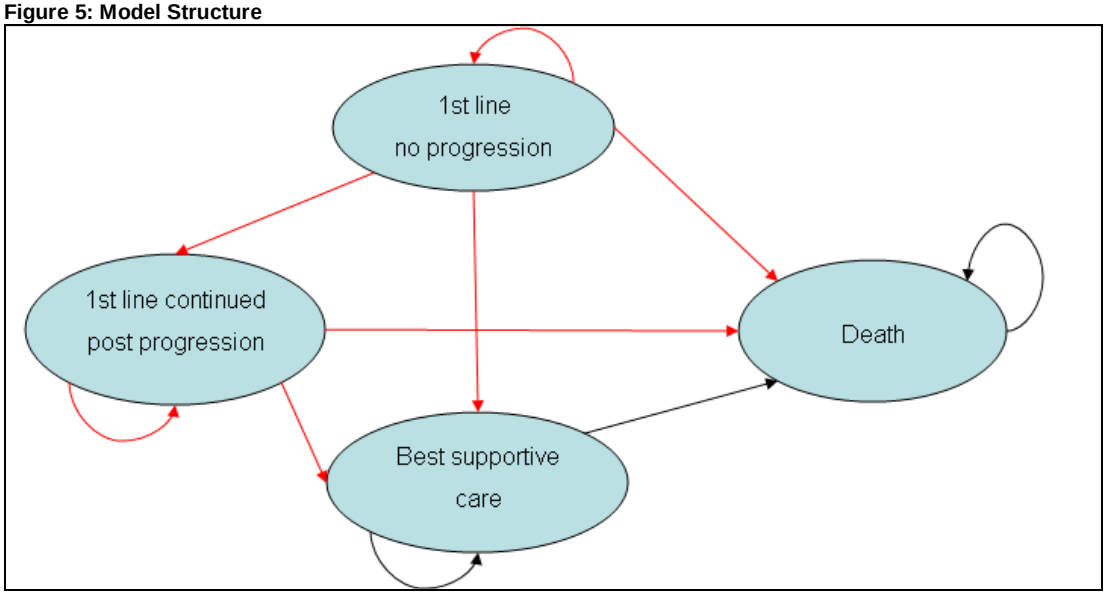
- The time-to-progression and overall survival observed in the treatment and the placebo group over 72 weeks can be extrapolated to the desired time horizons using a lognormal distribution;

- The rate of AEs is assumed to be constant over the time horizon; and

- The disutilities due to AEs are additive, i.e. can be estimated by subtracting the utility of a given health state with an AE from the utility of that health state without any AE

The model schematic below is Figure 5 from the submission.

Figure 5: Model Structure



Key parameters in the model include:

- disease progression, as measured by the time to progression and the overall survival (table 14 from the submission)
- The adverse events rates (table 16 from the submission)
- The utilities inputs (table 15 from the submission)
- The costs inputs (tables 17 and 18 from the submission)
- Probabilistic sensitivity analysis was undertaken and details were provided in tables 23-27 from the submission
- The manufacturers excluded ‘end of life care’ from the model as a separate state given that every patient – regardless of the comparator – had to pass through this state and there was no differentiation by prior treatment.

•

5.1.1 Natural history

From the submission:

> Hepatocellular carcinoma (HCC) is the dominant form of primary liver cancer, accounting for about 80-90% of liver cancer cases. It is the third most common cause of cancer-related death worldwide, and most prevalent in Asia and Africa. In accordance with NICE's 'End of Life' policy criteria, HCC affects a small population of patients in the UK with about 2751 new cases of liver cancer diagnosed in England and Wales in 2005 (ref 1 of the STA), this is approximately 2340 cases of HCC, a proportion of which will be eligible for sorafenib.
>
> The primary risk factor for HCC is cirrhosis (the replacement of normal liver cells by fibrous scar tissue, with patches of tissue regeneration). Whilst cirrhosis can have many causes, it is most commonly due to Hepatitis B; Hepatitis C; and alcohol. Unlike most other cancers, the incidence of HCC is rising in Western countries, probably as a direct result of the Hepatitis C virus (HCV) epidemic. In the UK the incidence trend has increased from 2.5 to 3.9 per 100,000 persons between 1993 and 2005.
>
> Less than 30% of patients are diagnosed in the early stages where liver tumours are considered more amenable to curative resection or transplantation. Some patients may be suitable for "loco-regional" treatments: ablation (radiofrequency ablation (RFA); percutaneous ethanol injection (PEI) or cryosurgery); (chemo)embolisation, and radiotherapy. For patients where surgical or loco-regional treatments have failed or are unsuitable (approximately 25-35% of HCC patients (ref 2 of the STA)), systemic therapy is the only active treatment option.
>
> Doxorubicin is used in a minority of patients, but low overall response rates (10-15%) and the risks associated with its use often outweigh any short-term benefits, and clinicians usually opt for a best supportive care (BSC) approach instead. Therefore, within the present therapeutic landscape, the prognosis for patients with advanced HCC is bleak, with 5-year survival rates of <5% (ref 18 of the STA). Consequently, there is a compelling clinical need for effective treatments in order to improve the outlook for these patients.
>
> Due to sorafenib being shown to prolong survival in this patient group, several guidelines and review papers, including the revised UK guidelines now include sorafenib as the standard of care systemic therapy for patients with advanced HCC for whom no potential curative option is available.

Comment:

- The natural history of HCC could have been more comprehensively described.
- In requests for clarification from the manufacturer, the ERG raised the issue of whether or not best-supportive care was the most appropriate comparator for use in the economic evaluation. The manufacturer replied that "*HCC as a specific indication i[s] not included in Therapeutic indications section 4.1 of the SmPC*" related to those companies manufacturing doxorubicin, although

in some SmPCs use in solid tumours is indicated. As stated in the submission, doxorubicin is currently used in practice by clinicians in the UK. Although the manufacturer's claim that most clinicians opt for best-supportive care, this claim is not supported by any empirical evidence. It remains unclear to the ERG what proportion of patients in the UK may be treated with doxorubicin and why this was not considered to be a valid comparison for the purposes of the economic evidence submission.

5.1.2 Treatment effectiveness within the submission

Treatment effectiveness of relevance to the economic model (overall survival and time to treatment progression) is dealt with in previous sections of this report.

According to the submission, effectiveness of sorafenib as a treatment is observed in patients with advanced HCC who have failed or are unsuitable for surgical or locoregional therapies. According to the submission, the proportion of diagnosed HCC patients that would be eligible for treatment with sorafenib was 25-35 %.

The efficacy inputs relied on log-normal distribution fits to observed overall survival and time to treatment progression. The table below from the submission details these parameters.

Table 14: Efficacy Inputs (lognormal distribution parameters from the SHARP trial) (28)

	TTP		OS	
	Mu	**Sigma**	**Mu**	**Sigma**
Total population (base case)				
Sorafenib	4.822	0.983	5.791	1.147
BSC	4.513	0.804	5.465	1.019
Age =>65				
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]
Child Pugh A				
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]
TNM Stage I-III				
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]
BCLC stage B				

	TTP		OS	
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]
BCLC stage C				
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]
Hepatitis C from lab				
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]
With macrovascular invasion				
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]
Without macrovascular invasion				
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]
No extra hepatic spread				
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]
No tumour burden				
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]

- With respect to disease progression, it is unclear how the lognormal distribution inputs translate across to the probabilities used within the Markov model. The distinction between TTP and time to symptomatic progression (TTSP) may also be important. In the SHARP trial there was a marked difference in TTP between the sorafenib and placebo arms, but little difference for TTSP.

- The submission outlines reasons why TTSP may not be an appropriate clinical measure (reproduced below). In light of this, the ERG consider TTP to be the appropriate clinical input into the model.

> The results for TTSP are not in line with the reported survival, TTP and other benefits of sorafenib and it is possible that the FHSI-8 tool may have been inadequate to discern treatment-related side effects or effects of underlying liver cirrhosis from progression of HCC. Indeed, an expert panel convened by the American Association for the Study of Liver Diseases (AASLD) concluded that this endpoint is particularly hard to measure in cirrhotic patients with cancer, in whom the impairment of quality of life may be a consequence of the natural history of cirrhosis and not tumour progression 34 and suggested that 'Time to Symptomatic Progression' as an endpoint in HCC studies is not 'ready for clinical research at this point.

5.1.3 Health related quality of life

From the submission:

Health effects were expressed as QALYs and in terms of life years gained, as quality of life is likely to be affected by the symptoms of underlying liver diseases, including liver failure, irrespective of whether the tumour has stabilised or regressed. As a result, it is not possible to demonstrate the impact of treatments in advanced HCC on quality of life, and no robust and reliable utility data is available that separates out the effect of the primary liver cancer from the underlying liver disease causes.

Utility scores are derived using an algorithm developed by Dobrez et al. (2007) (ref 63 of the STA): The health-related quality of life (HRQL) as measured by the FACT-G part of Functional Assessment of Cancer Therapy— Hepatobiliary (FACT-HEP) instrument was mapped to time trade-off (TTO) utilities.

Utility scores were obtained for first-line treatment with sorafenib and BSC before progression, and treatment with sorafenib and BSC after progression.

The model accounts for the disutility of treatment resulting from selected grade 3 or 4 adverse effects.

Table 21: Utility Scores Derived from the Mapping Study

Before progression		
Sorafenib	Mean	0.69
	SD	0.12
BSC	Mean	0.69
	SD	0.12
After progression		
Sorafenib	Mean	0.71
	SD	0.13
BSC	Mean	0.71
after progression	SD	0.13
Disutility for AEs		
Sorafenib	Mean	-0.012
	SD	0.00
BSC	Mean	-0.012
	SD	0.00

Comment:

- Quality adjusted life years (QALYs) and life years (LYs) gained are presented for the base case analysis (Table 28, p.79 of the submission). The model indicates that the improved clinical outcomes with sorafenib result in estimated discounted QALYs of 1.08 versus 0.72 with BSC whereas the estimated LYs gained is 1.54 for sorafenib compared to 1.03 for BSC. In addition to the base case analysis, a series of subgroups were considered - see section 5.3 for full details.

- The submission considers that "*HCC is a unique condition which poses methodological issues when evaluating the impact of new treatments on health related quality of life. Patients with HCC are heterogeneous, with a diverse range of underlying causes of cirrhosis. As a result it is particularly difficult to disentangle the effect of the advanced HCC, underlying live disease and interventions on quality of life*". It is argued that as a result the most appropriate metric for measuring benefit in the economic evaluation may be the use of (non-quality adjusted) life years gained.

- Using LYs in place of QALYs represents a significant departure from the NICE reference case. The submission argues that the condition is complex with substantial co-morbidities. Many conditions typically affecting patients in the age group modelled here will experience a range of co-morbidities

alongside the primary reason for treatment (see Tuominen et al[20] remarks about patients awaiting total joint replacement). For the economic model presented in the submission, information on the quality of life of patients was measured using the Functional Assessment of Cancer Therapy—Hepatobiliary (FACT-Hep) instrument. The manufacturer then mapped these responses using a previously published algorithm to obtain health state utility estimates for patients in the model (as discussed below). This mapping algorithm used the generic portion of the FACT instrument and did not include information gained from the HEP subset of questions. If the submission maintained that such detailed information was necessary a number of options would have been available to them in the context of both this submission and the SHARP study. They could for example of obtained health related quality of life information directly from patients using a validated measure such as the EQ-5D, or they could have developed a unique mapping algorithm using the Hep subscale as the source information.

Mapped estimates

From the submission:

> A search of the existing literature did not identify any relevant utility values for the model health states. A systematic review (Appendix 11) conducted by Bayer found 36 studies reporting utility weights for HCC. The utility values in the publications ranged between 0.10–0.95, and were mainly used in different subgroups of patients with hepatitis C or B or liver transplantation.

Comment:

- The submission contained a systematic review of the literature to identify sources for use in estimating the utility values for patients with HCC for use in the economic evaluation. The search terms used in the review are extensive and appropriate. Concurrent searches of EMBASE and Medline were undertaken. Other databases commonly searched for evidence sources in economic evaluations, such as the NHS Economic Evaluation Database and the Health Economic Evaluation Database were not

searched. The review identified 36 potential studies, although it was argued that none of these contained values appropriate for use in the population modelled here. Without re-reviewing all studies identified in the submission it is not possible to test the validity of this conclusion; such a review is beyond the scope of the ERG. The mapping approach taken in the submission is an acceptable practice according to the NICE reference case.

- Dobrez et al. (2007) developed an algorithm for mapping from the FACT-G questionnaire to a set of time trade off (TTO) utility values. The patients in the original study had received one of ten diagnoses – breast cancer, prostate cancer, colon cancer, non small cell lung cancer, head and neck cancer, non-Hodgkin's lymphoma, Hodgkin's, small cell lung cancer, other known cancer and unknown primary. No HCC specific sub-group was considered in the development of the algorithm. The authors selected questions from the FACT-G scale and used these in an algorithm based on a combination of a) the correlation between the responses to the item and b) TTO scores, together with a more subjective approach which aimed to include items from each domain in the instrument. An ordinary least squares approach was used to estimate the algorithm; this approach is commonly used in mapping studies.

- Although the Dobrez study is methodologically valid, it is questionable whether or not it is the most appropriate approach to estimating utility scores for this evaluation. The NICE methods guidance (2008) clearly states that preferences from the general population are preferred when estimating utility scores. The algorithm developed by Dobrez is based on preferences for a population with cancer. No justification is provided as to why it was not possible to provide utility scores based on the preferences of the general population. However, the utility scores used in the analysis showed no meaningful difference in the quality of life for the populations in the two different health states.

- In the submission, the pre-progression state does not include the disutilities from adverse events (a separate disutility was used). However, the post-progression state includes these disutilities, thus removing the need for

adverse events to be considered. The mean utility before disease progression was marginally lower (0.69) than the mean utility after disease progression (0.71). If we were to remove the effect of adverse events entirely, there will be a greater difference between before progression and after progression.

