NICE National Institute for Health and Care Excellence

CDF Rapid Reconsideration

Sorafenib for advanced hepatocellular carcinoma (review of TA189) [ID1012]

Committee Papers

NICE National Institute for Health and Care Excellence

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

CDF Rapid Reconsideration

Sorafenib for advanced hepatocellular carcinoma (review of TA189) [ID1012]

Contents:

Any information supplied to NICE which has been marked as confidential, has been redacted. All personal information has also been redacted.

Slides for public

Sorafenib for treating advanced hepatocellular carcinoma

Cancer Drug Fund Reconsideration of TA189

Second CDF committee meeting: 29 November 2016, London

First CDF committee meeting (July 2016, Manchester)

Evidence Review Group: NICE Decision Support Unit (DSU), University of Sheffield

Chair: Amanda Adler

Lead Team (NICE TA189, Committee C): Matt Stevenson, Philip Rutledge

NICE Technical Team: Wendy Gidman, Martyn Burke, Frances Sutcliffe

Company: Bayer

History of Appraisal

2009 Apr	1st appraisal committee meeting
	Appraisal Consultation Document (ACD) – not recommended
2009 Jun	2nd appraisal committee meeting
2009 Aug	3rd appraisal committee meeting
	2nd ACD – not recommended
2009 Oct	4th appraisal committee meeting: 14 October 2009
	Final appraisal determination issued: not recommended
2010 Feb	Appeal: 4 points. All dismissed
2010 May	Final guidance reissued: not recommended
2016 Jul	1st CDF reconsideration meeting - New price and new data to validate time beyond trial
	ACD: not recommended
Today	2nd CDF reconsideration meeting

Preview - Issues for discussion

- Did the company adequately address the outstanding issues of:
  - Extrapolation?
    - What is the most valid statistical function for extrapolating overall survival?
  - Treatment duration?
    - Source of data? Time-to-progression or time-to-treatment discontinuation?
    - Way to extrapolate?
  - Wastage?
- What is the committee's preferred resource use data?

Sorafenib and decision problem TA189

Sorafenib	
Marketing authorisation	‘for the treatment of hepatocellular carcinoma’ (and renal cell and thyroid carcinoma)
Mechanism	‘Multikinase’ inhibitor
Administration	Oral – twice daily
Indications	Renal cell carcinoma, differentiated thyroid carcinoma

Decision problem	
Population	Patients with advanced stage hepatocellular carcinoma who have failed or are unsuitable for surgical or loco-regional therapies
Intervention	Sorafenib
Comparators	Best supportive care

Evidence Randomised Controlled Trial

Sorafenib Hepatocellular Carcinoma Assessment Randomized Protocol (SHARP)

- 602 patients
- Not previously treated
- Life expectancy ≥ 12 weeks
- ECOG* 0 to 2
- Child-Pugh function** grade A

Sorafenib 400 mg

Placebo

2 1° endpoints: overall survival and time to symptomatic progression

Note:
Treat to radiographic progression (RECIST); 7.7% continued beyond
Trial stopped early
Utility: Functional assessment of cancer therapy - hepatobiliary [FACT-hep] mapped to EQ-5D

- **Eastern Cooperative Oncology Group performance status (0: fully active to 5: dead)*
- ***Child-Pugh based on serum bilirubin, serum albumin, prothrombin time, ascites, enchephalopathy; 96% of SHARP Child-Pugh function** grade A*

Committee conclusions TA189 final guidance

Population	Child–Pugh grade A liver function + good performance
Clinical effectiveness	Radiological disease progression differed by who assessed it;11.7 weeks longer when assessed independently, 5.1 weeks by investigator for sorafenib compared with placebo
Cost effectiveness	Company chose log-normal to extrapolate overall survival and progression free survival; key drivers of cost effectiveness
	Log-normal provided slightly better fit to observed data than Weibull, it could not be accepted as the definitive function to extrapolate beyond the data, therefore both distributions should be considered
	1st model included treating beyond progression (per SHARP); later model with PAS did not include costs beyond progression
	In SHARP a cycle lasted 31.5 days, but in model lasted 30.4 days
	Complex patient access scheme introduced
ICER	Lower range of ICER was £52,600 per QALY gained; key drivers OS, PFS, utility
End of life	Yes. Increased median survival >2.8 months, and the company's model predicted a mean gain in overall survival of 6.1 months

CDF reconsideration – 1st meeting

New evidence	
Effectiveness	1 GIDEON observational study; (‘data on file’) 2 UK real-life clinical data; (Palmer et al 2013)
Price of sorafinib	New price (Commercial Medicines Unit)
TA189 conclusions	
Radiological disease progression differed by who assessed it	All results assessed by investigator not independently (centrally). Independent was main measure in SHARP then changed
Extrapolation key driver; curves other than log-normal fit extrapolated portion better; consider Weibull and log-logistic	New observational data to validate company’s choice of lognormal for overall survival
Treating beyond progression	Included (again)
Cycle length	Amended

GIDEON
multinational post-marketing uncontrolled safety study n=3,202

Unresectable HCC, candidates for systemic therapy, life expectancy of > 8 weeks

All treated with sorafenib

1. Safety
2. Overall survival

GIDEON uncontrolled safety study

"of 3213 patients, median overall survival [AIC] days"

Overall survival	Median
SHARP Weibull	[AIC]
SHARP Lognormal	[AIC]
GIDEON	[AIC]*

*Kaplan Meier OS data ITT analysis visual inspection (approximate survivors)

* Median 330 or 300?

	80%
SHARP Weibull	[AIC]
SHARP Lognormal	[AIC]
GIDEON	[AIC]*

Source: Figure 3 of company's submission

Numbers at risk not provided

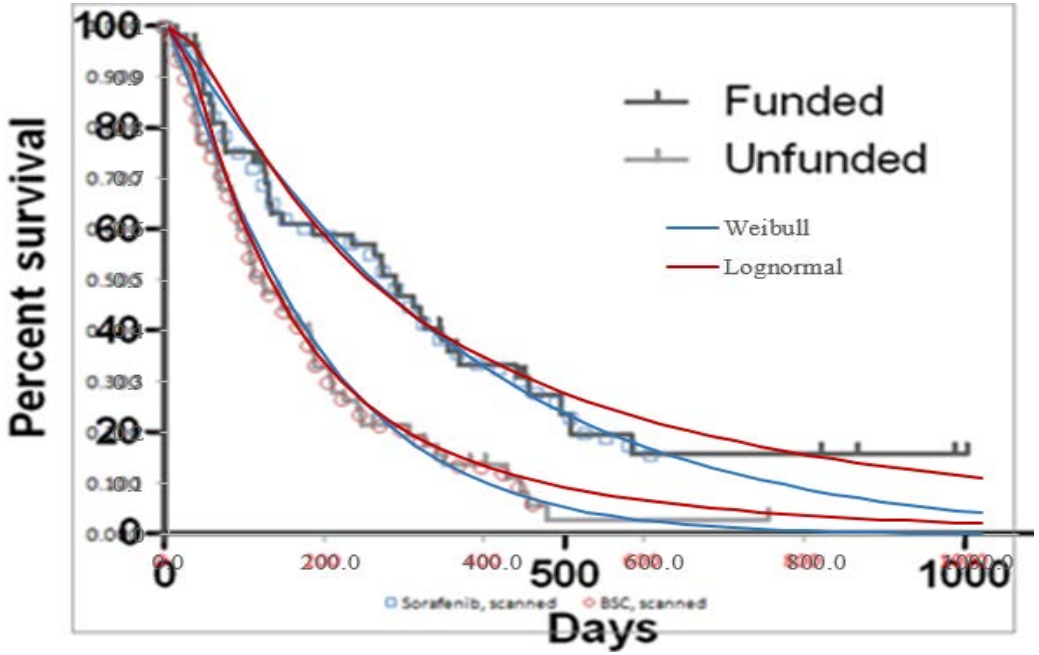
- ERG: Important differences between GIDEON and SHARP *e.g.* GIDEON 52% patients Barcelona Clinic Liver Cancer advanced disease vs SHARP 82%

Palmer *et al.* 2013 retrospective UK observational study

- Comparing ‘funded’ (n=57) vs. ‘unfunded’ (n=76)
- Numbers at risk not presented
- Statistical methods not presented

ERG’s critique

- High risk of bias (not suitable for decision-making)
- Small number (n=57 on sorafenib)
- On inspection, Weibull fits better
- Plateau at tail = high uncertainty
- Weibull likely to be well within confidence intervals