- One reason for this lack of face validity (better utility after progression) may be a possible error in the algorithm used to calculate utility values. In one table in Dobrez et al (2007) an answer at levels 0 and 1 on Question 3 is given a utility estimate that is 0.0431 lower than an answer at levels 2, 3 and 4. In a later equation provided by the Dobrez, and used in the submission, the utility value is 0.0431 higher for an answer at levels 0 and 1. Given that Dobrez is quite clear that 0 should refer to the worst health level, this appears to be incorrect. The consequences of this error may be to increase the estimate of health in more severe cases i.e. where disease progression has occurred. Given this concern, the utility estimates provided in the submission should be treated with caution. Although the ERG have identified this potential error, it is not possible within the scope of the STA process for additional analysis to try and correct it.

5.1.4 Resources and costs

Resource use and cost parameters were estimated from a wide range of primary and secondary sources. The tables from the submission specifying the resources used and their costs are provided in Appendix 21. The model includes costs for drug treatment (sorafenib) for HCC and the treatment costs for the different health states and AEs.

Reproduced below is a table from the submission which summarises the mean cost estimates per patient per cycle for a number of broad resource categories.

Table 22: Cost Estimates According to Type of Resource

Type of costs	Mean
Active treatment – routine care	
Hospitalisation	65
Medical staff visits	230
Lab tests	124
Radiological tests	61
Active treatment - after progression	
Hospitalisation	266
Medical staff visits	480
Lab tests	30
Radiological tests	78
BSC - first line	
Hospitalisation	151
Medical staff visits	225
Lab tests	124
Radiological tests	61
BSC – after progression	
Hospitalisation	386
Medical staff visits	364
Lab tests	30
Radiological tests	78
At progression - one off cost	
Hospitalisation	0
Medical staff visits	0
Lab tests	104
Radiological tests	134
End of life – one off cost	0

Comments:

- The estimates of resource use are not based on data collected alongside the clinical trial but rather on a survey of UK based clinicians. Using expert opinion as a primary source on such a wide range of resource use significantly increases the uncertainty associated with the model results. Although it is possible to include costs from commonly used sources within the model, if estimates of resource use are uncertain this will feed through into the overall results of the model.
- Uncertainty estimates around the costs are poorly reported. This increases the difficulty in establishing whether or not the uncertainty estimates of the overall model results can be considered reliable. Although the manufacturers state that gamma distributions have been fitted to the cost parameters in the model, it is not clear on what basis these estimates are made.

- The cost of adverse events is based on expert opinion. The questionnaire provided in Appendix 13 of the submission relates to the costs for HCC treatment only. As such, it is unclear on what basis the manufacturers give a higher weighted cost per cycle for adverse events under best supportive care. As this is not justified, it may be a source of bias in favour of sorafenib.
- During the clarifications stage the ERG sought further information about the treatment dosage. Sorafenib is supplied as tablets in packs of 112 200mg tablets with a recommended dose of 400mg taken twice daily (four tablets daily). The submission states that the mean daily dose is 710.5 mg/day. Consequently, costs are evaluated as regard to this latter dosage. The manufacturers replied that the average daily dosage will be 710.5 mg/day in accordance with the dose reductions and interruptions observed in the trial. These reductions occurred for a variety of reasons, including adverse events.
- The health consequences of these adverse events were included in the model, although only for common AEs. The health and direct cost consequences for AEs occurring in fewer than 10% of patients are not included in the model. This may mean that the impact of AE in reducing dosage is fully reflected in the model, but the impact of AE in increasing treatment costs and decreasing health utilities is not fully reflected. This would bias the model in favour of sorafenib and could not be explored in the ERG's sensitivity analyses.

5.1.5 Discounting

The economic model applied discounting to both costs and outcomes at the annual discount rate of 3.5%, in line with current guidance from NICE. Sensitivity an analysis is also conducted using alternative discount rates of 0% and 6%.

5.1.6 Sensitivity analyses

From the submission

Both probabilistic and one-way sensitivity analyses were explored in the model.

ONE WAY SENSITIVITY ANALYSES

DISCOUNTING

- differences in discount rates (cost and benefits at 0%; costs 6%, benefits 0%; costs 0%, benefits 6%)

COSTS

- zero drug costs

- allowing only differences in drug costs

- costs from Renal Cell Carcinoma (RCC) appraisal assessment report

- Inclusion of PSS costs from the resource use survey

- Inclusion of end of life costs from the literature (Source unclear)

UTILITIES

- No AE disutility

- AE disutility of 0.05

- AE disutility of 0.20

- Usiing separate utility values for adverse events in sorafenib and BSC. (Source unclear)

- Utility of 0.41 for all health states based on Levy et al, 2008

- Utility values from RCC assessment report

PROGRESSION

- On progression, no patients on sorafenib continue treatment (versus 7.7% treated in main study)

- On progression, all patients on sorafenib continue treatment for 3 months

- Progression defined using time to symptomatic progression (TTSP) rather than time to progression (TTP)

PROBABILISTIC ANALYSIS: Beta and gamma distributions were used according to the type of resource data. Gamma distributions were used for costs, with beta distributions for adverse event rates. The lognormal parameters for efficacy parameters were made probabilistic using Cholesky decomposition and the variance-covariance matrix generated from the patient level data.

Parameter distributions for the PSA are given on Tables 23-27 of the manufacturers' submission.

Table 23: Efficacy Inputs

Base Case	TTP		OS	
	Mu	**Sigma**	**Mu**	**Sigma**
Sorafenib	4.822	0.983	5.791	1.147
BSC	4.513	0.804	5.465	1.019
Distribution	Lognormal	Lognormal	Lognormal	Lognormal

Source: SHARP trial (25)

Table 24: Lognormal Covariance Matrices for TTP and OS

	Sorafenib		**BSC**	
TTP				
	Const	ln sigma	Const	ln sigma
Const	0.004267	-	0.002534	-
ln sigma	0.000836	0.002994	0.000283	0.002373
OS				
	Const	ln sigma	Const	ln sigma
Const	0.007019	-	0.00449	-
ln sigma	0.002415	0.004141	0.001211	0.003221

Table 24: Lognormal Covariance Matrices for TTP and OS

	Sorafenib		**BSC**	
TTP				
	Const	ln sigma	Const	ln sigma
Const	0.004267	-	0.002534	-
ln sigma	0.000836	0.002994	0.000283	0.002373
OS				
	Const	ln sigma	Const	ln sigma
Const	0.007019	-	0.00449	-
ln sigma	0.002415	0.004141	0.001211	0.003221

Table 25: Utilities Input

Before progression		
Sorafenib	Mean	0.69
	SD	0.12
	Distribution	Beta
BSC	Mean	0.69
	SD	0.12
	Distribution	Beta
After progression		
Sorafenib	Mean	0.71
	SD	0.13
	Distribution	Beta
BSC	Mean	0.71
after progression	SD	0.13
	Distribution	Beta
Disutility for AEs		
Sorafenib	Mean	-0.012
	SD	0.00
	Distribution	Beta
BSC	Mean	-0.012
	SD	0.00
	Distribution	Beta

Table 26: Adverse events

	Mean	SD	Distribution	Alpha	Beta
Rates					
Sorafenib	0.069	0.005	Beta	160	2174.67
BSC	0.056	0.005	Beta	118	1972.23
Weighted cost per cycle (£)					
Sorafenib	133.62	40.09	Gamma	11.11	12.03
BSC	220.77	66.23	Gamma	11.11	19.87

Table 27: Cost Inputs (£)

Type of costs	Mean	SD	Distribution	Alpha	Beta	Source
Active treatment – routine care						
Hospitalisation	65	19.43	Gamma	11.11	5.83	Expert Opinion
Medical staff visits	230	69.10	Gamma	11.11	20.73	Expert Opinion
Lab tests	124	37.24	Gamma	11.11	11.17	Expert Opinion
Radiological tests	61	18.22	Gamma	11.11	5.47	Expert Opinion
Active treatment - after progression						
Hospitalisation	266	79.87	Gamma	11.11	23.96	Expert Opinion
Medical staff visits	480	143.88	Gamma	11.11	43.16	Expert Opinion
Lab tests	30	9.11	Gamma	11.11	2.73	Expert Opinion
Radiological tests	78	23.44	Gamma	11.11	7.03	Expert Opinion
BSC - first line						
Hospitalisation	151	45.36	Gamma	11.11	13.61	Expert Opinion
Medical staff visits	225	67.61	Gamma	11.11	20.28	Expert Opinion
Lab tests	124	37.24	Gamma	11.11	11.17	Expert Opinion
Radiological tests	61	18.22	Gamma	11.11	5.47	Expert Opinion
BSC – after progression						
Hospitalisation	386	115.77	Gamma	11.11	34.73	Expert Opinion
Medical staff visits	364	109.34	Gamma	11.11	32.80	Expert Opinion
Lab tests	30	9.11	Gamma	11.11	2.73	Expert Opinion
Radiological tests	78	23.44	Gamma	11.11	7.03	Expert Opinion
At progression - one off cost						
Hospitalisation	0	0.00	Gamma	NA	NA	Expert Opinion
Medical staff visits	0	0.00	Gamma	NA	NA	Expert Opinion
Lab tests	104	31.34	Gamma	11.11	9.40	Expert Opinion
Radiological tests	134	40.09	Gamma	11.11	12.03	Expert Opinion
End of life – one off cost	0	0	Gamma	NA	NA	NA

Details of the resource use and unit costs can be found in Appendix 13.

Comment

- Within the submission, drug costs are modified according to subgroup but the figures given do not include any uncertainty estimates (Table 17, submission). If the subgroup figures reflect subgroup usage, then we would expect greater uncertainty for the subgroup costs than for the overall costs. If the model does not include drug cost uncertainty then model uncertainty will underestimate the true level of uncertainty.
- The choice of distributions used in the probabilistic sensitivity analysis are considered appropriate. However, there are concerns that insufficient information has been provided about the values used to estimate some of these distributions, particularly the log-normal distributions used to estimate the TTP and the gamma distributions fitted to the cost data.
- Regarding subgroups, the following subgroups were evaluated in submission:
    - [redacted].
    - [redacted].
    - [redacted].
- The ERG requested further details on the analysis done in relation to the various subgroups during the clarifications stage. Changes to the model to deal with subgroups were poorly reported in the initial submission. The following response was received from the manufacturers.

From manufacturers response to requests for clarification

For the subgroup analysis the model assumes that in the given subgroups the cost and utility of each health state/treatment phase and adverse event is the same as for the overall

population; the rate of adverse events, the probability of patients continuing on sorafenib after progression, and the length of this continuation is also the same as for the overall population.

For each sub-group the following changes were made to the economic model:

- Lognormal distribution was fitted to the data and new lognormal parameters were estimated for the given subgroup outside MS Excel using a statistical package (STATA®),

- TTP and OS was recalculated for the given subgroup,

- The average dose of sorafenib used was recalculated for the given subgroup,

Lognormal parameters for the subgroups are available in the model on 'Default_effect' sheet, and average doses on the 'Default_cost' sheet.

5.1.7 Model validation

The ERG were successfully able to replicate the results presented by the manufacturers in the submission using the excel model provided. All key cost and effectiveness parameters were compared between what was reported in the submission and what was entered into the executable model and no discrepancies were identified. The ERG conclude that the model is internally valid. Additional work undertaken by the ERG using the excel model also generated results consistent with expectations. See Section 6 for details of this additional work.

5.2 Critique of approach used

The literature review contained in the manufacturer's submission concluded that the literature contained no relevant papers. The only relevant item found was an abstract published May 2008 (Muszbek et al. 2008[21]). This review included all papers "to December 2008" and as such omits a paper published December 2008 (Musbek et al. 2008[22]), which appears: a) to be subsequent research from the authors of the abstract and b) to closely follow the methodology used in the submission. The authors of this second study include two employees of Bayer, including one UK-based employee and were supported by a grant from Bayer Healthcare Pharmaceuticals. Whilst this study was in a Canadian setting, the data, assumptions and even presentation of the information is very similar to the submission. Aside from context-specific differences, three clear differences can be identified: 1) the use of life years rather than QALYs, 2) the inclusion of a separate "procedures" cost category at disease progression and 3) a Tornado diagram outlining the sources of uncertainty in the model. The Tornado diagram in this study appears to suggest the greatest source of uncertainty in the model relates to the parameters for overall survival with sorafenib and best supportive care. Whilst there are references to such parameters being important in the submission no Tornado diagram is presented.

Regarding the use of clinical evidence, the lognormal fitted function for sorafenib appears to fit the observed data more closely than the lognormal fitted function for the BSC data. In both cases, the later values appear to underestimate TTP actual observations, and this under fitting appears more extreme in the BSC data, possibly indicating a bias in favour of sorafenib. For overall survival, the lognormal fitted functions appear to fit similarly, and both appear to overestimate survival for the later observed values. This may suggest an overestimate of potential lifespan in both groups, and may overestimate the absolute benefit of sorafenib against best supportive care.

Regardless of the fit of the individual models as discussed above there remains doubt over whether or not the data chosen to estimate survival is

appropriate. The excel model provided by the manufacturers include options to consider the estimates of survival curves based on the investigators assessments of the TTP or

[redacted]. No option was provided for considering the independent assessment alone. Given disagreement in TTP that exists between the two assessments, it would have been appropriate for the independent assessment to be considered as part of the sensitivity analyses. This work has been done by the ERG and is presented in section 6.

The use of a Markov model is an appropriate approach given the decision problem. The Markov model allows for the estimation of costs and effects when these values are likely to change over time. The manufacturers have correctly chosen to model a life-time time horizon to estimate the cost-effectiveness of the treatment. Analyses based on time horizons of 2, 5 and ten years are also included to illustrate the cost-effectiveness of treatment over shorter time periods. This is appropriate given the degree of uncertainty of survival benefit for sorafenib when compared with best-supportive care.

The choice of utility values is informed by the systematic review conducted by the manufacturers. It has not been possible within the scope of the STA to individually examine all studies included in this review. The use of mapping for the estimation of utility values is acceptable within the NICE methods guidance, although it is commonly accepted in the literature that mapping represents an second-best alternative to utility estimation when compared with either direct valuation or the use of previously developed and validated general preference estimates.

The manufacturers have chosen to use a mapping algorithm (Dobrez et al) that is based on the health state valuations of patients with cancer. The NICE reference case is clear that utility values used in submissions should reflect societal preferences and not patient preferences. It is not clear from the submission what values might be appropriate estimates of societal values health states associated with HCC or whether the use of such value would be likely to have a significant impact on the results of the economic evaluation.

Probabilistic sensitivity analysis is the appropriate method for addressing uncertainty in the model parameters. The choice of distributions used to estimate the uncertainty in the model is appropriate and consistent with decision modelling good practice (Briggs et al, 2006[23]). However, as highlighted in this report, in some cases inadequate information was presented in the submission to establish how the distributions were defined and as a result it is difficult for the ERG to be certain that appropriate values for distributions have been selected.

5.3 Results included in manufacturer's submission

The base case analysis indicates that total lifetime discounted costs are £32,971 per patient, against £9,739 for best supportive care. The majority of the cost difference is the cost of sorafenib (£19,673) and other treatment costs (£857) at first line, plus additional costs at second line which are largely due to the increased survival (£11,457 sorafenib vs £7,576 best supportive care). Life expectancy and total QALYs are higher under sorafenib (1.08 QALYs, 1.54 LYG) than under best supportive care (0.72 QALYs, 1.03 LYG). The ICER is given as £64,754 per QALY, with a cost per additional life year gained of £45,502.

For the subgroup analyses, the equivalent figures are summarised in Table 30 (reproduced below).

Table 30: Results from Subgroup Analyses

	Incremental LYG	Incremental QALYs	Incremental cost (£)	Cost/ LYG(£)	Cost/ QALY(£)
Total Population (base case)	0.51	0.36	23,232	**45,502**	**64,754**
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Costs and benefits discounted 3.5%

The results from the sensitivity analyses are summarised in the manufacturer's submission in Tables 31-32, which is reproduced below.

Table 31: Scenario Analysis Discounted Results

Analyses description	Incremental LYG	Incremental QALYs	Incremental cost (£)	Cost/ LYG (£)	Cost/QALY (£)
Base Case	0.51	0.36	23,232	45,502	64,754
Discount rates					
Discount rate: costs 0%, benefits 0%;	0.58	0.41	24,399	41,883	59,545
Discount rate: costs 6%, benefits 0%;	0.58	0.41	22,524	38,666	54,972
Discount rate: costs 0%, benefits 6%	0.47	0.33	24,399	52,040	74,108
Cost data					
Zero drug costs	0.51	0.36	4,029	7,891	11,230
Same patient management costs	0.51	0.36	23,759	46,533	66,221
Management costs taken from the RCC assessment report^	0.51	0.36	21,158	41,440	58,973
Inclusion of PSS costs	0.51	0.36	24,249	47,494	67,589
Cost of death included *(£3,923)	0.51	0.36	23,147	45,334	64,515
Alternative utility assessment					
a) Separate Sorafenib and BSC**	0.51	0.36	23,232	45,502	63,739
b) AEs disutility 0.05	0.51	0.36	23,232	45,502	64,930
c) AEs disutility 0.2	0.51	0.36	23,232	45,502	65,380
d) Utility of 0.41 for all health states	0.51	0.21	23,232	45,502	110,904
e) No AE disutility	0.51	0.36	23,232	45,502	64,780
f) Utilities from RCC assessment report~~	0.51	0.36	23,232	45,502	63,992
Length of sorafenib treatment after progression					
0 months	0.51	0.36	22,296	43,668	62,144
3 months	0.51	0.36	22,949	44,948	63,965
Time horizon					
2 years	0.19	0.13	18,844	97,962	141,425
5 years	0.38	0.27	21,779	56,833	81,171
10 years	0.48	0.34	22,945	47,420	67,526
Outcomes assessment					
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

LYG= life-years gained, [redacted]

*Assumed a cost of £3,923, taken from Coyle et al (1999), averaged over hospital and hospice stays = £2,701, revalued to 2007/8

**Using the following mapped utilities: First line – no progression with sorafenib: 0.6957, First-line treatment continued – post progression with sorafenib: 0.7132, First line – no progression with BSC: 0.6818, BSC - post progression: 0.7094 (see Appendix 12)

The manufacturer's submission also presents uncertainty using cost-effectiveness planes and CEACs. The latter figures for cost per QALY and cost per LY gained are presented below.

Figure 13: Probabilistic Analysis, CEAC for BSC vs. Sorafenib, Cost per QALY

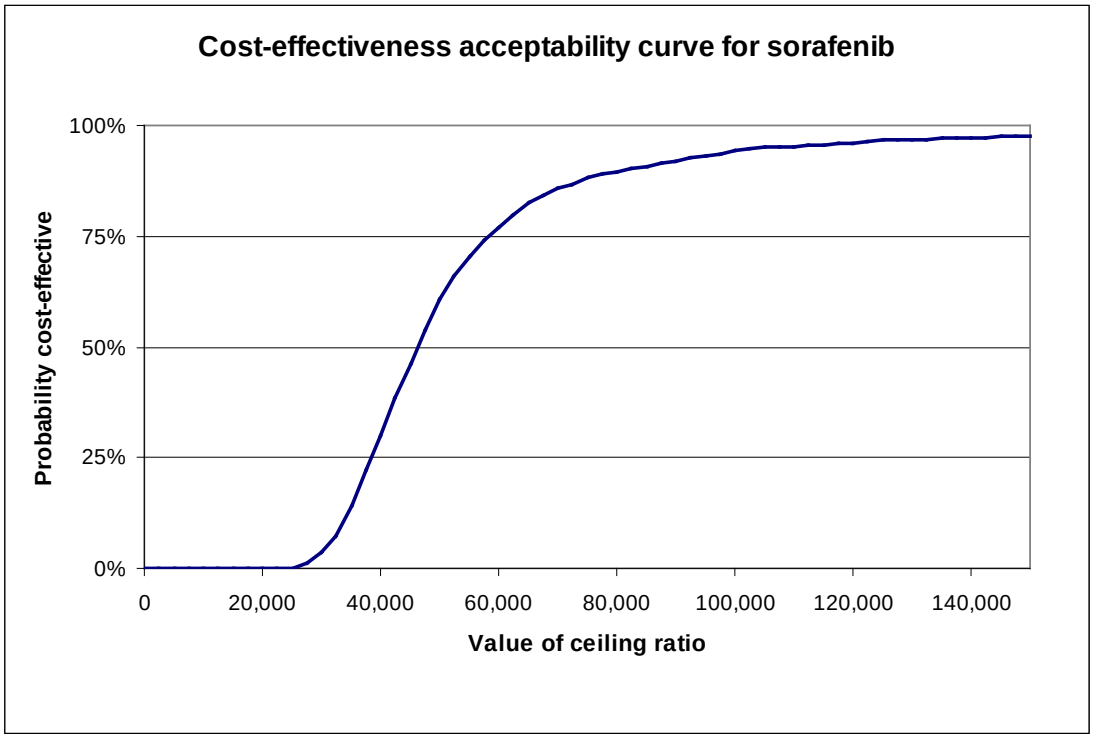


Figure 13: Probabilistic Analysis, CEAC for BSC vs. Sorafenib, Cost per QALY

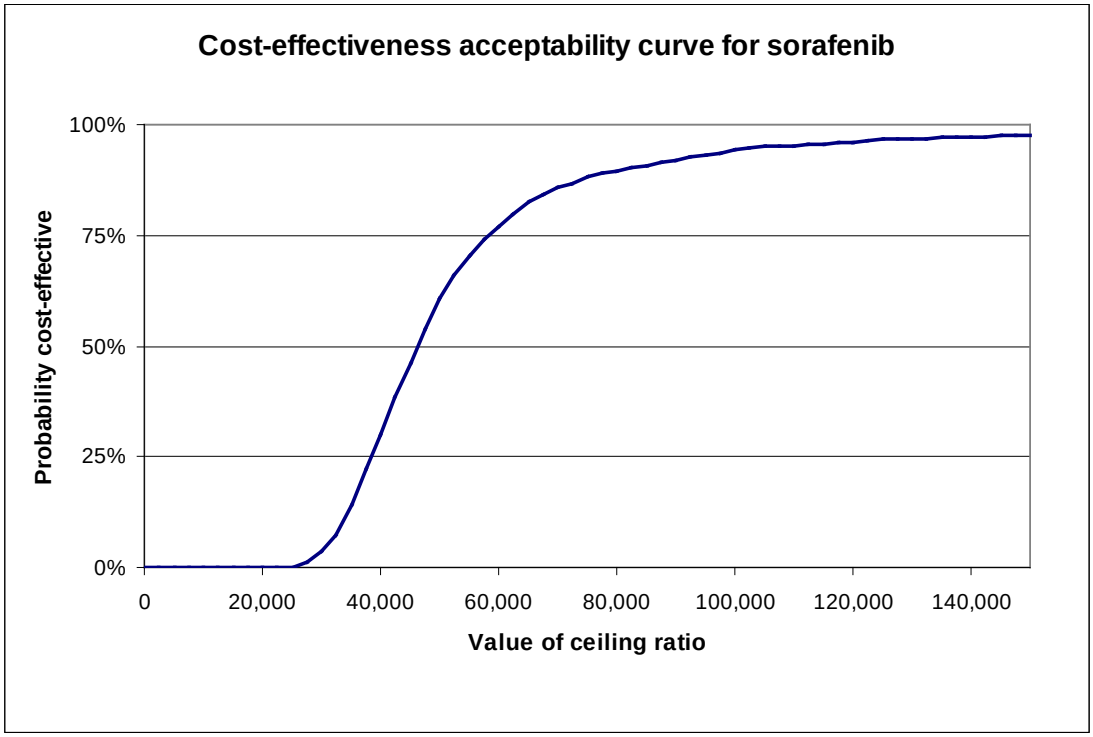
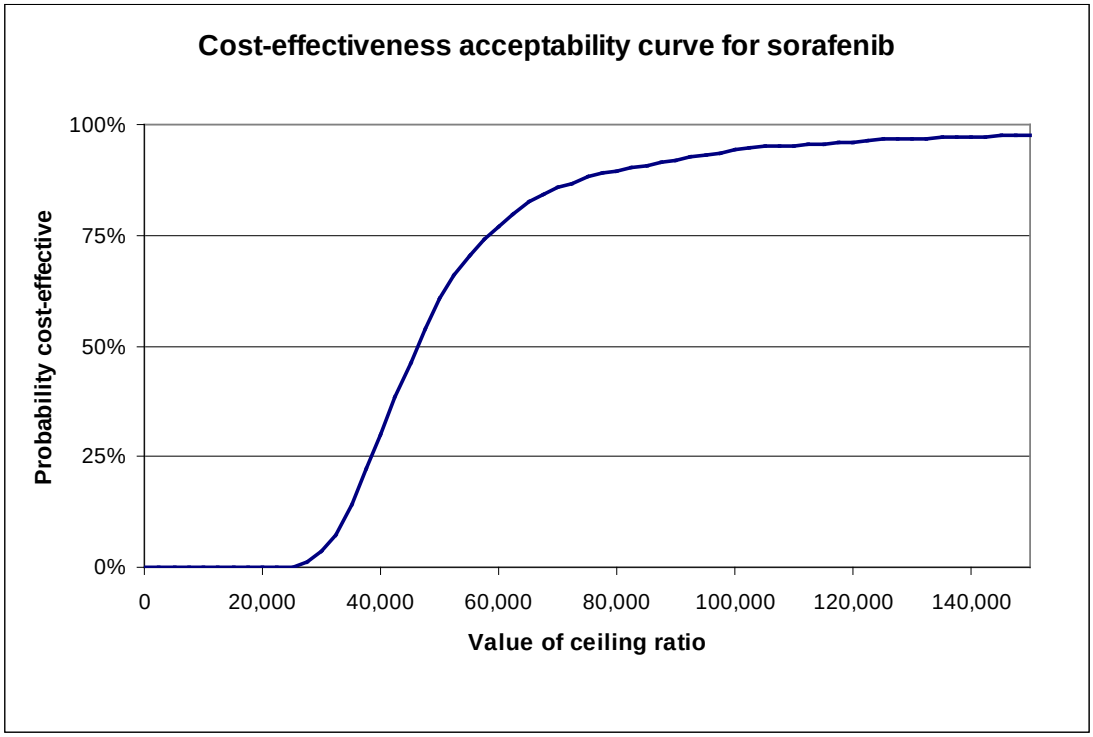


Figure 14: Probabilistic Analysis, CEAC for BSC vs. Sorafenib, Cost per LY gained



- In the base-case analysis sorafenib generates an additional 0.51 life years and 0.36 QALYs than best supportive care, at an incremental cost of £23,232. The resulting ICERs are £45,502/LY and £64,754/QALY.
- No standard deviations are reported for the cost and outcome figures.
- The subgroup analysis suggests a range of cost per QALY figures that approach cost-effectiveness in only one case ([redacted]).
- The probabilistic sensitivity analysis provides very similar cost-per-QALY and cost-per-LYG figures to the base case (£65,244 versus £64,754 per QALY; £45,832 vs £45,502 per LYG). In the one way analyses, the cost per QALY remains above the standard £30,000 threshold except in the case where drug costs are ignored, and otherwise remain above £50,000 per QALY.