Source: Figure 1, page 16 of ERG report

ERG's exploratory analyses

Source: table 16, page 46 of the ERG report		QALYs		Costs(£)		ICER (£/QALY)
		Total	Inc.	Total	Inc.	
Company's base case (investigator assessment + updated costs + resource use)	BSC	[AIC]		[AIC]		
	Sorafenib	[AIC]	[AIC]	[CIC]	[CIC]	£39,162
Weibull	BSC	[AIC]		[AIC]		
	Sorafenib	[AIC]	[AIC]	[CIC]	[CIC]	£58,287
Independent assessment	BSC	[AIC]		[AIC]		
	Sorafenib	[AIC]	[AIC]	[CIC]	[CIC]	£45,468
Using pooled resource use estimates						
Lognormal	BSC	[AIC]		[AIC]		
	Sorafenib	[AIC]	[AIC]	[CIC]	[CIC]	£45,372
Weibull	BSC	[AIC]		[AIC]		
	Sorafenib	[AIC]	[AIC]	[CIC]	[CIC]	£66,873
ERG's preferred assumptions (independent assessment + pooled resource use estimates)						
Lognormal	BSC	[AIC]		[AIC]		
	Sorafenib	[AIC]	[AIC]	[CIC]	[CIC]	**£51,208**
Weibull	BSC	[AIC]		[AIC]		
	Sorafenib	[AIC]	[AIC]	[CIC]	[CIC]	£71,276

CDF reconsideration 1st committee meeting: Committee's key conclusions

Validation of the overall survival extrapolation	Log normal fitted GIDEON data better than Weibull function, but because the populations between SHARP and GIDEON differed in the company's current analysis of GIDEON, the Weibull function still had some plausibility. Palmer likely confounded.
Treatment duration	Estimate the duration of treatment with sorafenib based on patient-level data from SHARP rather than using the proxy measure of progression-free survival
Resource use	Original and revised estimates of resource use should be pooled because they are based on the opinion of a small number of clinicians
Treatment wastage	Account for drug wastage
Most plausible ICER	Would not be lower than £51,200 per QALY gained, and could be higher taking into account the uncertainty in extrapolating overall survival and treatment costs

ACD recommendation

Sorafenib is **not recommended** for treating advanced hepatocellular carcinoma in adults when surgical or locoregional therapies have failed or are not suitable

ACD consultation responses

- Consultees:
  - Bayer (sorafenib)
  - British Liver Trust
  - Department of Health (no comments)
  - NHS England
- Commentators:
  - National Cancer Research Institute Hepatobiliary
- Web comments
  - NHS professional x1
  - Healthcare other x1

Comments on ACD: British Liver Trust

- People with advanced HCC have a very poor prognosis and given lack of alternatives, sorafenib fulfils a unmet clinical need
- Only systemic treatment to increase survival and quality of life
- Relatively small population, so overall cost to the NHS small
- Patient access in England has been possible through the CDF and 968 patients had sorafenib from April 2013–September 2015
- At the committee meeting the company gave a “very poor representation of the desperate need for sorafenib”
- Provision of sorafenib should be maintained for patients with advanced HCC, Child-Pugh A and performance status <2

Comments on ACD: National Cancer Research Institute Hepatobiliary Subgroup

- An audit involving 15 centres and 448 patients suggests:
  - median time on treatment in UK is 15.6 weeks (not 24)
  - actual daily dose is 590 mg rather than 800 mg
  - sorafenib is more effective in people with Child-Pugh A (than Child-Pugh B) and good performance status (than poor performance status)
- Suggests sorafenib may be more cost effective in people with Child-Pugh A disease and good performance status
  - *n.b. same population as SHARP trial*
- Sorafenib is the only approved drug for advanced HCC and is defined as the current standard-of-care in international guidelines
  - *n.b. guidelines vary; do not take into account cost*
- Removing access bad for patient care and future research

NHS England

- ‘NHS England urges the NICE Technology Appraisal Committee to use the same patient source of information on which to base their preferred estimates of both treatment duration and overall survival.’
- ‘Trusts will not waste sorafenib.’
- ‘Regorafenib has been shown to offer a survival benefit as second line treatment and there are other promising drugs such as cabozantinib and nivolumab in the pipeline. If any marketing authorisations state that any of these new drugs can only be used after previous treatment with sorafenib, then these new drugs will be disqualified from NICE appraisal’ if NICE does not recommend sorafenib.
- ‘In the past, the Cancer Drugs Fund placed a special emphasis on those drugs that were the only proven systemic therapies for a particular cancer. This latter thinking now plays no part in NHS England in the decision making’
- ‘What matters now is whether sorafenib is cost effective in this indication or not’

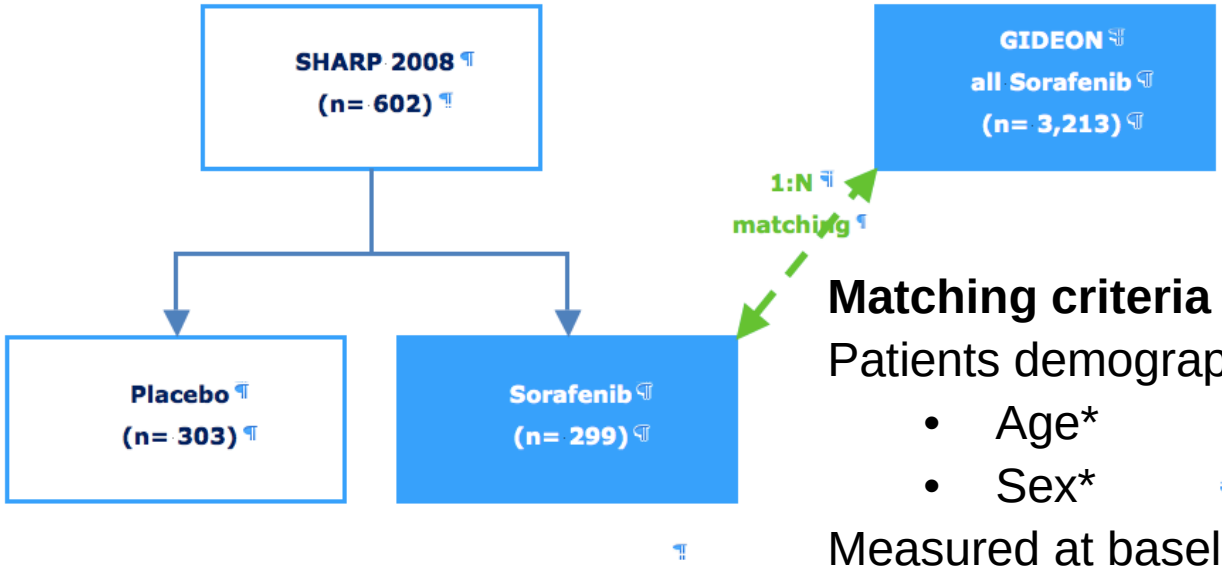
Comments on ACD: Web comments

- Individual patient data (IPD) meta-analysis of 3 large prospective randomised trials ‘alternative therapies’ vs. sorafenib; n = 3256
- Suggests that that the drug may be more cost-effective in people with hepatitis C
- “Little evidence of impact in other aetiologically-defined sub-groups”
- Median unadjusted survival is 12.6 (11.15, 13.8) months for sorafenib
- With sorafenib’s funding currently under review, patients and physicians are left with very few treatment options in intermediate and advanced HCC

Comments on ACD: Company

- New evidence to address uncertainties:
  1. Extrapolating overall survival: Analysis of matched GIDEON population to SHARP to inform distribution
  2. Treatment duration: based on individual patient level data from SHARP
  3. Treatment wastage – evidence from 2 hospitals
- Contests committee's decision that Palmer confounded
- Pooling of resource use data not correct because 'clinical practice has changed'
- Further discount to Commercial Medicines Unit (CMU)
  - List price £3575.56 per 28 day supply vs. CMU price £[CIC] per 28 day (a further [CIC]% discount)

GIDEON – matching to SHARP Bayer Propensity Scoring



Source page 37, Figure 7, of company response to ACD

Matching criteria of interest *=matched

Patients demographics:

- Age*
- Sex*

Measured at baseline:

- Child-Pugh status (Class A, Class B)*
- BCLC (stage B, Stage C)
- ECOG Status (Status 0/1, Status 2)*
- Vascular invasion presence
- Extrahepatic spread presence
- Hepatitis B presence
- Hepatitis C presence

BDLC = Barcelona Clinic Liver Cancer
ECOG = Eastern Cooperative Oncology Group

Baseline characteristics SHARP/GIDEON in propensity scoring matched sample

Variable mean and standard deviation or percentage	SHARP_n (N = 299)		GIDEON_n (N= 895)	
Age	[AIC]	[AIC]	[AIC]	[AIC]
Male	[AIC]	[AIC]	[AIC]	[AIC]
ECOG				
- 0 or 1	[AIC]	[AIC]	[AIC]	[AIC]
- 2	[AIC]	[AIC]	[AIC]	[AIC]
Child-Pugh Status				
- Class A	[AIC]	[AIC]	[AIC]	[AIC]
- Class B	[AIC]	[AIC]	[AIC]	[AIC]
Barcelona Clinic Liver Cancer				
- Intermediate (Stage B)	[AIC]	[AIC]	[AIC]	[AIC]
- Advanced (Stage C)	[AIC]	[AIC]	[AIC]	[AIC]

n.b. Hepatitis?