- The key drivers identified by the manufacturers were said to be the estimates for TTP and OS from the SHARP trial. However, this references a section of the submission (7.3.3.1) that does not appear to exist in the current version.

A best case scenario analysis of results was requested by the ERG as part of the clarification process. This is given in the table below:

Table 4: Best case sensitivity analysis

Per Patient	LYG	QALYs	Total Costs (£)	ICER	
				Cost/LYG (£)	Cost/QALY (£)
Best case analysis for the total population					
Sorafenib	1.54	1.09	23,229	**39,627**	**55,729**
BSC	1.03	0.73	2,997		
Best case analysis for subgroups					
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]		

Comment:

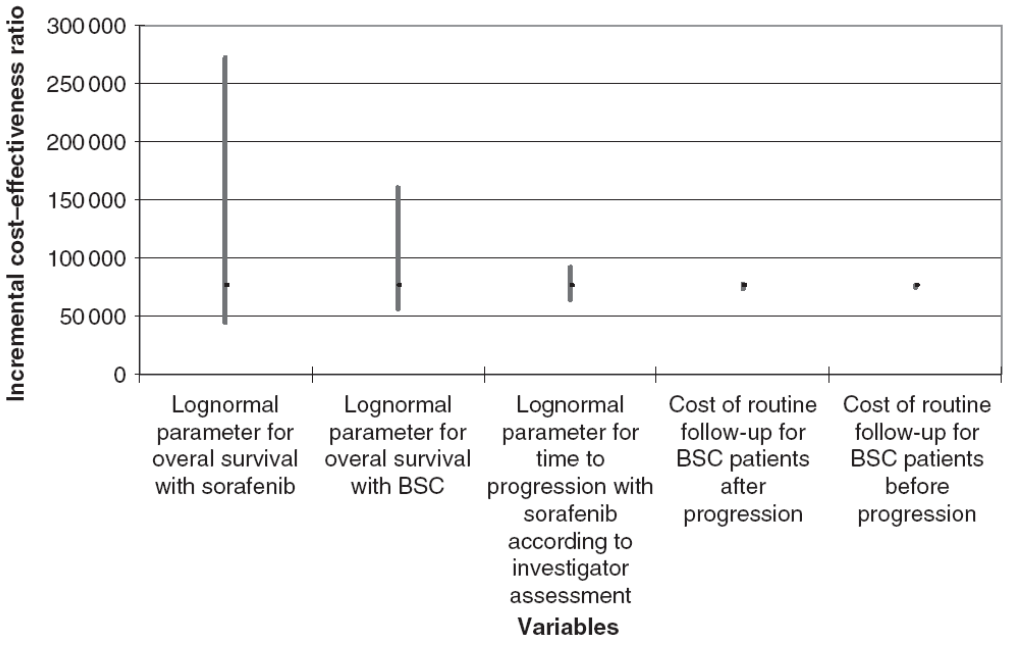
- In the case of the total population, it is clear that the treatment does not approach the commonly accepted upper threshold of £30,000/QALY to be considered cost-effective. The manufacturers also included the best case result for the best performing sub-group, [redacted].

5.4 Comment on validity of results presented with reference to methodology used

Quality assessment of the economic model using the ScHARR-TAG check list can be found in Appendix 22.

The results of the analysis provided in the submission can be considered a reliable estimate of the likelihood that sorafenib is a cost-effective treatment for HCC. The submission has used an appropriate approach to model the cost-effectiveness of treatment, although the caveats introduced in the preceding discussion about the validity of some parameter estimates must be borne in mind when drawing this conclusion. However, it is unlikely in most instances that improved parameter estimates would meaningfully reduce the uncertainty in the results. The base case results quite clearly suggest that sorafenib is not cost-effective for the treatment of HCC within the framework of the decision problem as set out in this submission.

The submission provides a list of limitations on pages 85 and 86. In addition to these, we note that the submission does not include a tornado diagram but that the very similar (or related) Canadian model does include such an analysis. The excel model provided would allow for the calculation of a tornado diagram but this information is not presented in the submission. Given the similarity between the Canadian model published December 2008 and the current model, the tornado model from that study may provide a useful indication to the committee. This tornado diagram is reproduced below, in which cost figures are varied by +/- 30% and efficacy parameters are varied by the endpoints of 95% confidence intervals.



Tornado diagram from Muszbek et al (2008b); ICERS in Can$/LY gained

Here, the base case has a cost per LY gained of around Can$75,759 with uncertainty varying between around 60%-360% of this figure. The Canadian base case can be converted to approximately £41,909 per LY gained (at 0.5532 £/$) compared to £45,502 in the UK base case, which suggests some general comparability in results. A similar level of sensitivity in the UK case would suggest a cost per LY gained between £27,000 and £164,000 using the confidence interval for the lognormal parameter for overall survival with sorafenib. Using the relatively stable relationship between LY gained and QALYs in the model noted above, this suggests that the ICER lies between £39,000 and £234,010 per QALY whilst changing this parameter. The one-way sensitivity analyses did not include this parameter.

5.5 Summary of uncertainties and issues

The resource use in the treatment of patients with HCC relied on the opinion of experts. The validity of the resulting estimates was not commented upon and nor was it assessed.

The adverse event rate modelled was also derived from expert opinion. It is possible that further information could have been obtained on the occurrence of adverse events (particularly concerning best supportive care). Furthermore, the cost of adverse events was also estimated from expert opinion. It is unclear what sources of information were used in this exercise.

It is also not clear on what basis the submission allocated a higher weighted cost per cycle for adverse events related to best supportive care. As this is not justified, it may be a source of bias in favour of sorafenib.

The model may favour sorafenib by exclusion of adverse effects of grade 3 or 4 that occur in less than 10% of the treated patients. Some adverse events of grade 3 or 4 can occur less than 10% of a sample of individuals but have an appreciable health and cost consequence. Therefore it can be argued that these should be included.

Health-related quality of life assessment plays a key role in cancer research. A published algorithm was used to map values from a cancer specific instrument (FACT-G) in order to obtain utilities. An issue is that the submission did not discuss or convincingly explain the absence of a generic measure of health-related quality of life (EQ-5D or SF-36). Furthermore the ERG found a potential error in the algorithm used.

6 Additional work undertaken by the ERG

Much of the additional work undertaken by the ERG is encompassed in the relevant sections above. In addition, the ERG explored the impact of using the independent assessment of TTP in preference to either the investigator assessment (as reported in the submission) [redacted]. The results are provided below followed by comment on the relevance this may have for the estimates of cost-effectiveness presented in the submission.

As described in the effectiveness section individual patient data was used to generate Kaplan-Meier curves for placebo and sorafenib groups for TTP. These were then fitted with log normal distributions using the same methodology as that in the submission. The log normal parameters are shown in the table below and were used in the model for deterministic analysis.

[redacted]

		Investigator assessment	[redacted]	Independent assessment
BSC	mu	[redacted]	[redacted]	[redacted]
	sigma	[redacted]	[redacted]	[redacted]
sorafenib	mu	[redacted]	[redacted]	[redacted]
	sigma	[redacted]	[redacted]	[redacted]

Tables reporting the model results using the independent [redacted] parameters for TTP are presented below. The investigator assessment – the base case – is also presented for reference

Results based on the investigator assessment – Base case

Per Patient	LYG	QALYs	Total Costs (£)	ICER	
				Cost/LYG (£)	Cost/QALY (£)
Sorafenib	1.54	1.08	**32,971**	**45,502**	**64,754**

BSC	1.03	0.72	**9,739**

Results based on independent assessment of TTP.

Per Patient	LYG	QALYs	Total Costs (£)	ICER	
				Cost/LYG (£)	Cost/QALY (£)
Sorafenib	1.54	1.06	**37,166**	**53,284**	**76,067**
BSC	1.04	0.73	**9,650**		

[redacted]

Per Patient	LYG	QALYs	Total Costs (£)	ICER	
				Cost/LYG (£)	Cost/QALY (£)
Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]		

The base-case scenario gives ICERs of £64,754/QALY and £45,502/LY. Using the survival curves estimated by the ERG for the independent assessor results, the ICER is £76,067/QALY or £53,284/LY, considerably higher than those estimated using the investigator estimated survival curves. For comparison, the ICER reported [redacted].

There is a substantial difference in the ICERs estimated using each of the investigator, [redacted] and independent based estimates of TTP, when holding all other base-case assumptions equal. That these results were not included in the manufacturers submission is an important omission when making the argument for sorafenib to be considered as a cost-effective treatment for use in the NHS. What the results of this analysis show is the wide range in ICERs that can be expected when examined using the full range of plausible values for the crucial clinical effectiveness parameters relating to TTP.

An important caveat however is that the excel model provided as part of the submission was not specifically designed to include the option of considering only the independent assessment of TTP. As a result, modifications made to the model cannot necessarily be fully tested and validated by the ERG, and so these results should be interpreted with caution. The submission lacked an economic analysis that used [redacted] even though this was a feasible option in the excel model provided. By only including [redacted] and investigator assessments as options within the excel model, the manufacturers reduce the reliability that the ERG can have in relation to their overall estimates of cost-effectiveness.

7 Discussion

7.1 Summary of clinical effectiveness issues

The main issues regarding the clinical effectiveness of sorafenib are:

Does the gain in overall survival for patients receiving sorafenib (versus no active systemic therapy) observed for Child-Pugh grade A patients extend to other patients with poorer liver function, especially those with Child-Pugh grade B classification.

Is best supportive care (i.e. no active systemic intervention) the appropriate comparator for investigating the clinical effectiveness of sorafenib.

Although sorafenib clearly extends the median time to radiographic disease progression there is uncertainty about the precise size of this advantage.

From the trials conducted there is very little good information about the quality of life of patients with advanced HCC receiving sorafenib or no active systemic therapy.

7.2 Summary of cost effectiveness issues

The choice to focus on best-supportive care as the only comparator in the economic model is not unequivocally supported by evidence that doxorubicin is not used in clinical practice in the UK.

The economic evaluation relies heavily on the use of expert opinion for estimating resource use for the treatments in the model. As a result, significant uncertainties remain over the cost of treatments that are not adequately expressed in the model.

The economic evaluation relies on the use of mapped estimates of health related quality of life. Although this is acceptable according to the NICE methods guidance, it is a second best option and primary data collection of health related quality of life information within the clinical trial using a validated preference based instrument would provide a more robust estimate.

The submission does not include cost-effectiveness estimates using the independent assessment of TTP. This is an important omission, and the ERG have undertaken some sensitivity analysis to address this.

The inclusion of adverse events in the model was unsatisfactory and inconsistent between the treatment and placebo arms of the model.

7.3 Implications for research

To consolidate the evidence base about the clinical and cost-effectiveness of sorafenib used alone or in combination for patients with advanced HCC future research should be directed toward:

- Investigation of patients with worse liver function than Child-Pugh grade A
- Head to head trials comparing sorafenib with established and new systemic agents
- Trials that may identify the most effective and cost effective combinations of sorafenib with other potentially therapeutic agents.

A value of information analysis could be considered before further research is undertaken. No significant difference was shown in health related quality of life between sorafenib and best supportive care and the resulting ICER exceeded the maximum willingness to pay of £30,000 by a considerable margin. Value of information analysis will assist in determining whether or not

any further reduction in uncertainty regarding the cost-effectiveness of sorafenib is worth the additional cost of undertaking the research.

8 Appendices

Appendix 1 The BCLC (Barcelona Clinic Liver Cancer) staging system

STAGE	PST	Tumour stage	Okuda stage	Liver functional status
Stage A: early HCC				
A1	0	single	I	No PH & normal bilirubin
A2	0	single	I	PH & normal bilirubin
A3	0	single	I	PH & abnormal bilirubin
A4	0	3 nodules < 3 cm	I – II	Child-Pugh A -B
Stage B: intermediate HCC	0	Large multinodular	I – II	Child-Pugh A -B
Stage C: advanced HCC	1 – 2*	Vascular invasion or extrahepatic spread	I – II	Child-Pugh C
Stage D: end-stage HCC	3 – 4	Any	III	Child-Pugh A -B

PH portal hypertension
Stage A and B: all criteria should be fulfilled
Stage C at least one criterion; * PST 1 – 2 or vascular invasion / extrahepatic spread
Stage D at least one criterion ** PST 3 -4 or Okuda stage III/Child-Pugh C

Appendix 2 Child-Pugh grading of cirrhosis

Criteria taken from the submission appendix are shown below:

Measure	Score		
	1 point	2 points	3 points
Ascites	Absent	Slight	Moderate
Bilirubin (mg/dL) **(µmol/L)**	<2.0 <34	2.0-3.0 34-50	>3.0 >50
Albumin (g/dL) **(g/L)**	>3.5 >35	2.8-3.5 28-35	<2.8 <28
PT prolonged (sec) **PT prolonged (%)** **INR**	<4 >60 <1.7	4-6 40-60 1.7-2.3	>6 <40 >2.3
Encephalopathy	Stage 0-Absent	Stage 1-2 – Moderate	Stage 3-4 - Severe

Child-Pugh A: 5 or 6 points

Child-Pugh B: 7-9 points

Child-Pugh C: >9 points

Appendix 3 Further comments on submission description of current service provision

1. The systematic review searches (reference 22) were updated to November 2008 and thus this review should be described as “concurrent with” rather than “prior to” the introduction of sorafenib (marketing authorisation October 2007).

2. Reference 7 is used to support the claim that “*placebo / BSC is justified as being a relevant comparator arm*” but searches in this study were completed in 2003. Studies published in the last 6 years are not represented and therefore this reference may not accurately reflect current practice.

3. It is unlikely that the systematic review (ref 22) of trials aimed at identifying novel or other beneficial agents for treatment of HCC would provide insight about what constitutes current UK service provision as implied in the submission; inevitably such a review will reflect heterogeneity in terms of dosage and treatment regimens, study population characteristics and outcome measures. It may be that there is no information about current provision in the UK, however if this is the case this should be stated in the submission.