Validation overall survival extrapolation: GIDEON

Company

- Suggests longer survival benefit in ‘clinical practice’ than in SHARP (median overall survival: 324 vs. [AIC] days)
- log-normal distribution fits better statistically than Weibull

Evidence Review Group (ERG)

- Statistical goodness of fit should not be used in isolation to inform choice of survival function
- Visual inspection of the fitted curves plotted against the Kaplan-Meier curve, the log normal function **overestimates** overall survival while Weibull **underestimate** overall survival (see later slide)
  - Should consider both curves

Validation overall survival extrapolation: Overall survival Kaplan–Meier graph: matched (3:1) GIDEON dataset to SHARP patients

Source: Figure 3 of company's response to the ACD

Choice of survival extrapolation: GIDEON

Source:
page 9,
figure 1,
ERG critique
of company
response to
ACD

n.b. no
numbers at
risk

⊙ *What is the best statistical function for extrapolating overall survival?*

Validation of overall survival extrapolation: Palmer (1)

Committee (ACD conclusion)

- Not suitable to validate the extrapolation of overall survival beyond SHARP because the results were likely to be confounded

Company

- ACD incorrectly stated Palmer unpublished
- Object to ACD statements that “patients who did not receive funding did not live as long as patients who did have funding”
  - n.b. but this is what the Palmer study showed
- Decisions to fund “not apparently based on clinical variables”
- “No statistically significant difference in patient characteristics at baseline”
  - n.b. ‘funded’ (n=57) vs. ‘unfunded’ (n=76)
- Company provides Kaplan Meier curves without numbers at risk

Validation of overall survival extrapolation: Palmer (2)

ERG

- Patients not randomly allocated to funded and non-funded groups
- Potential imbalances in unknown confounders in non-randomised studies
- Selection bias could explain higher efficacy of sorafenib in Palmer compared with SHARP
- Palmer provides no robust evidence to favour one curve over the other

⊙ ***Did the committee see evidence to change its decision and to include Palmer data in decisions making?***

Treatment duration:
Time-to-progression as a proxy for treatment duration

Company

- Continue to prefer investigator-determined time-to-progression
- Maintain investigators determine actual time on treatment in SHARP
  - n.b. At odds with NEJM publication:
    - "Time to radiologic progression was defined as the time from randomization to disease progression on the basis of independent radiologic review"
    - "Treatment continued until the occurrence of both radiologic progression, as defined by RECIST, and symptomatic progression …"
- Claim investigator assessment in line with UK practice
- Uses a larger sample size and longer period of data collection
  - n.b. than what?

ERG

- Prefer individual patient data on time to treatment discontinuation (next slide) compared with using time-to-progression as a proxy or UK clinical practice data (see later slide):
  - consistency between treatment duration and efficacy estimates

SHARP data on time to treatment discontinuation for time to treatment discontinuation

Kaplan-Meier curve incomplete so company extrapolated using 5 parametric models – company preferred log normal on statistical fit						
Months	KM	Expo-nential	Weibull	Log logistic	Gom-pertz	Log-normal
Median	[AIC]	[AIC]	[AIC]	[AIC]	[AIC]	[AIC]
Mean	[AIC]	[AIC]	[AIC]	[AIC]	[AIC]	[AIC]

Treatment duration: Individual patient data from SHARP – 'hybrid'

Kaplan Meier
SHARP duration of treatment

Log normal

Duration of treatment using hybrid approach extrapolates from last observed event - *"smoothing effect of parametric curve reduces accuracy of observed events ([AIC]%)" Company does not provide results for exponential, Weibull, log logistic and Gompertz*		
months	KM	Log-normal
Median	[AIC]	[AIC]
Mean	[AIC]	[AIC]

Source: Table 8 and figure 5, pages 16 and 17 company's response to ACD. Key: NR = not reported.

Treatment duration data from SHARP ERG comments

- Favours using individual patient level data on time to treatment discontinuation because it avoids use of surrogates such as time to progression
- Company did not apply hybrid approach appropriately and its justification to use this approach was not appropriate
- Prefer fully parametric approach over hybrid approach
- AIC/BIC statistical fit criteria should not be used in isolation
- Differences in Weibull, Gompertz and log normal in AIC and BIC criteria are relatively small
- Weibull and Gompertz appear to provide more plausible extrapolation of treatment duration than log normal based upon visual inspection and consideration of external data

⊙ ***What is the committee's preferred approach for estimating the treatment duration?***

Treatment duration: UK clinical practice

Company

- Individual patient data from SHARP overestimate mean treatment duration in clinical practice

ERG

- Evidence provided by company is inconclusive
- CDF data confounded. Not all notifications result in treatment. Sorafenib being used in other indications
- Results of King et al. include only median treatment duration (3.2 months), not the mean, and the baseline characteristics of patients predict worse prognosis than those in SHARP
- Palmer only reported median treatment duration and this was close to that observed in SHARP (5.1 compared 5.3 months)
- Mean treatment duration reported for GIDEON ([AIC] months) is based on a population with a worse prognosis to that of SHARP
- NCRI commented that median time on treatment is 15.6 weeks

⊙ ***What is the appropriate data source for treatment duration?***

Treatment wastage

Company

- Wastage does not fall within the scope of the CDF re-consideration
- Hospital statements support that wastage is minimised in clinical practice, for example:
  - University Hospital Birmingham: 1 month at a time; very rarely 1 week. Splitting packs. Cannot eliminate waste, but ‘not considered to be a major issue within the Trust’.
  - Christie: 1 month at a time; some pack splitting, waste exists but is small.
- However, exploratory analyses include wastage of up to 7 days’ worth of treatment presented

Evidence Review Group

- Unlikely that treatment would incur absolutely no wastage
- Company’s approach to incorporating wastage is acceptable

⊙ ***Should wastage be included in the base case?***

Resource use estimates: Updated versus pooled

Company

- Pooling of the resource use not valid or appropriate as clinical practice has changed since initial resource data gathered in 2007
- Original resource use estimates based on no experience of using sorafenib in clinical practice compared with updated resource use based on actual experience through use in the CDF
- Prefer the updated estimates

Evidence Review Group

- Company did not provide any evidence to support its claims
- Further new estimates with a bigger sample size would have provided a more reliable estimate
- Biggest drivers of the ICER were mostly parameters in the best supportive care arm, and were estimated by only 2 clinicians
- Continue to prefer pooling resource use estimates

⊙ **Are the most recent resource use estimates or the pooled resource use estimates preferred?**

Company's ICERS revised price

Scenario	Details	Cost/QALY
Company base case	• Treatment costs from investigator-determined time to progression (TTP) • Log normal for extrapolating overall survival • Updated resource use data only • No wastage	£35,695
Appraisal committee's preferred assumptions for treatment costs	• Treatment duration based on SHARP time to treatment discontinuation with company's choice of 'hybrid' extrapolation • Independent assessment of progression • Pooled resource use • No wastage	£47,852
Scenario for treatment costs; including wasting	• Duration of treatment based on SHARP data (fully parametric curve, log normal) • Independent assessment of progression • Pooled resource use • 7 days wastage	£49,060
Company's base case plus Weibull for overall survival	• Treatment costs based on investigator-determined time to progression (TTP) • Weibull for extrapolating overall survival • Updated resource use data only • No wastage	£52,056

ERG's exploratory base case analysis: Based on the "ACD preferred assumptions"

Base case assumptions:

- Extrapolating overall survival on log normal distribution
- Time to progression based on independent reviewer assessment
- Treatment duration extrapolation based on patient level data for treatment duration from SHARP
  - fully parametric curve = log normal
- Resource use: pooled estimates from the original appraisal and the new submission
- Up to 7 days of wastage

ERG's exploratory analyses

Scenario (source: pages 17–18, table 2 of the ERG's critique of the company's response to the ACD)		ICER (£/QALY)
	ERG's base case (log normal, 7 days wastage)*	£49,299
	ERG's base case* (probabilistic)	£49,239
1	Extrapolation of overall survival: Weibull	£87,091
2	Extrapolation of treatment duration: Weibull	£41,935
3	Combining 1 and 2: Weibull overall survival and duration of treatment	£72,596
4	Wastage: half a pack (14 days)	£50,884

* Equivalent to the "ACD preferred assumptions" in the company's response to the ACD, but with an adverse costing error fixed.

Abbreviations: ACD = appraisal consultation document; ICER = incremental cost effectiveness ratio; QALY = quality adjusted life year.

Equality considerations

British Liver Trust

- “If NICE does not recommend sorafenib there will be variation in access and standard of care across UK with English patients being disadvantaged”

Web Comments

- “Deaths from liver cancer in England are more prevalent in people living in the most deprived areas and historically this patient group has been overlooked when funding decisions are being made – “little commitment” from the NHS to financially address this inequality”

Issues for discussion

- Did the company adequately address the outstanding issues of:
  - Extrapolation?
    - What is the most valid statistical function for extrapolating overall survival?
  - Treatment duration?
    - Source of data? Time-to-progression or time-to-treatment discontinuation?
    - Way to extrapolate?
  - Wastage?
- What is the committee's preferred resource use data?

Response to the Appraisal Consultation Document (ACD)
Sorafenib for treating advanced hepatocellular carcinoma
Bayer plc, 14th October 2016

Executive Summary

Bayer are disappointed that the preliminary recommendation from the National Institute for Health and Care Excellence (NICE) does not recommend sorafenib (Nexavar®) for the treatment of patients suffering from advanced hepatocellular carcinoma (HCC).

Advanced hepatocellular carcinoma is associated with reduced survival and at this stage of the disease treatment options for patients are very limited. Sorafenib is currently the only available treatment option for advanced HCC patients, and after 35 years of research is the only available systemic treatment shown to improve overall survival (1). In 2006 sorafenib received EMA orphan designation, highlighting its now established use in small populations with a high level of clinical unmet need.

Future treatments for advanced HCC patients are reliant on the availability of sorafenib as a first-line treatment option. The preliminary recommendations, if implemented without amendment, will leave no treatment option for the small number of patients with advanced HCC, for whom surgical or loco-regional therapies have failed or are not suitable.

Bayer welcome the opportunity to respond to the Committee's initial conclusions and believe the uncertainties raised in the ACD have been addressed. The ACD response addresses the following issues:

1. **Reduction in sorafenib treatment cost:** The current list price for a pack of sorafenib is £3,576.56[1] with a reduced CMU price of [redacted] per pack made available for the re-consideration submission. In response to the ACD a further price reduction has been offered, resulting in a new pack price of [redacted]. This represents a [redacted] discount to the list price and a further [redacted] discount to the price considered at the last committee meeting.

2. **Validating the extrapolation of overall survival:** The ACD acknowledged that the log-normal function used to extrapolate overall survival fitted the long-term GIDEON study data better than the Weibull function. It was the view of both the ERG and Appraisal Committee that this analysis was valid to the extent that the GIDEON population aligned with patients enrolled in SHARP(2;3).

   In line with guidance in the ACD, an analysis considering the overall survival of patients in the GIDEON study aligned to those enrolled in SHARP (matched on patient baseline characteristics via propensity scoring) was conducted to address the differences in baseline prognostic factors between the two studies. Results show that when prognostic factors are controlled for the log-normal function continues to provide a statistically superior fit based on AIC/BIC criteria for the extrapolation of overall survival when considered versus the Weibull function.

[1] 112-tablet pack, 28-day supply (200mg)

Findings from the analysis also show that matched GIDEON patients have markedly improved overall survival versus patients in SHARP (median OS: 324 vs [redacted] days), despite receiving a lower mean duration of treatment of [redacted] months and lower mean dose intensity of [redacted] (vs. 710.5mg in SHARP), adding to the clinical plausibility of the log-normal function. This analysis adds to the weight of evidence in support of the log-normal extrapolation and should be considered in all cost-effectiveness scenarios.

3. **Duration of treatment:** In line with the Appraisal Committee's request, patient level data from the SHARP trial has been used to model the unrestricted mean duration of treatment. This leads to a treatment duration estimate of [redacted], exceeding that observed in clinical practice. The company believes that both investigator and independent assessment of trial based discontinuation at progression lead to more realistic estimates than a trial based extrapolation. However when considered in conjunction with the committee's preferred assumptions, incorporation of this estimate in the model results in sorafenib remaining a cost-effective use of NHS resources.

4. **Resource use:** The Company does not consider the pooling of resource use collected pre-launch in 2007 with the updated resource use survey to be valid. Increasing the sample size does not increase accuracy or robustness if underlying clinical practice has changed. Whilst this change to clinical practice was acknowledged to be possible by the ERG for sorafenib treated patients (and was the basis of a 'hybrid' analysis in the ERG model, omitted from the report showing cost-effectiveness under their preferred assumptions) the company argues that introduction of sorafenib as standard of care and the 9 years since the original survey has implications for patients' resource use in the best-supportive care arm. This variation means that data collected via the updated resource use survey is more relevant and therefore robust than the pooled estimates.

5. **Treatment wastage:** Treatment wastage has not been considered an uncertainty in any of the previous Appraisal Committee meetings for sorafenib and as such does not fall within the scope of the CDF re-consideration process. Evidence has been presented from two of the largest HCC treating hospitals, explaining how wastage of sorafenib is minimised. These processes include pack-splitting and clinical assessment prior to and aligned with date of prescription. Analyses considering theoretical wastage have been conducted to address comments in the ACD.

For further information on the issues above please see Table 1 and the following detailed response.

Conclusion

Bayer hope that having addressed the committee's concerns, the additional evidence presented, as well as a further reduction to the cost of sorafenib, will allow the committee to reconsider its decision and approve sorafenib for the treatment of patients with advanced HCC. Base case cost-effectiveness provides a cost per QALY of £35,695 with this increasing to £49,060 when reflecting a 'worst-case' scenario. Clinical experience gained over the last 6-7 years supports the significant benefit sorafenib

provides for this small patient group, whom in the absence of sorafenib have no treatment option available.

Table 1: Summary of incremental cost-effectiveness ratios based on manufacturers/ Appraisal Committee's preferred assumptions.

Description of scenario	Scenario details	Cost per QALY
Manufacturer base case: As per original base case with inclusion of new treatment cost (section 5.1)	• Reduced treatment cost of sorafenib • Log-normal extrapolation of OS • Treatment duration based on investigator TTP • Updated resource use • No wastage.	**£35,695**
DSU's preferred base case (previously £51,208): The previously preferred ERG base case has been updated to include the reduced treatment cost of sorafenib (Section 5.2)	• Reduced treatment cost for sorafenib • Log-normal extrapolation of OS • Treatment duration based on independent TTP • Pooled resource use • No wastage.	**£46,863**
Appraisal Committee preferred assumptions: Appraisal Committee's preferred assumptions as presented in the ACD (Section 5.3)	• Reduced treatment cost for sorafenib • Log-normal extrapolation of OS • Duration of treatment based on SHARP patient level data (extrapolated from last observed point on KM) • Independent assessment of disease progression • Pooled resource use • Inclusion of wastage	**£47,852**
Appraisal Committee preferred assumptions: Appraisal Committee's preferred assumptions as presented in the ACD (Section 5.3)	• Reduced treatment cost for sorafenib • Log-normal extrapolation of OS • Duration of treatment based on SHARP patient level data (fully parametric curve) • Independent assessment of disease progression • Pooled resource use • Inclusion of wastage	**£49,060**

Bayer's detailed response to issues raised in the ACD

1 Reduced treatment cost for sorafenib (ACD section 4.24)

The Appraisal Committee stated that "sorafenib did not have plausible potential to be cost effective at the current Commercial Medicines Unit price"(2). In the interest of patient access a further discount to sorafenib has been offered in response to the Appraisal Consultation Document (ACD).

The new pack price of [redacted] represents an additional [redacted] discount to the price previously considered by the committee and a total discount of [redacted] to the list price. All analyses presented in this response use this new price.

2 Validating the extrapolation of overall survival (ACD section 4.19)

To date through the NICE submission process, an increasing weight of evidence has been considered that supports the use of the log-normal function for the extrapolation of overall survival. Table 2 considers evidence presented to date which follows steps to validating a survival model as outlined in NICE TSD 14 (4). It is hoped consideration of previous supporting evidence, in addition to analyses presented to directly address the committee's concerns will demonstrate that uncertainty has been addressed, with the weight of evidence heavily favouring the log-normal function.