Appendix 4 The manufacturer's definition of the decision problem

The manufacturer's definition of the decision problem compare to the scope is shown below:

	Final scope issued by NICE	Decision problem addressed in the submission
Population	Adults with advanced hepatocellular carcinoma whose disease is unsuitable for local or loco-regional curative therapy or has progressed after those types of therapy	Sorafenib is indicated for the treatment of hepatocellular carcinoma (Appendix 1), the main type of primary liver cancer. In April 2006, on account of the small number of cases and lack of alternative therapies in HCC, sorafenib was granted European and US orphan drug status. The decision problem addressed in the submission is the clinical benefit and cost-effectiveness of sorafenib as a treatment in those patients with advanced stage hepatocellular carcinoma disease who have failed or are unsuitable for surgical or locoregional therapies. There are approximately 2340 new cases of HCC diagnosed in England and Wales each year. Of these the population eligible for Nexavar is around 700.
Intervention	Sorafenib (Nexavar)	Sorafenib tosylate (Nexavar®), a multi-kinase inhibitor, is an oral therapy for HCC, targeting both tumour angiogenesis (vasculature) and tumour cell proliferation. Sorafenib is administered orally in the form of 200 mg film coated tablets. The recommended dose of sorafenib in adults is 400 mg twice daily (bd; equivalent to a total daily dose of 800 mg). Treatment should continue as long as clinical benefit is observed or until unacceptable toxicity occurs.
Comparator(s)	Standard care which may include doxorubicin, cisplatin or biological agents, depending on performance status and severity	Sorafenib will be compared to best supportive care. Due to the underlying liver disease and lack of effective treatments, patients diagnosed with advanced HCC have a bleak prognosis. Sorafenib is the only treatment to have demonstrated a survival benefit in advanced HCC for over 30 years. No systemic agent has shown survival benefit versus placebo in HCC in more than 75 randomised controlled trials and, in most cases, such treatments are associated with a high rate of side effects. As a result, there are no treatments, other than sorafenib, with FDA and/or EMEA approval for advanced HCC. Furthermore, because of the advanced nature of the disease, surgery is not a treatment option. Guidelines (BSG 2003) recommend that systemic chemotherapy with standard agents have a poor response rate and should only be offered in the context of clinical trials of novel agents. Best supportive care is the most appropriate comparator for these patients. This is supported by various reviews, meta-analyses and systematic reviews published over the past decade which

		conclude that no anti-cancer treatment has clearly been identified as either a 'gold standard' or to demonstrably improve overall survival.
Outcomes	The outcome measures to be considered include: • Overall survival • Progression free survival • Time to symptomatic progression • Tumour response • Health related quality of life • Adverse effects of treatment	The outcomes listed will be presented in the submission. Advanced HCC is a unique condition which poses methodological issues when evaluating the impact of new treatments on health related quality of life. Patients with hepatocellular carcinoma are heterogeneous, with a diverse range of underlying causes of cirrhosis, including hepatitis B, hepatitis C, alcoholism and haemochromatosis. In some patients, typically younger women, HCC may develop where cirrhosis is not present. Due to this diverse liver disease, it is particularly difficult to disentangle the effect of the advanced HCC, underlying liver disease and interventions on quality of life. More specifically, quality of life is likely to be affected by the symptoms of the underlying liver disease, including liver failure, irrespective of whether the tumour has stabilised or regressed. As a result, it is not possible to demonstrate the impact of treatments in advanced HCC on quality of life, and no robust and reliable utility data is available that separates out the effect of the primary liver cancer from the underlying liver disease causes.
Economic Analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality adjusted life year. The time horizon for the economic evaluation should be sufficiently long so as to incorporate all the important costs and benefits related to the condition. Where the evidence allows, any likely dose adjustment during the treatment should be taken account of. Costs will be considered from and NHS and Personal and Social Services Perspective	The economic evaluation will be a cost effectiveness analysis, with the results presented as incremental cost per quality adjusted life year and life year gained. Taking this uniqueness of confounding co morbidities into consideration, the QALY would not be an appropriate outcome to measure the health benefit of patients with advanced HCC, therefore the cost per life years gained figures should also be given consideration. Due to the advanced nature of the disease, the model will be a lifetime model, consisting of three health states; non-progressive advanced disease, progressive disease, and death. The model will also consider dose adjustments during the treatment period. Costs will be considered from an NHS perspective.
Special considerations, including issues related to equity or equality Subgroups to	If the evidence permits, the appraisal will seek to identify subgroups of individuals for whom sorafenib may be particularly clinically and cost effective, for example by age, performance status or degree of underling	Patients with advanced HCC have a heterogeneous co morbidity profile that affects their prognosis, quality of life and treatment. Given that this is an end of life medicine, with small patient numbers, a demonstrable survival benefit and no alternative treatments, sorafenib should be considered under the End of Life Policy. Applying a single estimate of cost-effectiveness to the

be considered	cirrhosis. Guidance will only be issued in accordance with the marketing authorisation.	overall advanced HCC group of patients is unreliable because of the unique large variation in underlying disease (e.g. liver cirrhosis), rarely seen in other cancers, it is therefore of utmost importance to base decisions on patient sub-groups where the health and economic outcomes are most likely to vary considerably from the overall mean. It is acknowledged there is a high degree of variability around the point estimate of cost effectiveness due to the heterogeneous nature of the disease and the difficulty disentangling the underlying liver disease and treatment effects. For these reasons it would be appropriate to collect further evidence as recommended under the end of life scheme.

Appendix 5 Further comments on definition of comparator

Minor issues are:

Reference 3 is the NEJM report of the SHARP RCT of Sorafenib; this publication does not consider treatments implemented over the last 30 years as implied in the submission.

Reference 4 (Lopez 2006[24]) is cited for evidence that BSC is the best comparator. Lopez reviewed only six trials of systemic treatments (review of 75 such trials is implied in the submission). Furthermore not all of these six employed BSC/placebo as the comparator. The review did indeed conclude that systemic treatments failed to deliver any survival benefit.

Appendix 6 The manufacturer's description of search strategy

The searches in the main submission (appendices 2, 3 and 11) are in effect the same as those detailed in the systematic review on which the submission was based (Reference 22).

Searches were quite extensive since the review is described on p12 of 97 as "*a systematic review which identified studies involving sorafenib, doxorubicin, placebo or BSC in advanced HCC…… and informed on the heterogeneity in terms of dosage and treatment regimens, study population characteristics and outcome measures*." The searches therefore comprised separate strategies for sorafenib, doxorubicin, placebo/best supportive care and natural history of advanced HCC. The results of these searches were also used to inform on the most appropriate comparators to use in the economic modelling.

Four bibliographic databases were searched (MEDLINE Dialog Datastar 1950 – 21 November 2008, EMBASE Dialog datastar 1974 – 21 November 2008, The Cochrane Library (DARE CDSR CENTRAL HTA) 2008 Issue 4. Additional studies were identified in a search of abstracts from key Oncology conferences (American Society of Clinical Oncology (ASCO) 2000-2008, European Cancer Conference (ECCO) ASCO, ECCO 2003-2008). One clinical trials database was searched – National Cancer Institute's Clinical Trials Database which identified 6 studies relevant to use of sorafenib in HCC. The reference lists of relevant articles identified in the database searches were also hand-searched.

The searches were restricted to English language with no date limits.

Appendix 7 ERG search strategies

Database: Ovid MEDLINE(R) <1950 to November Week 3 2008>

Search Strategy:

1 liver cancer.mp. or exp Liver Neoplasms/ (100443)
2 ((liver or hepatocellular or hepatic) and (carcinoma or cancer)).mp. [mp=title, original title, abstract, name of substance word, subject heading word] (95196)
3 exp Carcinoma, Hepatocellular/ or hcc.mp. (44646)
4 or/1-3 (131449)
5 sorafenib.mp. (555)
6 sorafinib.mp. (2)
7 nexavar.mp. (30)
8 or/5-7 (556)
9 4 and 8 (72)
10 limit 9 to (english language and humans) (63)
11 from 10 keep 1-63 (63)

Database: EMBASE <1980 to 2009 Week 03>

Search Strategy:

1 liver cancer.mp. or exp Liver Cancer/ (68033)
2 hcc.mp. (12365)
3 ((hepatic or hepatocellular or liver) and (cancer or carcinoma)).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (104994)
4 or/1-3 (111292)
5 sorafenib.mp. or exp Sorafenib/ (2680)
6 nexavar.mp. (612)
7 sorafinib.mp. (8)
8 or/5-7 (2684)
9 4 and 8 (486)
10 limit 9 to (human and english language) (419)
11 from 10 keep 1-419 (419)

Appendix 8 Ongoing studies identified in the submission

Ongoing studies identified in the submission are shown below

Table 3 : Ongoing studies involving sorafenib in HCC

Title	Arm A	Arm B	Expected Accrual	Status	Data source
Phase IV					
GIDEON – Post Marketing Surveillance Study in HCC	Sorafenib		3000	Ongoing Expected closure September 2013	NCI / Bayer
Phase III					
Sunitinib vs sorafenib in patients with inoperable liver cancer	Sorafenib	Sunitinib	1200	Ongoing Expected closure July 2012	NCI
Study 11721 Phase III Study of BAY 43-9006 in Patients With Advanced Hepatocellular Carcinoma Treated After TACE (Japan)	Sorafenib	Placebo	414	Ongoing Expected closure March 2010	NCI / Bayer
Phase II					
Study 11546 A randomised controlled study of BAY 43-9006 in combination with doxorubicin versus doxorubicin in patients with advanced hepatocellular carcinoma.	Sorafenib + doxorubicin	Doxorubicin	96	Closed	NCI
Dose Escalation of Sorafenib in Patients With Advanced HCC (Italy)	Sorafenib	-	100	Ongoing	NCI
Phase I & unspecified					
Sorafenib in locally advanced or metastatic liver cancer with Child B cirrhosis	Sorafenib	-	30	Ongoing Expected closure September 2010	NCI

The study BAY 43-9006 listed in the submission as ongoing is the Abou-Alfa et al[13] study and is completed. A presentation entitled "*Final results of phase II, randomised, double blind study of sorafenib plus doxorubicin and placebo plus doxorubicin in patients with advanced hepatocelluar carcinoma*" by Abou-Alfa et al is downloadable from the internet.

Although only one clinical trials database was searched (see 6.2.5 - National Cancer Institute's Clinical Trials Database which is a subset of ClinicalTrials.gov) it is unlikely trials would have been missed, given searches of key Oncology conferences (American Society of Clinical Oncology - ASCO, European Cancer Conference -ECCO) were also conducted. Searches of additional trials databases were conducted by the ERG (the whole of

ClinicalTrials.gov database plus the ISRCTN Register on Current Controlled Trials and CRN Portfolio) which did retrieve additional references. These were screened by reviewers and one ongoing study of potential relevance was found.

Appendix 9 Submission table of included studies.

Two studies compare single-agent sorafenib with placebo. At the time of the systematic review the Asia-Pacific study was published only in abstract form, while the SHARP study, had been analysed and fully published. Since then the Asia-Pacific study has been published on-line (17th December 2008).

Table 1: RCTs involving sorafenib as a single-agent identified during the systematic review

Author	Study Title	No of patients / Interventions
Llovet 2008 (3) , ASCO abstract 2007 (25)	A Phase III randomised, placebo-controlled study of sorafenib in patients with advanced hepatocellular carcinoma *[also known as the SHARP (Sorafenib HCC Assessment Randomised Protocol) study]*	n=602 Sorafenib 400mg bd n=299 vs placebo n=303
Cheng 2008 (23,36)	Randomised phase III trial of sorafenib versus placebo in Asian patients with advanced hepatocellular carcinoma (Study 11849 Asia Pacific trial)	n=226 Sorafenib 400mg bd n=150 vs placebo n=76

The Asian- Pacific study (Cheng 2008) (23,36) will be provided as supporting data because this is based on a different patient population with different underlying characteristics and aetiologies.

6.2.3 List of relevant RCTs

Table 2: Relevant RCTs involving sorafenib as a single-agent identified during the systematic review

Author	Study Title	No of patients / Interventions
Llovet 2008 (3) , ASCO abstract 2007 (25)	A Phase III randomised, placebo-controlled study of sorafenib in patients with advanced hepatocellular carcinoma *[also known as the SHARP (Sorafenib HCC Assessment Randomised Protocol) study]*	n=602 Sorafenib 400mg bd n=299 vs placebo n=303

The SHARP study(3,28), which has been analysed and fully published will provide the evidence for the clinical effectiveness of sorafenib in HCC in this submission. The Asia-Pacific study (Cheng 2008) (23,36) will be provided as supporting data because this is based on a different patient population with different underlying characteristics and aetiologies.

6.2.4 List of relevant non-randomised controlled trials

One phase II study examines the use of sorafenib in an open multicentre study (24,37) and will be used where appropriate to support the SHARP study results.

Appendix 10 Critical appraisal of the SHARP randomised controlled trial

The submission's critical appraisal of the SHARP trial is shown below.