Table 2: Evidence submitted in the consideration of most appropriate parametric survival model

Report	Evidence considered	Appraisal Committee comments
TA189:2009 (Initial submission) (5)	Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values supported log-normal as providing a statistically superior fit for the extrapolation of overall survival of both sorafenib and placebo arms of the SHARP trial.	*"the base-case lognormal extrapolation probably produced the most robust ICER for sorafenib"(6)*
TA189:2009 (Response to first ACD) (7)	Analysis of published long-term overall survival of patients with advanced HCC based on patient level data from the National Cancer Institute of New South Wales (NCI NSW) was presented and found: • **AIC/BIC statistical fit:** Log-normal provided a statistically superior fit (when compared with Weibull)	Not reported in FAD
TA189:2009 ERG report (8)	**ERG analysis of survival of advanced HCC patients:** The ERG analysed the OS Kaplan-Meier curves of two published HCC articles: **Camma et al 2008 (9):** "Both Weibull and log-normal distributions provide good fits to the data with the latter (log-normal) arguably superior"	Not reported in FAD

	Greten et al 2005 (10): "Very little difference between fits, log-normal being closer to observed data at early times but less satisfactory than Weibull at longest times".	
TA189:2009 (Response to second ACD) (11)	**NCI NSW:** Further survival analysis based on patient level data considering 3,280 patients found: • **AIC/BIC statistical fit:** Log-normal provided a statistically superior fit (when compared with Weibull) • **Cumulative hazard plots:** Log-normal displays less divergence from index lines (when compared with Weibull) **Greten et al 2005 (10):** As assessment of statistical fit was not presented by the ERG. AIC/BIC assessment presented by the company supported log-normal as providing a statistically superior fit (when compared with Weibull)	Not reported in FAD

2.1 CDF rapid re-consideration submission: February 2016

In the resubmission Bayer presented supporting evidence from two observational studies. The GIDEON study (12) (a large prospective observational study considering outcomes of 3,213 patients treated with sorafenib) and Palmer et al 2013 (13) (an independent comparative study conducted in the two largest specialist hepatobiliary oncology units in the United Kingdom).

2.2 Palmer et al. 2013 (section 4.19 of ACD)

In the ACD it was incorrectly stated that Palmer (2013) was an *"unpublished UK observational study".* Palmer et al (2013) was published in the British Journal of Cancer in 2013 (13).

The ACD goes on to state that *"The committee agreed that Palmer was not suitable to validate the extrapolation of overall survival beyond SHARP because the results were likely to be confounded".*

Whilst comments surrounding potential biases often attributed to observational studies may be valid, the decision to dismiss this evidence was based on the potential confounding effect that funding (used to apportion patients to the sorafenib and best-supportive care arms) was linked to patient prognosis, hence the Committee concluding that *"patients who did not receive funding did not live as long as patients who did have funding"(2)*

The study considers applications to local funding bodies made between July 2007 and May 2009. At this time Strategic Health Authorities (SHAs) were responsible for local funding decisions with regional variation at this time documented in a recent systematic review (14). The review identified variation specific to sorafenib, showing pre-CDF variation to be higher across 5 SHAs than post-CDF formation (sorafenib 0.08 to 2.5mg per head pre-CDF, 0.45 to 1.3mg post CDF)(15). The Palmer et al publication

outlines how potential biases were controlled for, and provides reasons to doubt that the provision of funding was a confounder:

- **Inclusion criteria were applied to all patients (funded and non-funded):** Criteria for application were uniform across each centre and comprised clinical information to indicate that, in the treating clinician's opinion, sorafenib was the most appropriate therapy. The variables are listed in the publication as, performance status (WHO PS 0-2); well compensated background chronic liver disease and lack of suitability for loco-regional therapies. All patients met these inclusion criteria which can be considered both aligned with the decision problem for this appraisal and reflective of patients treated in UK clinical practice.

- **Decisions to fund were *"not apparently based on clinical variables" (13)*:** Subgroup results from randomised trials indicate a similar benefit across all patient groups, with no known predictive variables that funding bodies could have used to select patients more likely to benefit from treatment, the authors concluded that decisions to fund was likely not based on clinical variables, other than the broad inclusion criteria detailed above which all patients met.

- **No statistically significant difference in patient characteristics at baseline:** Analysis was conducted on each reported baseline variable showing no statistically significant difference between groups. Whilst numerical differences are likely to exist due to the sample size, these were addressed via a series of sensitivity analyses once identified by the authors as potentially negatively influencing survival in the unfunded group. Controlling for these differences did not significantly affect the results, indicating differences in survival are likely due to treatment effect rather than due to imbalances in prognostic variables between the two groups. Results of these sensitivity analyses are presented in Table 3, showing when differences in baseline characteristics are accounted for HRs for overall survival exceed those reported in SHARP (HR: 0.69) (16).

Table 3: Palmer (2013) Overall survival results and sensitivity analyses conducted

Analysis	Result (overall survival)
All patients	Funded *vs* unfunded; 9.5 *vs* 4.1 months; **HR 0.48**, 95% CI 0.32-0.72; p=0.0005
Sensitivity analysis: Non-metastatic patients only	Funded *vs* unfunded; 8.95 *vs* 3.7 months; **HR 0.51**, 95% CI 0.32-0.82; p=0.0061
Sensitivity analysis: Exclusion of fibrolamellar patients	Funded *vs* unfunded; 8.98 *vs* 3.68 months; **HR 0.45**, 95% CI 0.29-0.69; p=0.0002

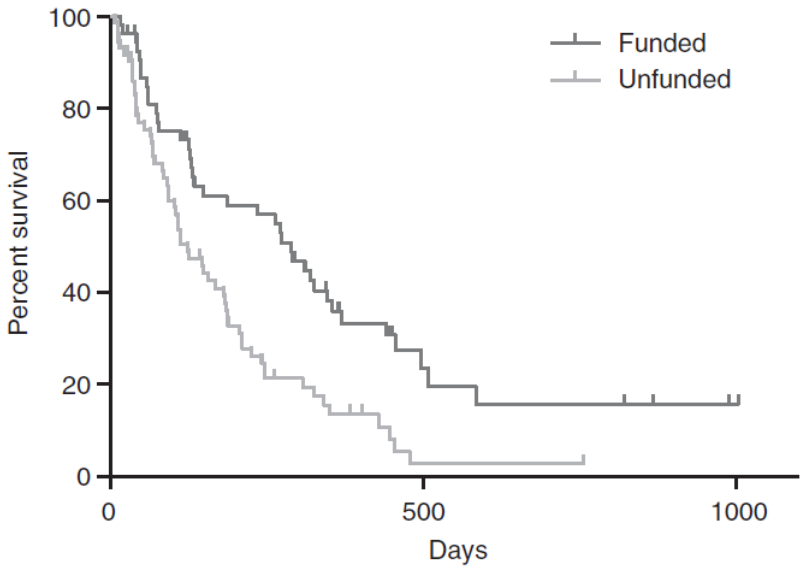
2.3 Palmer et al: Supporting evidence

Evidence from Palmer 2013 was submitted to support:

1. Overall survival over a longer duration than the SHARP study

Palmer et al (2013) presents a plateau in the overall survival Kaplan-Meier at approximately 600 days, with approximately 18% of patients surviving from this point up to 1,000 days (the final observation). This adds to the weight of evidence that the log-normal function, which is characterised by a small percentage of patients living for significantly longer than the average, is clinically plausible.

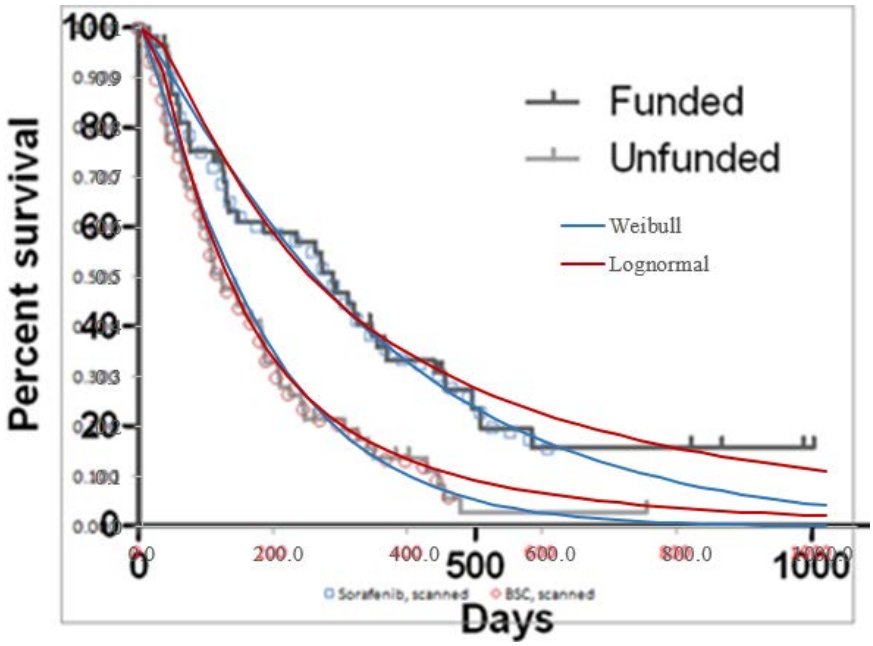
Figure 1: Palmer (2013) overall survival Kaplan Meier



When Weibull and log-normal parametric curves were applied to the model the ERG agreed with the company's methodology that the plateau of the KM was not to be included in the analysis as no events (in this case death) occurred and that inclusion of the plateau would reflect a mortality rate less than that of the general population (3). Whist this is logical, patients are censored at time points exceeding 600 days up to the point of last observation, indicating survival of patients over this period that could not be captured in the curve fitting analysis. This underestimates patient survival and therefore the appropriateness of the log-normal function as shown in Figure 2.

The Palmer study found that survival benefit versus placebo exceeded that observed in SHARP (HR: 0.48 vs 0.69) and may reflect experience of two high-volume liver units and the evolution of experience in using sorafenib (13). Results from Palmer et al indicate both support for the log-normal function and increased survival versus placebo in UK clinical practice.

Figure 2: ERG survival analysis (reproduced from Figure 4 in ERG report)



2. Scenario analysis: considering the cost-effectiveness outcomes using the Palmer 2013 data

In a scenario analysis conducted, outcomes from the Palmer study were used with assumptions to estimate the likely cost-effectiveness of treatment with sorafenib observed in UK clinical practice.

When the model is edited into a two state model (removing the utility benefit from post-progression survival) the analysis leads to an ICER of £18,870 and when one way sensitivity analysis is conducted, the ICER does not exceed £21,244. Additionally when dosing is taken from SHARP patient level data (a conservative assumption considering Palmer reports a lower median duration of treatment) and uses all other Committee preferred assumptions the ICER stays below £25,965 per QALY (see section for 5.4 full results).

Evidence from Palmer et al (2013) suggests a substantial degree of downwards uncertainty, highlighting that based on outcomes that have been observed in UK clinical practice on the CDF, sorafenib can be a cost-effective use of NHS resources.

2.4 GIDEON (Global Investigation of therapeutic decisions in hepatocellular carcinoma and of its treatment with sorafenib) (section 4.19 of ACD)

Survival analysis conducted on GIDEON (12), a long-term observational study (unpublished at the time of submission) considering the overall survival of 3,213 advanced HCC patients treated with sorafenib, was presented to examine the log-normal extrapolation of overall survival applied to the SHARP RCT (16).

With a longer follow-up than the SHARP study (39 vs. 19 months) and a larger sample of sorafenib treated patients (n=3,213 vs. n=299), this was considered a robust dataset in which to validate the

long-term extrapolation of overall survival applied to SHARP. Close alignment was observed in terms of patient survival times from extrapolated SHARP data using the lognormal distribution and actual patient survival in the GIDEON study (2).

The ERG acknowledged that the "log-normal distribution extrapolated from the SHARP trial data predicted long-term survival in GIDEON remarkably better than the Weibull does"(3). However there was concern that this alignment may be explained by "the heterogeneity of the population in the study" with the ERG concluding that "the better fit of the log-normal compared with the Weibull distribution to the KM curve of the GIDEON study, was relevant only to the extent that its studied population is considered representative of the target population"(3).

The ERG acknowledged that patients in the GIDEON study population were more severe at baseline than patients enrolled in the SHARP study. As uncertainty had previously surrounded the extrapolated section of the overall survival distribution, use of a more severe population in this analysis was likely to be conservative.

In the committee's view the *"log-normal function used by the company to extrapolate survival beyond the SHARP study fitted the GIDEON data better than the Weibull function, but because the populations between SHARP and GIDEON differed in the company's current analysis of GIDEON, the Weibull function still had some plausibility". As such the "committee stated that it would have been appropriate for the company to modify the GIDEON population to reflect characteristics of the population enrolled into SHARP when attempting to use GIDEON to validate SHARP" (2)* this analysis has now been conducted.

2.4.1 GIDEON: Matching of characteristics and prognosis factors to patients in SHARP

Rationale for analysis:

When comparing long-term overall survival (OS) from the non-randomised GIDEON study to the treatment arm of the SHARP RCT, selection bias is more likely to occur when the prognostic factors are unevenly distributed between both treatment groups (17). In accordance with the NICE TSD (17), propensity score matching based on individual patient level data from both GIDEON and SHARP was conducted to reduce potential bias due to cross-study differences in baseline patient characteristics. The analysis aimed to explore alignment in overall survival between matched SHARP and GIDEON patients and examine the extrapolation of the overall survival analyses previously used in the SHARP RCT. The full methodology for this analysis is presented in Appendix A.

2.5 Results

2.5.1 Statistical Fit

Assessing the statistical fit of each parametric model provides the most objective and robust assessment of goodness-of-fit and is a key assessment (in additional to clinical plausibility) as documented in TSD 14 (4).

The SHARP trial ended early, underestimating the survival benefit of sorafenib; however GIDEON's follow-up period is over double that of the SHARP trial, and incorporates over 10 times more patients treated with sorafenib. As the statistical assessment incorporates a longer period in the GIDEON trial in which to assess overall survival, results presented below show objectively that the log-normal provides a superior fit based on both AIC and BIC assessment.

Figure 3: Kaplan-Meier graph using 3 to 1 matched dataset

Table 4: AIC/BIC results for GIDEON matched analysis and SHARP RCT

Parametric	AIC			BIC		
	GIDEON Matched	SHARP RCT	SUM	GIDEON Matched	SHARP RCT	SUM
Weibull	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lognormal	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

2.5.2 Clinical plausibility

The matched GIDEON patients (n=895) show markedly increased overall survival when compared with the SHARP study (median OS: 324 vs. [redacted] days). This is a strong conclusion given the mean duration of treatment in GIDEON was [redacted] months and mean dose intensity was [redacted] (vs. 710.5mg in SHARP).

Table 5 compares days at which given percentages of patients remain alive in both the SHARP RCT and the patient matched GIDEON analysis, using both log-normal and Weibull functions.

Results show when patients from GIDEON are matched to those enrolled in SHARP, the log-normal extrapolation applied to SHARP may be conservative and in clinical practice survival benefit would exceed that considered in the economic model.

Table 5: GIDEON/SHARP: Days at which a given percentage OS is met in SHARP and GIDEON

Parametric Distribution	**SHARP RCT**			**GIDEON matched survival**		
Percentage survival	**50%**	**30%**	**20%**	**50%**	**30%**	**20%**
Weibull (days)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lognormal (days)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

2.5.3 Scenario analysis: 1 to 1 matching

To explore the potential impact on results of using a 3 to 1 matching approach, alternative analyses were also undertaken based on propensity score matching conducted using 1 to 1 matching between the SHARP and GIDEON populations. Results of this scenario analysis are closely aligned with those generated using the 3 to 1 matching, with AIC and BIC statistics also concluding the log-normal curve is the most appropriate. This suggests that the base case findings from the analysis using 3 to 1 matching are robust, and are not driven by the approach taken to patient matching. Full results from the scenario analysis are available in Appendix B:

2.6 Conclusion

The Appraisal Committee concluded that the *"Weibull retained plausibility due to the difference in populations between the SHARP and GIDEON studies"*, whilst the ERG attributed the superior fit of the log-normal to be relevant to the extent it matches the target population (the SHARP RCT). Both of these statements suggest the analysis previously presented, if adjusted to consider population differences, would address the uncertainty surrounding the choice of distribution to extrapolate overall survival.

The analysis requested by the Committee confirms that when study populations are aligned, the log-normal function provides a superior fit to extrapolate overall survival and this has been demonstrated based on both statistical fit and clinical plausibility. This evidence should be considered in

combination with that presented in Table 2, which shows to date the weight of evidence considered heavily favours extrapolation using the log-normal function.

The matched GIDEON population sample (n=895) is substantially larger than that of the SHARP RCT and shows markedly improved overall survival (median OS: 324 vs [redacted] days). With a shorter mean duration of treatment [redacted] months and a lower mean dose intensity was [redacted] (vs. 710.5mg in SHARP). These results suggest the model currently underestimates the benefit of overall survival as seen in clinical practice.

3 Duration of treatment (Section 4.20)

The SHARP study measured time to progression (TTP) using two different assessments. The investigator assessment (based on a local radiology assessment conducted by the treating physician) and the independent assessment (an assessment conducted subsequently by a central panel).

3.1 Investigator assessment of time to progression

The following outlines the rationale for employing the investigator assessment and reasoning as to why this is a valid assessment:

1. Investigator assessment determined actual time on treatment in the SHARP study

In the SHARP trial the decision to discontinue treatment with sorafenib was based on the investigator assessment of time to disease progression. Subsequent central independent assessment of progression had no bearing on treatment decisions made during the study. Patients treated post-progression were modelled separately in line with the committee's previous preferred assumptions, which notably caused confusion when aligning mean duration of treatment and TTP in the previous meeting.