6.3.6 Critical appraisal of relevant RCTs

	SHARP study, Llovet 2008
How was allocation concealed?	Bayer prepared computer-generated randomisation list. The randomisation number for each patient was provided through telephone interactive voice response system (IVRS). The unique randomisation number of each patient was used on all medication labels (placebo & active treatment). Placebo & active treatments were identical in appearance and given under identical conditions. Randomisation codes kept in individual sealed envelopes.
Randomisation Technique	Computer-generated randomisation list. Randomisation was done stratified by region, ECOG performance status (0 versus 1 or 2), and 'tumour burden (presence or absence of macroscopic vascular invasion (as determined through radiological assessment) and / or extrahepatic spread. The randomisation number for each patient was provided through telephone interactive voice response system (IVRS)
Was a justification of sample size provided?	Yes, see section 6.3.5 Power of study/sample size
Was follow-up adequate?	Yes. Period of recruitment: March 2005 to April 2006 During the follow-up period patients were assessed every 3 months until death for survival status and receipt of any new cancer treatment.
Were the individuals undertaking outcome assessment aware of allocation?	Independent assessors of response / progression were blinded to the treatment.
Parallel group or cross-over?	Parallel Group. At the second interim analysis, sorafenib was found to significantly prolong survival, which meant that all patients ongoing in the double-blind phase, as well as patients in follow-up were unblinded and given the opportunity to enter into an 'extension with crossover' study phase. After this point only safety data were collected.
Location effects	UK participants n=16 (3%)(n=7 placebo, n=9 sorafenib). Majority of subjects were noted as White (n=273 placebo; n=261 sorafenib). Eighty seven per cent of the placebo patients (n=263) were from 'Europe & Australasia' as were 88% (n=263) of the sorafenib patients. No location effect likely.
Dosage regimens	As per SPC (see Appendix 1) Sorafenib 400mg b.d. for as long as a clinical benefit is observed or until unacceptable toxicity occurs.
Were study groups comparable?	Yes, demographic, baseline and surgical characteristics were similar across treatment groups.
Were the statistical analyses used appropriate?	Yes
Was an intention-to-treat (ITT) analysis undertaken?	Intention-to Treat analysis used See section 6.3.2
Confounding factors?	None identified. The study design and selection and measurement of endpoints were discussed and agreed with the US and European licensing authorities prior to study initiation.

6.8.3 Critical appraisal of relevant non-RCTs

The phase II study is an open, single arm, uncontrolled study. It is therefore not possible to directly compare results from this study with other RCTs.Patients enrolled in the study came from the USA, Belgium, France, Italy and Israel.

Patients from the UK would be expected to have similar baseline and demographic characteristics to the study population.

The dose of sorafenib utilised in this study is as described & indicated within the Nexavar Summary of Product Characteristics (see Appendix 1).

Appendix 11 RECIST and WHO criteria for assessment of tumour response

Response Evaluation Criteria In Solid Tumors (RECIST) criteria taken from the submission appendix are shown below:

Complete response (CR)	Disappearance of all clinical and radiological evidence of tumour
Partial response (PR)	At least a 30% decrease in the sum of longest diameter (LD) of target lesions taking as reference the baseline sum LD
Stable disease (SD)	Neither sufficient shrinkage to qualify for PR, nor sufficient increase to qualify for PD
Progressive disease (PD)	At least a 20% increase in the sum of LD of measured lesions taking as reference the smallest sum LD recorded since the treatment started, or the appearance of 1 or more new lesions

WHO response Criteria for tumour response taken from the submission appendix are shown below:

Objective response:	
Complete response (CR)	Disappearance of all known lesion(s); confirmed at 4 weeks
Partial response (PR)	At least 50% decrease; confirmed at 4 weeks
Stable disease (SD)	Neither PR nor PD criteria met
Progressive disease (PD)	An increase of 25% or more in the sum of all target lesions area; no CR, PR or SD documented before increased disease

Appendix 12 Comparison of SHARP OS results from submission with published and trial report results

The ERG requested and were granted access to the full SHARP trial report in this document the time scale used is days. The time scale for the results in the published report for SHARP was in months. The submission chose the time scale to be weeks. These different scales slightly complicated cross-checking of results, for example the numbers at risk shown below the time axis on the Kaplan-Meier plot in the submission and in the publication are different. The ERG have checked the results from the submission against those in the SHARP publication and those in the SHARP trial report and found the results correspond closely. Any differences can be attributed to rounding procedures or assumption of days / month. The comparison is tabulated below.

Measure	Submission	Publication	[redacted]
Hazard Ratio (95%CI)	0.69 (0.55, 0.88)	0.69 (0.55, 0.87)	[redacted]
Hazard Ratio p value	0.000583	<0.001	[redacted]
Median survival sorafinib (95%CI)	46.3 weeks (40.9, 57.9)	45.85 weeks (40.29, 57.0)	[redacted]
Median survival placebo (95%CI)	34.4 weeks (29.4, 39.4)	33.86 weeks (29.14, 39.0)	[redacted]
Median survival difference sorafenib - placebo	11.9 weeks	12 weeks	[redacted]

Appendix 13 Use of hazard ratio to calculate overall survival advantage

A] Interpretation of Hazard Ratios.

General

Hazard Ratio
= Hazard of death on Sorafenib / Hazard of death on Placebo = 0.69
i.e. there is a 31% reduction in the hazard (or instantaneous risk) of death with Sorafenib compared to Placebo.

Generally 1/HR just inverts the comparison i.e. it becomes:
Hazard of death on Placebo / Hazard of death on Sorafenib = 1/0.6931 = 1.44
i.e. there is a 44% increase in the hazard of death with Placebo compared to Sorafenib.

Special Case of Exponential Distribution for Survival Times

If the survival times can be assumed to have an exponential distribution then the following is true.

For Sorafenib:
Survivor function = $S_s(t) = \exp(-\lambda_s t)$
Hazard rate = λ_s
Mean survival time = $1 / \lambda_s$
Median survival time = $(\ln 0.5) / (-\lambda_s)$

For Placebo:
Survivor function = $S_p(t) = \exp(-\lambda_p t)$
Hazard rate = λ_p
Mean survival time = $1 / \lambda_p$
Median survival time = $(\ln 0.5) / (-\lambda_p)$ (note: ln 0.5 = -0.6931)

So comparing Sorafenib with Placebo:
Hazard Ratio = $\lambda_s / \lambda_p = 0.69$
i.e. there is a 31% reduction in the hazard of death with Sorafenib compared to Placebo.

Relative difference in mean survival time = $(1 / \lambda_s) / (1 / \lambda_p) = \lambda_p / \lambda_s$ = 1/HR = 1/0.6931 = 1.44
i.e. there is a 44% increase in the mean survival time with Sorafenib compared to Placebo.

Relative difference in median survival time = $(\ln 0.5 / -\lambda_s) / (\ln 0.5 / -\lambda_p) = \lambda_p / \lambda_s$ = 1/HR = 1/0.6931 = 1.44
i.e. there is a 44% increase in the median survival time with Sorafenib compared to Placebo.

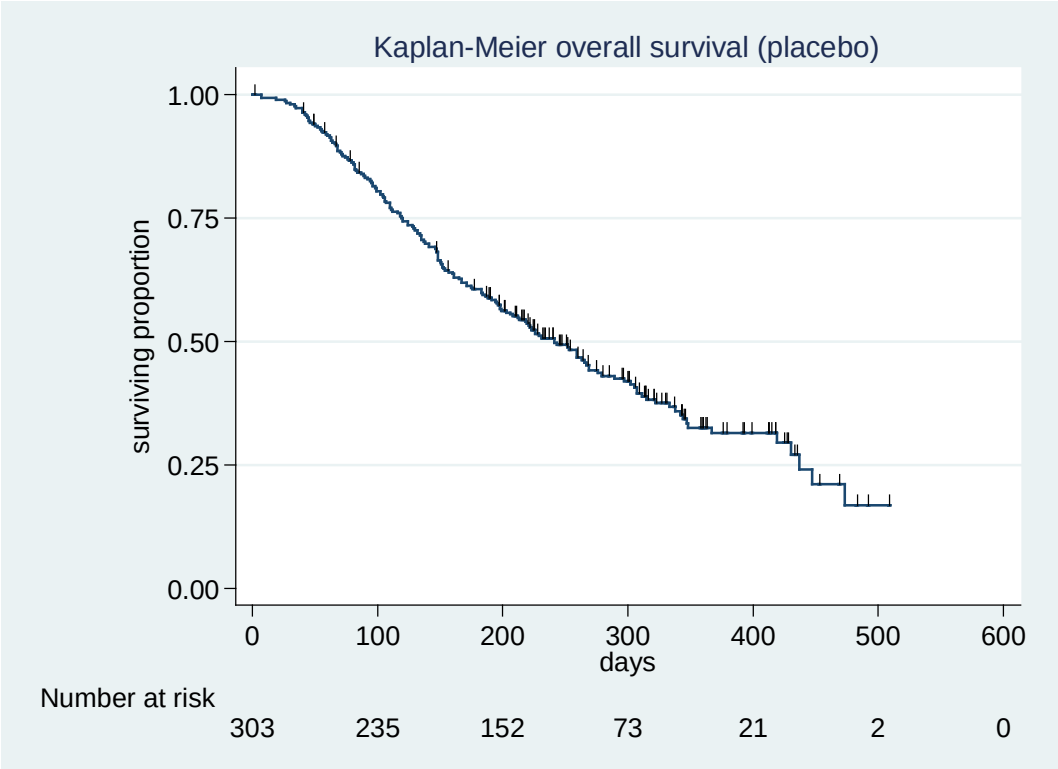
Conclusion

In the special case when survival times have exponential distribution then the inverse of the hazard ratio is equivalent to the relative difference in median and mean survival time and this is true more generally at any point on the survival curve.

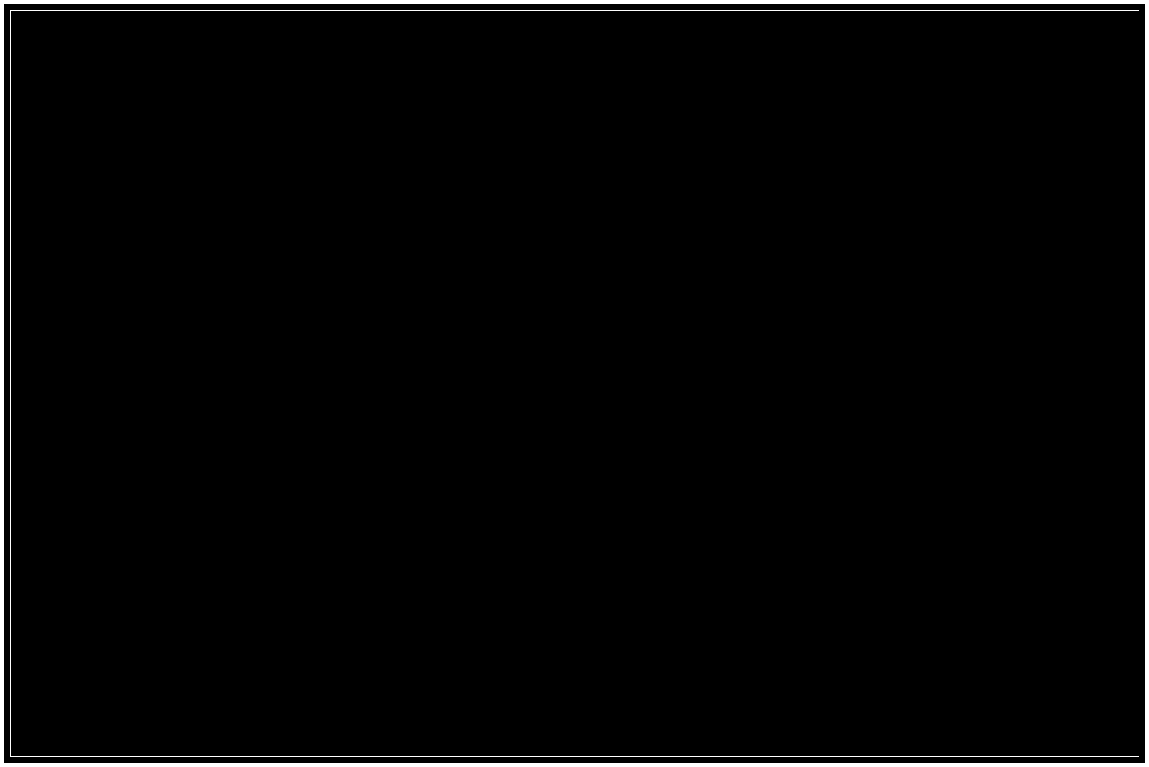
Therefore the hazard ratio for sorafenib versus placebo of 0.69 can be interpreted as 'sorafenib gives a 44% increase in survival time' BUT this is wholly reliant on the survival times having an exponential distribution.

Note: the observed medians give relative difference in median survival of 46.3 / 34.4 = 1.35 i.e. 35% not 44%.

B] **Examination of the exponential assumption for overall survival in the placebo group of SHARP.**
The ERG extracted the individual patient data from the SHARP trial report and using STATA (v 10) software constructed a Kaplan-Meier plot (shown below).



The plot below is taken from the SHARP trial report (the upper and lower curves are for the sorafenib and placebo groups respectively)



In this figure the ERG placebo survival plot is superimposed (solid line) on that from the SHARP trial report; an exact correspondence is seen.

Parameters for the placebo group provided in the trial report and derived independently by the ERG are shown below

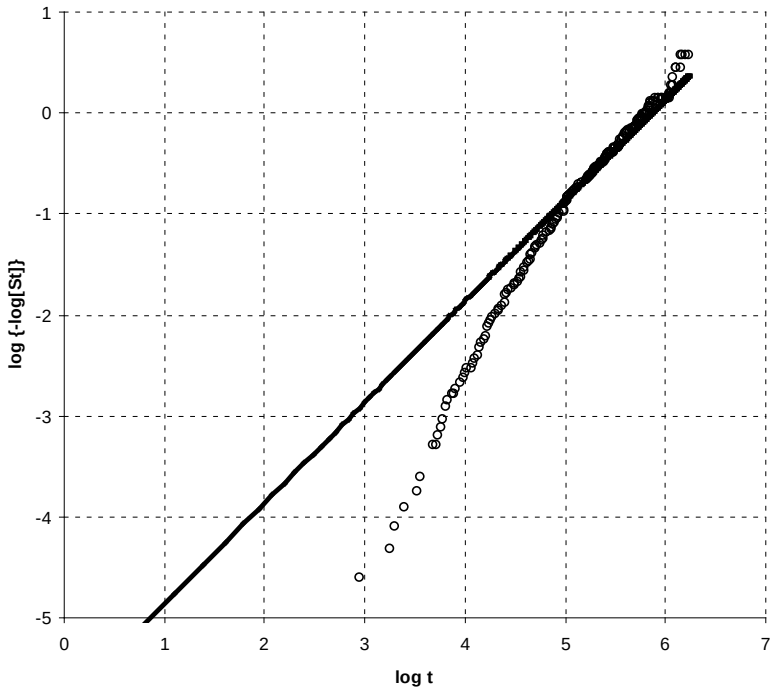
	SHARP report	ERG
Number analysed	[redacted]	303
Number failed		178
Number censored		125
Median survival	[redacted] days	241 days

The submission provides lognormal fit parameters for the Kaplan-Meier survival data for overall survival in the placebo group; below these are compared with those obtained by the ERG from the ERG independently determined survival probabilities.