Consideration of the independent assessment, which suggests patients have a longer period of progression-free survival when treated with sorafenib, leads to an overestimation of time on treatment accounting for cost of treatment in the model that study participants never received. This additional time on treatment is not reflected in the clinical outcomes. For consistency with clinical efficacy presented from the SHARP study, the investigator assessment which drove time on treatment and therefore clinical outcomes observed should be used.

2. Investigator assessment by the treating physician is in line with clinical practice in the UK

Progression is important in the respect that it determines the decision as to when treatment is discontinued. For regulatory purposes, such as drug licensing, an independent assessment provides a consistent and central evaluation, however in clinical practice it is the investigator (synonymous with physician) who conducts the assessment and bases patient need for treatment upon this assessment.

In clinical practice scans are not sent for review by an independent body and therefore an investigator assessment where progression and discontinuation of treatment is determined by the treating physician is in line with UK clinical practice as observed in the NHS.

3. The investigator assessment results in a larger sample size , a longer period of data collection and from a statistical standpoint represents a more robust dataset

The central, independent assessment was stopped at the first interim analysis as pre specified (data cut off May 2006). The local assessment, performed by the investigators, continued until second interim analysis (data cut off October 2006). Thus the numbers of events differs:

- Independent: [redacted] progressions in total ([redacted] sorafenib/ [redacted] placebo),
- Investigator: [redacted] progressions in total ([redacted] sorafenib/[redacted] placebo)

From a statistical point of view the investigator assessment, with a larger sample size and period of data collection represents the more robust dataset.

Results using the investigator assessment

The investigator assessment, leads to a cost per QALY of £35,695, after accounting for the reduced cost of sorafenib. This approach factors in post-progression treatment for [redacted] of patients as observed in the trial for an average duration of [redacted].

Description of scenario	Scenario details	Incremental cost per QALY
Manufacturer base case As per original base case with reduced treatment cost of sorafenib	• Reduced treatment cost of sorafenib • Log-normal extrapolation of OS • Investigator time to progression (TTP) assessment • Updated resource use • No wastage	**£35,695**

3.2 Independent assessment of time to progression

The ERG previously stated a preference for the independent assessment of progression on the basis that it was the assessment used in the SHARP publication and that there was a possibility of bias with the investigator approach (3). This 'bias' is more a reflection of the variability in investigator assessment, which is both reflective of clinical practice and treatment decisions made in the SHARP trial that determined the efficacy observed.

Recalculation of the assessment above using the independent assessment and new cost of sorafenib leads to an ICER of £41,123. When combined with the ERG's preferred assumptions which formed their most plausible ICER of £51,208, sorafenib is cost-effective with an ICER of £46,863.

Description of scenario	Scenario details	Incremental cost per QALY
Independent assessment of TTP: As per original base case with inclusion of new treatment cost	• Reduced treatment cost of sorafenib • Log-normal extrapolation of OS • Independent assessment of time to progression (TTP) assessment • Updated resource use • No wastage	**£41,123**
DSU's preferred base case (previously £51,208): After addressing the uncertainty surrounding log-normal as the most appropriate extrapolation of overall survival, the previously preferred ERG base case has been updated to include new treatment cost	• Reduced treatment cost for sorafenib • Log-normal extrapolation of OS • Independent assessment of TTP • Pooled resource use • No wastage	**£46,863**

3.3 Scenario analysis: Mean duration of treatment based on patient level data from SHARP

3.3.1 Rationale for conducting analysis

This analysis was conducted in line with the ACD conclusion that *"it was possible and appropriate for the company to estimate the duration of treatment with sorafenib based on the actual patient-level data from SHARP rather than using the proxy measure of progression free survival" (2)*

3.3.2 Methodology

An alternative approach to modelling treatment costs was developed whereby sorafenib treatment administration was not linked to progression status, but modelled using data on patient discontinuation based on events from the trial.

Patient level data from the pivotal SHARP study was used to perform a survival analysis to estimate time on treatment. The duration of treatment endpoint was defined as the time from the date of randomisation to date of discontinuation of treatment due to any cause. In total, there were [redacted] events/failures in the survival analysis with the last observed exit at [redacted]. The analysis captures all patients treated with sorafenib.

Since the Kaplan-Meier curve was not complete, parametric extrapolation methods were applied in line with other extrapolated parameters in the model. Five parametric models were explored: exponential, Gompertz, loglogistic, log-normal and Weibull.

A search of the literature also identified an alternative approach to modelling treatment costs based on patient level data. A hybrid approach whereby extrapolation of the KM is applied from the last observed discontinuation event, thus removing the parametric smoothing from the observed portion of KM was explored. For consistency this hybrid approach to extrapolation was applied to other extrapolated model parameters in a scenario analysis.

3.4 Results

Table 6 presents the AIC and BIC values based on extrapolation of the full parametric curve. Considering these criteria the log-normal provided the closest fit and was therefore used to extrapolate over the time horizon of the model for both extrapolation approaches employed.

Table 6: Duration of treatment: AIC/BIC statistical assessment

		Extrapolated models				
	KM	Exponential	Weibull	Loglogistic	Gompertz	Lognormal
AIC		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BIC		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

3.4.1 Full parametric approach to extrapolation

Application of a full parametric curve reflects how OS and TTP are modelled and is therefore presented. In this approach a parametric curve is fitted to the entire KM. The smoothing effect of the parametric curve reduces the accuracy of observed events ([redacted] of discontinuations) and as such is considered less robust than the hybrid method reported below.

Based on the AIC/BIC results presented above, the log-normal results in mean duration of treatment of [redacted] as presented in Table 7. The digitised survival function of the KM with the five parametric models applied is presented in Figure 4 below.

Table 7: Mean and median duration of treatment using full parametric approach to extrapolation

		Extrapolated models				
	KM	Exponential	Weibull	Loglogistic	Gompertz	Lognormal
Median DoT (months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean DoT (months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Figure 4: Digitised survival curves of patients on sorafenib treatment

3.4.2 Hybrid approach to extrapolation

The committee noted that the mean duration of treatment from the SHARP study was unrestricted and recommended that patient level data was used to extrapolate duration of treatment for all patients (2). In this approach the KM is used where available to accurately calculate treatment costs (covering [redacted] of discontinuations) and is then extrapolated from the last observed event (based on all the available data from the KM) as presented in Figure 5. This approach is robust as all observed events are modelled in line with the trial, this minimises the assumptions associated with parameterisation of the DoT curve and leads to a mean duration of treatment of [redacted] months using the log-normal.

When the hybrid approach to extrapolation is applied to the other extrapolated parameters in the model (overall survival and TTP) the ICER using the committee's preferred assumptions falls to [redacted] full details of this analysis are reported in section 5.3.2.

Table 8: Mean and median treatment duration based on hybrid extrapolation

	KM	Lognormal with hybrid approach
Median DoT (months)	[redacted]	[redacted]
Mean DoT (months)	[redacted]	[redacted]

Figure 5: Survival curves of patients on sorafenib treatment using hybrid method

Consideration of both possible approaches, in addition to all the Appraisal Committee's preferred assumptions leads to ICERs less than £50,000. Full results can be found in section 5.3.

Description of scenario	Scenario details	Incremental cost per QALY
Last observed event carried forward (with Appraisal Committee's preferred assumptions)	• Reduced treatment cost for sorafenib • Log-normal extrapolation of OS • Duration of treatment based on SHARP patient level data (extrapolated from last observed point on KM) • Independent assessment of disease progression • Pooled resource use • Inclusion of wastage (up to 7 days)	**£47,852**
Full parametric curve (with Appraisal Committee's preferred assumptions)	• Reduced treatment cost for sorafenib • Log-normal extrapolation of OS • Duration of treatment based on SHARP patient level data (full parametric curve) • Independent assessment of disease progression • Pooled resource use • Inclusion of wastage (up to 7 days)	**£49,060**

3.5 Trial based treatment estimate

There are reasons to believe that this trial based extrapolation resulting in a mean duration of [redacted] is not aligned with clinical practice:

Treatment post-progression: The trial protocol states that *"treatment beyond the point of disease progression is allowable as long as the investigator assesses the patient is continuing to derive benefit from study drug treatment".* In the phase III SHARP study [redacted] of patients continued treatment for an average of [redacted] months as observed at the end of the trial.

At the time of the SHARP study no other treatment had displayed a survival benefit for patients with advanced HCC and due to a lack of experience in treating post-progression patients it is likely that investigators would have had an ethical consideration to treat as long as there may have been a benefit and treatment could be tolerated. It is now understood that post-progression treatment is less efficacious with the Committee in agreement with the ERG previously stating that with sorafenib "benefit from post-progression treatment is likely to be small" (11). The extent and duration of post-progression treatment when extrapolated is likely to lead to an overestimate of the treatment duration when compared with current clinical practice.