	Submission	ERG analysis
Lognormal mu	5.465	5.464731
Lognormal sigma	1.019	1.018836
Log likelihood		-345.599

The ERG results are identical to those in the submission connforming that the ERG extracted data faithfully from the trial report.

Using the independently derived K-M estimates for survival ERG explored the relationship between “log time” and “log (- log survival)” which can be used as a diagnostic tool to test for an exponential relationship. When the exponential assumption is upheld the data points have a slope equal to one and tend to follow a straight line (a constant hazard rate). The result is shown below.



The solid line derives from an exponential distribution fitted to the survival data and has a slope of one. It can be seen that the slope of the empirical data (circles) deviates considerably from one (indicating poor support for the exponential assumption) and that the data deviates somewhat from a straight line indicating the lack of a constant hazard rate.

The economic section of the submission provides information about exploration of parametric fits to survival data used for economic modelling. Of the five models for overall survival in the placebo group (lognormal, loglogistic, exponential, Weibull, and Gompertz) the least satisfactory was the [redacted].

Appendix 14 Comparison of TTP results reported in different documents

TTP results from submission, SHARP trial report and SHARP publication.

	Median time TTP, weeks (95% CI)			Hazard ratio, sorafenib v. placebo (95% CI)		
	Independent assessment	Investigator assessment	Hybrid assessment	Independent assessment	Investigator assessment	Hybrid assessment
SUBMISSION sorafenib	24 (18, 30)	17 (13, 18)	[redacted]	.58 (.45,.74)	.6889 (.5634,.8423)	[redacted]
TRIAL REPORT sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PUBLICATION sorafenib	24	NR	[redacted]	.58 (.45, .74)	NR	NR
SUBMISSION placebo	12.3 (11.7, 17.1)	11.9 (11.1, 12.4)	[redacted]			
TRIAL REPORT placebo	[redacted]	[redacted]	[redacted]			
PUBLICATION placebo	12	NR	NR			

Appendix 15 Comparison of Independent and Hybrid TTP

Appendix 16 Comparison of the ERG and trial report independent TTP analyses.

The table below compares the parameters from survival analysis of TTP presented in the trial report with those obtained from the ERG survival analysis of individual patient data extracted from the trial report. All parameters correspond.

Independent assessment	PLACEBO (N=303)		SORAFENIB (N=299)		HR
	Median (days)	*Number censored*	Median (days)	*Number censored*	
ERG analysis	86	*147*	168	*192*	0.576
Trial report/ submission	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

The following figure superimposes the ERG K-M plot over that presented in the trial report; an exact correspondence was observed.

Appendix 17 Subgroup analysis of overall survival for HCV positive patients

The submission presented a Kaplan-Meier plot for overall survival of hepatitis virus C positive patients from SHARP (sorafenib group n=93 placebo group n=85). This is shown in below.

Comment

The overall survival benefit of sorafenib for this subgroup is evident.

Appendix 18 Supporting RCT data presented in the submission

From the submission:

Supporting RCT data Asia Pacific Study (23,36)

Results from the SHARP study are supported by the Asia-Pacific RCT, which showed superiority for sorafenib over placebo for overall survival (OS) and time to progression (TTP), thus demonstrating efficacy in a different population in patients with different leading aetiologies.

In the Asia Pacific study (study 11849), 226 patients from China, Korea and Taiwan with advanced HCC were randomised to receive either sorafenib (n=150) or placebo (n=76). The study was designed in parallel with the SHARP study and inclusion and exclusion criteria were similar.

Sorafenib significantly prolonged overall survival (OS), despite more advanced disease compared to patients enrolled in SHARP. The median OS was 18.2 weeks in placebo patients compared to 28.2 weeks in sorafenib--treated patients. The hazard ratio for this improvement was 0.68 (P=0.014) representing a 47% increase in OS with sorafenib. The 6-month overall survival rate was 53.3% in the sorafenib group and 36.7% in the placebo group.

The absolute increase in median overall survival rates were smaller (although the HR differentials almost match) when compared to results from the SHARP trial. This is most likely explained by the fact the patients in the Asia-Pacific trial had a poorer status and more advanced tumour stage as exemplified by a higher rate in extrahepatic spread (48). This is in accordance with the SHARP data where patients with poorer status (43) and extra-hepatic spread and/or macroscopic vascular invasion (44) also showed lower survival rates, although the significant difference and benefit between sorafenib and placebo was maintained throughout the subgroups.

The Asia-Pacific study also measured TTSP using the Functional Assessment of Cancer Therapy - Hepatobiliary Symptom Index (FSHI8) questionnaire with similar results (15.2 vs 14.8 weeks) (23) . Reasons for lack of differences between the arms are highlighted elsewhere.

Sorafenib significantly prolonged TTP in the Asia-Pacific study. Median TTP was 6.1 weeks in placebo patients and 12.2 weeks with sorafenib. The hazard ratio for this improvement was 0.57 (P=0.0005) representing a 76% improvement in TTP. Sorafenib was well-tolerated and had manageable side effects. The reduced benefit compared to the SHARP study, also seen with overall survival, can again be explained by patients in the Asian trial having poorer performance status and more advanced tumour stage.

DCR was 35% [95% CI 28,34] in the sorafenib arm and 16% [95% CI 8, 26] in the patients receiving placebo. Five of 150 patients in the sorafenib group (3.3%) achieved a partial response and 81 of 150 patients (54%) had stable disease. In the placebo group, one patient achieved a partial response (1.3%) and 21 patients had stable disease (27.6%).

In the preplanned subgroup analysis, sorafenib provided clinical benefit in all groups, despite some patients having characteristics associated with poor prognosis e.g. extrahepatic spread, macroscopic vascular invasion and HBV infection.

Appendix 19 Eastern Cooperative Oncology Group performance status

Criteria taken from the submission appendix are shown

Grade	Description
0	Fully active, able to carry on all pre-diseases performance without restriction (Karnofsky 90-100)
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature (eg, light housework, office work). (Karnofsky 70-80)
2	Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours. (Karnofsky 50-60)
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours. (Karnofsky 30-40)
4	Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair. (Karnofsky 10-20)

Appendix 20 Effectiveness of sorafenib in Child-Pugh grade B advanced HCC

The ERG requested clarification regarding the effectiveness of sorafenib for Child-Pugh grade B relative to grade A patients. The response received is shown in below followed by the ERG comments.

The SHARP data population have not been refined to address the decision problem. Based on the spectrum of underlying liver conditions, the population studied within the SHARP trial reasonably represented the HCC condition encountered in the western world with the exception that enrolment was predominantly Child Pugh A score. However based on exploratory analyses, the FDA highlighted that clinical benefit with sorafenib did not appear to depend on underlying liver disease or Child Pugh score A or B.

The data for use of sorafenib in Child Pugh B patients is limited; however existing pharmacokinetic, safety and efficacy data make it reasonable for consideration as an option in this setting.

- Yau et al in a phase II trial looked at factors predictive of clinical benefit with sorafenib. The authors assessed 36 patients with Child Pugh A and 13 patients with Child Pugh B and concluded that Child Pugh status was not predictive of clinical benefit with sorafenib

- Shim et al assessed 34 patients with Child Pugh A and 23 with Child Pugh B. The authors showed that Child Pugh class had no significant effect on time to progression (TTP) after treatment with sorafenib.

Within our internal dataset we have analysed [redacted]

Based on these data we can conclude the following:

- Patients with HCC and Child-Pugh B liver dysfunction have similar sorafenib exposures as those with Child-Pugh A liver function
- Incidence rates of common sorafenib toxicities are similar between HCC patients with Child-Pugh B and Child-Pugh A
    - The exception to this general conclusion appears to be sorafenib-related increases in bilirubin occurring in Child Pugh B patients and patients with baseline hyperbilirubinemia. The bilirubin elevations appear to be isolated and not associated with other clinical sequelae or elevations of transaminases. Additional analyses are ongoing to understand this preliminary finding further.
- Comparison of pERK staining intensities from baseline tumor samples of HCC patients with Child-Pugh B were similar to those from HCC patients with Child-Pugh A, suggesting that HCC tumor biology was not dramatically different for both groups

[redacted]

Comment

1. The Shim et al and Yau et al studies were from Korea and China respectively. Shim was a retrospective analysis of 57 consecutive patients of whom 34 and 23 were Child-Pugh grades A and B respectively. Yau was an open label Phase 2 prospective study with 51 advanced HCC patients (BCLC rating not reported) of whom 36 were Child-Pugh grade A and 13 grade B and 2 grade C (an abstract of this study reported different numbers: N= 58, C-P A = 30 and C-P B = 15).

2. In Shim 43 patients were evaluable for disease response; none achieved a complete response, 3 (all Child-Pugh grade A) achieved a partial response and 20 achieved stable disease but the proportions that were classes A and B were not reported. This publication contains the following apparently contradictory statements: *"the marginal results we observed with sorafenib may have been because of the inclusion of many patients with worsening underlying cirrhosis (Child-Pugh grade B) or infiltrating, far advanced HCC"* and *"surprisingly however Child-Pugh class …could not be correlated with disease stabilisation".*

3. In Yau median overall survival was 13 weeks, considerably less than in the Asia-Pacific study (28 weeks) in which the population were almost exclusively (97%) Child-Pugh grade A. There were no complete responses, 4 (8%) partial responses and 9 patients (18%) achieved stable disease. On this basis 13 patients (26%) were calculated to have derived clinical benefit from sorafenib (the abstract of this study reported different numbers for response: 4 partial responses and 5 stable disease responses). Of 38 Child-Pugh grade A patients 7 (19%) derived benefit while of 13 Child-Pugh grade B+C patients 5 (40%) derived benefit. The CHI squared P value for difference between Child-Pugh grades was 0.125.

4. The populations in these studies were smaller than that in the Abou-Alfa uncontrolled study and less likely to be similar to that in the UK.

Appendix 21 Resource use and costs tables reproduced from appendix 13 of submission

RESOURCE USE AND UNIT COSTS

Mean (standard deviation)

Table 1: Monthly Physician visits

Medical Contact	First line – no progression with sorafenib	First line – no progression with BSC	First-line treatment continued – post progression with sorafenib	BSC - post progression
1. Oncologist	0.75 (0.50)	0.38 (0.48)	1.00 (0)	0.38 (0.48)
2. Hepatologist	0.17 (0.19)	0.50 (0.58)	0.58 (0.96)	0.50 (0.58)
3. Macmillan Nurse	0.50 (0.58)	0.50 (0.58)	1.00 (1.15)	1.00 (1.15)
4. Gastroenterologist	0.08 (0.17)	0.25 (0.50)	0.13 (0.25)	0
5. Radiologist	0.08 (0.17)	0	0	0
6. Clinical Nurse Specialist	0.50 (0.58)	0.13 (0.25)	0.50 (0.58)	0.25 (0.50)
7. Palliative Care Physician / Nurse	0.13 (0.25)	0	1.00 (2.00)	0.75 (0.96)

Table 2: Monthly laboratory tests

	First line – no progression with sorafenib		First line – no progression with BSC		First-line treatment continued – post progression with sorafenib		BSC - post progression	
Laboratory Test	**Mean proportion of patients utilising**	**Mean # of tests**	**Mean proportion of patients utilising**	**Mean # of tests**	**Mean proportion of patients utilising**	**Mean # of tests**	**Mean proportion of patients utilising**	**Mean # of tests**
1. AFP test	0.75 (0.50)	0.83 (0.29)	0.75 (0.50)	0.83 (0.29)	0.38 (0.48)	1.00 (0)	0.38 (0.48)	1.00 (0)
2. Liver function test	0.50 (0.58)	0.67 (0.47)	0.50 (0.58)	0.67 (0.47)	0.25 (0.50)	1.00 (0)	0.25 (0.50)	1.00 (0)
3. INR	0.50 (0.58)	0.67 (0.47)	0.50 (0.58)	0.67 (0.47)	0	0	0	0
4. Complete blood count	0.75 (0.50)	1.00 (0)	0.75 (0.50)	1.00 (0)	0.50 (0.58)	1.00 (0)	0.50 (0.58)	1.00 (0)
5. Biochemistry	0.50 (0.58)	1.00 (0)	0.50 (0.58)	1.00 (0)	0.25 (0.50)	1.00 (0)	0.25 (0.50)	1.00 (0)
Other								
1. Endoscopy	0.25 (0.50)	0.33 (0)	0.25 (0.50)	0.33 (0)	0	0	0	0

Table 4: Monthly hospitalisation

Hospitalisation	First line – no progression with sorafenib	First line – no progression with BSC	First-line treatment continued – post progression with sorafenib	BSC - post progression
Proportion of patients requiring hosp	0.46 (0.31)	0.39 (0.35)	0.42 (0.32)	0.48 (0.30)
Number of hospitalisations	0.16 (0.10)	0.16 (0.10)	0.32 (0.21)	0.40 (0.32)
General ward stay (days)	2.5 (2.89)	7.00 (0)	5.50 (4.20)	5.67 (5.13)
Proportion of A&E admissions	0.11 (0.16)	0.18 (0.18)	0.14 (0.15)	0.35 (0.15)

Table 5: Monthly social care

Social Care	First line – no progression with sorafenib			First line – no progression with BSC			First-line treatment continued – post progression with sorafenib			BSC - post progression		
	Proportion utilising	Mean days	Proportion funded by NHS	Proportion utilising	Mean days	Proportion funded by NHS	Proportion utilising	Mean days	Proportion funded by NHS	Proportion utilising	Mean days	Proportion funded by NHS
1. Residential care	0.02 (0.04)	0	0	0.00	0.00	0.00	0.03 (0.06)	0	0	0.03 (0.05)	6.43 (0)	1.00 (0)
2. Day care	0.02 (0.04)	0	0	0.00	0.00	0.00	0.03 (0.06)	0	0	0.23 (0.26)	5.36 (1.51)	0.00
3. Home care	0.07 (0.05)	4.00 (0)	0.50 (0)	0.09 (0.10)	12.86 (0)	1.00 (0)	0.27 (0.25)	4.00 (0)	0.50 (0)	0.28 (0.10)	12.86 (8.57)	1.00 (0)
4. Hospice	NA	NA	NA	0.09 (0.10)	6.47 (0.05)	0.15 (0.21)	0.10 (0.10)	1.00 (0)	0.50 (0.71)	0.18 (0.10)	14.00 (3.50)	0.43 (0.51)

Table 6: Monthly follow-up associated with hospitalisation

Follow-up visit	First line – no progression with sorafenib	First line – no progression with BSC	First-line treatment continued – post progression with sorafenib	BSC - post progression
1. Specialist	0.25 (0.50)	1.00 (1.41)	0.67 (0.58)	3.00 (0)
2. GP	1.50 (2.38)	0.67 (1.15)	0.50 (0.58)	1.50 (2.12)
3. Nurse	1.75 (2.36)	2.00 (2.83)	1.00 (1.00)	2.00 (2.83)

Table 7: Unit costs for medical staff visits

Resource use item	Unit	Mean unit cost (£)	Mean unit cost (2008 £)	Source	
Oncologist	per contact	151.00	156.04	NHS National Schedule of Reference Costs 2006-07	Consultant Led First Attendance Outpatient Face to Face (TCLFASFF); specialty code 800; Clinical Oncology (attendance without treatment) Total Attendances; LQ £71; UQ £243
Hepatologist	per contact	191.00	197.38	NHS National Schedule of Reference Costs 2006-07	Consultant Led First Attendance Outpatient Face to Face (TCLFASFF); specialty code 306; Hepatology Total Attendances; LQ £150; UQ £251
Specialist Nurse	per hour	30.00	31.00	PSSRU 2007	Schema 9.4 Nurse specialist (Community); costs including qualifications
GP	per consultation	34.00	35.14	PSSRU 2007	Schema 9.8b General practionner, per surgery consultation lasting 11.7 min; costs including qualifications
District Nurse	per hour	30.00	31.00	PSSRU 2007	Schema 9.1 Community nurse (district nurse); costs including qualifications
Palliative care team (1 consultant, 4 nurses, 1 social worker)	per contact	124.00	128.14	NHS National Schedule of Reference Costs 2006-07	Consultant Led First Attendance Outpatient Face to Face (TCLFASFF); speciality code 191;Pain Management Total Attendances ; LQ £202; UQ £255
Specialist visit	per half hour	87.50	90.42	PSSRU 2007	schema 14.4 Consultant: medical, per contract hour; costs including qualifications
Dietician	per contact	32.00	33.07	PSSRU 2007	Schema 12.4 Dietician; costs including qualifications

Table 8: Unit costs for laboratory and radiology tests

Resource use item	Mean unit cost per test (£)	Mean unit cost per test(2008 £)	Source
Laboratory tests			
AFP test	7.12	7.36	Newcastle Upon Tyne Hospitals NHS Trust Diagnostic Services Tariff 2006/2007
Liver Function test	5.90	6.10	Meavy Clinic Tariff Charges April 2006-March 2007
INR	3.84	3.97	Newcastle Upon Tyne Hospitals NHS Trust Diagnostic Services Tariff 2006/2007
Complete blood count	2.29	2.37	Newcastle Upon Tyne Hospitals NHS Trust Diagnostic Services Tariff 2006/2007 (Haematology Laboratory Services - Full blood count)
Complete metabolic panel	96.40	99.62	Newcastle Upon Tyne Hospitals NHS Trust Diagnostic Services Tariff 2006/2007
Radiological tests			
CT scan: abdominal	156.00	161.21	Newcastle Upon Tyne Hospitals NHS Trust Diagnostic Services Tariff 2006/2007
MRI: abdominal	230.00	237.68	Newcastle Upon Tyne Hospitals NHS Trust Diagnostic Services Tariff 2006/2007
Ultrasound: abdominal	96.00	99.21	Newcastle Upon Tyne Hospitals NHS Trust Diagnostic Services Tariff 2006/2007

Table 9: Unit costs for hospitalizations and social care

Resource use item	Unit	Mean unit cost (£)	Mean unit cost (2008 £)	Source
ICU	day	1410.00	1457.08	NHS National Schedule of Reference Costs 2006-07
General Ward	day	323.00	357.20	CIFPA 2004-2005
A&E Admission	admission	90.00	93.00	NHS National Schedule of Reference Costs 2006-07
Residential Care	per day	99.00		Marillac, nursing home,2006
Day Care	per day	130.00	105.58	NHS National Schedule of Reference Costs 2006-07
Home Care	per day	74.00	134.34	NHS National Schedule of Reference Costs 2006-07
Hospice	per episode	84.00	76.47	NHS National Schedule of Reference Costs 2005-06

Table 10: Unit costs for other tests

Resource use item	Mean unit cost per test (£)	Mean unit cost per test(2008 £)	Source
Microbiological examination	23.33	24.11	UCL lab tariff 2007
IV rehydration	2.10	2.10	BNF, 2008
Urea and electrolytes (blood urea nitrogen)	5.90	6.10	Meavy Clinic Tariff Charges April 2006-March 2007
Urea and electrolytes (urine)	23.00	23.00	Mullhaven Medical Laboratory 2008
Endoscopy	750.00	775.04	GI Endoscopy. Meavy Clinic Tariff Charges April 2006-March 2007

Table 11: Unit costs for medications

Medication	Mean drug cost, package price (£)	Source
1. Ferrous sulphate (200mg)	2.10	BNF 2008
2. Dexamethasone	3.27	BNF 2008
3. Loperaminde	0.61	BNF 2008
4 Codeine	0.97	BNF 2008
5. Cyclizine	1.48	BNF 2008
6. Metoclopramide	0.44	BNF 2008
8. Domperidone	1.36	BNF 2008
9. Paracetamol	0.17	BNF 2008
10. Cholestyramine	17.28	BNF 2008
11. Atenolol	0.30	BNF 2008
12. Morphine sulfate	1.87	BNF 2008

Appendix 22 Quality Assessment using ScHARR_TAG economic modelling checklist

Title

Single Technology Appraisal (STA) of sorafenib (Nexavar®) for the treatment of hepatocellular carcinoma (HCC)

A statement of the problem

Yes, a statement of the problem has been given.

A discussion of the need for modelling

No discussion of the need for modelling was included.

A description of the relevant factors and outcomes

A clinical pathway and outcomes on overall survival and time to progression are provided but there are outcomes related to health related quality of life that are not clearly reported in the model.

A description of model including: type of model; time frame; perspective; and setting

The model is a lifetime state transition (Markov) model. This is appropriate to the decision problem. The time frame is 14 years. The model was populated using the SHARP trial data and extrapolating the 72-week data to the longer timeframe using a lognormal distribution.

Patients enter the model if they are diagnosed with advanced HCC, and have failed or are unsuitable for surgical or locoregional therapies. They are thus suitable to receive first-line treatment. They continue to be modelled until they finally exit the evaluation due to death.

A description of data sources, with description of respective strengths and weaknesses

Clinical data come from the phase III SHARP study, said to be the largest and most relevant data source for the decision problem being addressed. A description of those data is provided but strengths and weaknesses have not been discussed.

Data on adverse events and cost rely on expert opinions.

Key assumptions relating to model structure and data stated

Yes, key assumptions relating to the model structure and data used are stated.

Disease specific factors included within modelling (Items to be specified in conjunction with expert clinical input)

Validation

There was no validation of the model results in the submission.

Results

Model results are reported in the submission in an appropriate format.

Sensitivity analysis results

Results of sensitivity analyses are reported in the submission.

References

1 Llovet JM, Burroughs A, Bruix J. Hepatocellular carcinoma. *Lancet* 2003; **362**(9399):1907-1917.

2 Ryder SD. Guidelines for the diagnosis and treatment of hepatocellular carcinoma (HCC) in adults. *Gut* 2003; **52 Suppl 3**:iii1-iii8.

3 Llovet JM, Di Bisceglie AM, Bruix J, Kramer BS, Lencioni R, Zhu AX, *et al*. Design and endpoints of clinical trials in hepatocellular carcinoma. *J Natl Cancer Inst* 2008; **100**(10):698-711.

4 Ryder,SD. UK Guidelines for the management of suspected hepatocellular carcinoma (HCC) in adults (Revised guidelines, in draft, British Society of Gastroenterology). (Personal communication)

5 Camma C, Di M, V, Cabibbo G, Latteri F, Sandonato L, Parisi P, *et al*. Survival of patients with hepatocellular carcinoma in cirrhosis: a comparison of BCLC, CLIP and GRETCH staging systems. *Aliment Pharmacol Ther* 2008; **28**(1):62-75.

6 Kung JW, MacDougall M, Madhavan KK, Garden OJ, Parks RW, Kung JWC, *et al*. Predicting survival in patients with hepatocellular carcinoma: a UK perspective. *European Journal of Surgical Oncology* 2007; **33**(2):188-194.

7 European Medicines Agency (EMEA). Nexavar - European Public Assessment Report (EPAR), Scientific discussion document (HCC submission). Access date: 4 Dec. 2008.

8 Fleming ID. AJCC/TNM cancer staging, present and future. *J Surg Oncol* 2001; **77**(4):233-236.

9 Llovet JM, Bru C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. *Semin Liver Dis* 1999; **19**(3):329-338.

10 Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, *et al*. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med* 2008; **359**(4):378-390.

11 Cheng AL, Kang YK, Chen Z, Tsao CJ, Qin S, Kim JS, *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.

12 Abou-Alfa GK, Schwartz L, Ricci S, Amadori D, Santoro A, Figer A, *et al*. Phase II study of sorafenib in patients with advanced hepatocellular carcinoma. *J Clin Oncol* 2006; **24**(26):4293-4300.

13 Abou-Alfa GK, Johnson P, Knox J, Davidenko I, Lacava J, Leung T, *et al*. Final results from a phase II (PhII), randomized, double-blind study of sorafenib plus doxorubicin (S+D) versus placebo plus doxorubicin (P+D) in patients (pts) with advanced hepatocellular carcinoma (AHCC). *J Clin Oncol ASCO 2008 meeting: Gastrointestinal cancers symposium* 2008;Abstract No: 128.

14 National Cancer Institute. CTCAE v3.0 Online Instructions and Guidelines (Updated August 9, 2006). *https://webapps ctep nci nih gov/webobjs/ctc/webhelp/welcome_to_ctcae htm* 2006.
15 Abou-Alfa,GK, Amadori,D, Santoro,A, Figer,A, De Greve,J, Lathia,C, *et al.* Is sorafenib safe and effective in patients with hepatocellular carcinoma (HCC) and Child-Pugh B cirrhosis. (Abstract) *J Clin.Oncol.* **26**[15S (May 20 suppl)]
16 Gish RG, Porta C, Lazar L, Ruff P, Feld R, Croitoru A, *et al*. Phase III randomized controlled trial comparing the survival of patients with unresectable hepatocellular carcinoma treated with nolatrexed or doxorubicin. *J Clin Oncol* 2007; **25**(21):3069-3075.
17 Spruance SL, Reid JE, Grace M, Samore M. Hazard ratio in clinical trials. *Antimicrob Agents Chemother* 2004; **48**(8):2787-2792.
18 Chu D, Lacouture ME, Fillos T, Wu S. Risk of hand-foot skin reaction with sorafenib: a systematic review and meta-analysis. *Acta Oncol* 2008; **47**(2):176-186.
19 Wu S, Chen JJ, Kudelka A, Lu J, Zhu X. Incidence and risk of hypertension with sorafenib in patients with cancer: a systematic review and meta-analysis. *Lancet Oncol* 2008; **9**(2):117-123.
20 Tuominen U. BMHJSSLMPPHKRPSH. The effect of co-morbidities on health-related quality of life in patients placed on the waiting list for total joint replacement. *Health Quality Life Outcomes* 2007; **5**(16).
21 Muszbek N. Economic Evaluation of Sorafenib vs best supportive care in Hepatocellular Carcinoma. *Journal of Clinical Oncology* 2008; **26**(15 S May 20 Supplement):6527.
22 Muszbek N, Shah S, Carroll S, McDonald H, Dale P, Maroun J, *et al*. Economic evaluation of sorafenib in the treatment of hepatocellular carcinoma in Canada. *Curr Med Res Opin* 2008; **24**(12):3559-3569.
23 Briggs A, Claxton K, Sculpher M. Decision Modelling for Health Economic Evaluation. *Oxford University Press (2006)* 2006.
24 Lopez PM, Villanueva A, Llovet JM. Systematic review: evidence-based management of hepatocellular carcinoma - an updated analysis of randomized controlled trials. *Alimentary Pharmacology & Therapeutics* 2006; **23**:1535-1547.