3.6 Other empirical sources of treatment duration

SHARP is not the only source available in which to estimate average treatment duration and the cost that is likely to fall on the NHS.

3.6.1 Treatment duration from UK clinical practice

Data from the Cancer Drugs Fund: Between July 15 and June 16 there were [redacted] patient notifications for sorafenib. Over the same period [redacted] of sorafenib were sold in England accounting for [redacted] per patient. These figures are confounded by an indication in differentiated thyroid cancer which makes up [redacted] of sales. In the pivotal DECISION trial for the thyroid indication the average treatment on sorafenib was for 10.6 months (18), which is longer than that in SHARP resulting in a potential overestimation of treatment duration. Accounting for dosing intensity [redacted] packs is equivalent to [redacted] months at SHARP dosing intensity (710.5mg).This estimate is less than half of that observed in the duration of treatment obtained via patient level data from SHARP and reflects the likely cost to the NHS.

An independently conducted analysis from March 2011(15) found that if every application to the CDF led to a treatment dose and duration similar to that used in clinical trials or studies describing intended treatment, use of sorafenib should be almost four times higher than the actual observed use via the CDF.

Trial based treatment versus CDF sorafenib treatment: J King et al (19) conducted an independent retrospective audit of UK patients treated with sorafenib via the CDF and compared outcomes with patients treated for HCC in first-line systemic therapy trials. All UK centres that treat HCC were invited to participate and data was obtained from 279 sorafenib treated patients and 111 trial treated patients.

Median time on treatment for patients treated with sorafenib was 3.2 months (this compared to 4.2 months for patients treated with trial based systemic therapies). Findings here are very relevant to the UK and suggest treatment is significantly lower than the median of 5.3 months reported in SHARP, and the extrapolated mean of [redacted].

Palmer et al 2013 (13): Median treatment duration (mean not reported) was reported as 5.1 months, this is lower than the 5.3 months as observed from the SHARP study and reflects UK clinical practice.

3.6.2 Treatment duration from long term sorafenib studies

GIDEON: Estimates from a large population (n=3,213) over an extended period when compared with the SHARP study provides a robust estimate of treatment duration. In GIDEON the mean duration of treatment was [redacted] This is substantially lower than the estimate obtained via the extrapolation of SHARP RCT data of [redacted].

3.7 Conclusion

Assessing duration of treatment based on patient level data from the trial does not lead to an accurate reflection of the treatment costs that would be observed through use in the NHS. At the time of the study the benefit of treatment post-progression would not have been known and there would have been a lack of on-going clinical trials or alternative treatments for patients to commence. Due to the lack of alternative options or knowledge of post-progression treatment benefit, there may have been an ethical obligation to treat progressed patients as long as the patient could tolerate, as opposed treating in line with the clinical benefit observed.

Assessment via TTP leads to a treatment duration reflective of that seen in clinical practice, corroborated with the sources presented. Previous alignment issues between the mean duration of treatment and TTP at the previous meeting were due to post-progression being modelled separately and the company urges the committee to revisit these analyses.

Prescribing in the SHARP trial is not reflective of current practice; with the evidence sources presented (CDF, GIDEON, Palmer, King) highlighting the substantial downward uncertainty from both UK clinical practice and large long-term studies. Under each of the scenarios presented, including one reflecting the ERGs plausible ICER and another reflecting the Committee's preferred assumptions, sorafenib is a cost-effective use of NHS resources.

4 Costs and resource use estimates

4.1 Resource use (ACD section 4.22)

In the reconsideration submission up to date revised resource use estimates were presented, these were used to replace the original resource use questionnaire which was conducted in 2007 (prior to the launch of sorafenib). In the ACD the committee *"concluded that it was appropriate to pool the original and revised estimates of resource use due to the small number of clinicians" (2)*

Bayer would contest this. Increasing the sample size does not necessarily increase accuracy or robustness if underlying clinical practice has changed for either patients receiving sorafenib or best-supportive care.

4.1.1 Resource use for patients treated with sorafenib

In the original resource use survey, respondents would have had no experience in clinical practice of administering sorafenib or any other TKI to treat advanced HCC. Any experience would have been limited to participation in a clinical trial, or inferred otherwise. Many of the questions posed would have required a degree of experience prescribing sorafenib to identify both the regularity of an event and the related resource consequences. In contrast the updated resource is based on actual clinician experience through use in the CDF.

It is not plausible that resource use estimates for sorafenib use have not changed from pre-launch, with the ERG acknowledging the updated survey may provide better results for these questions.

4.1.2 Resource use for patients receiving BSC

It was claimed by the ERG that "estimates of the clinicians that took part in the new survey might have produced better results for the sorafenib arm due to the learning curve but the estimates for the BSC arm from the original survey should be equally as valid when compared with those of the new survey"(3).

In the past 10 years care of cancer patients has changed and it would be expected that out of the many questions asked to clinicians underlying resource would have changed to some extent . In 2007, when the original questionnaire was conducted there was no alternative treatment for advanced HCC patients, processes of care would have orientated around BSC and would look very different to those observed now.

4.1.3 Total sample size

The sample size is reflective of the fact that advanced HCC is a relatively rare condition, treated by a small number of specialists at centres with relatively high volume. All respondents were of consultant level and at the time of the updated survey all respondents would have had sufficient experience using sorafenib and BSC in today's clinical environment.

4.1.4 Hybrid resource use (as included in the DSU analysis)

As previously noted the ERG had accepted that resource use from the updated survey might have produced better results for the sorafenib arm due to the learning curve (3), but that estimates from the BSC arm should be equally as valid across the two surveys. Whilst the company refute the second point of the argument, it is noted that in the ERG model an analysis considering the pooling of questions in only the BSC arm was conducted (but not subsequently presented) reflecting this rationale[2]. This would have showed the ERGs previous most-plausible ICER of £51,208 to reduce to £49,666. Should the Committee conclude that pooling of surveys is appropriate it should be on the basis of the BSC arm only. Results reflecting the hybrid resource use a presented in section 5.5.3.

4.1.5 Conclusion

Given the acknowledgement by the ERG that sorafenib resource use could feasibly change, pooling of resource use estimates adds more uncertainty than it addresses. Bayer believe pooling undermines robustness of this data and do not think it is appropriate to combine results of the resource survey without evidence suggesting this is appropriate.

4.2 Treatment wastage (Section 4.21)

The committee stated *"It was also aware that in clinical practice, the company charges the NHS for a full pack of sorafenib at the start of each treatment cycle. Some patients do not complete their treatment cycle, and therefore the company may have underestimated the cost of treatment in its economic modelling. The committee concluded that it was appropriate for the company to use updated unit cost data and account for any drug wastage because this reflected the price relevant to the NHS"(2)*

Treatment wastage has not been considered an uncertainty in any of the previous Appraisal Committee meetings for sorafenib and as such does not fall within the scope of the CDF re-consideration process. This was a surprising inclusion to the ACD.

The manufacturer knows of no published evidence considering the extent of treatment wastage due to disease progression. As this submission is an evidence based process, this would be the minimum required to inform such an evaluation.

To inform potential approaches to incorporate wastage a review of previous oncology technology appraisals, through both the STA and CDF reconsideration process, found no precedent for the inclusion of wastage for oral oncology treatments in the final most plausible ICER.

The company understands that in clinical practice the vast majority of patients would have monthly consultations. At this consultation an assessment would be conducted to consider the appropriateness of treatment with sorafenib, and upon meeting treatment criteria sorafenib would be prescribed, usually for a month, but in some cases on a weekly basis. Upon return to the oncologist a similar assessment would be conducted. If a patient was viewed to have progressed or was deemed unsuitable

[2] Hybrid resource use is an option on the resource use drop down

for treatment this would lead to termination of treatment and fall in line with the date of prescription resulting in no waste.

The company approached two of the largest HCC treating trusts to understand their practice and the implications of treatment wastage. Both are large trusts with high volume use, and years of experience with sorafenib, full statements can be found in Appendix B.

4.2.1 Hospital statements

The Christie (Manchester)

Chief pharmacist: [redacted]

- It is Christie policy to issue **only one month of sorafenib** (in all indications including HCC) therapy at a time

- **Prescribing of sorafenib is aligned with a patient's monthly follow-up** where a clinical decision is made in regard to patient suitability for treatment for the following month. Only when this is satisfied is sorafenib prescribed

- It is advised that patient supply of sorafenib is actively managed where possible i.e. **through pack splitting where appropriate**. The clinician, pharmacist and patient work closely to reconcile what medicines were used within the month and where the patient has not used some tablets, only the remainder of another month's supply will be issued to reduce wastage (i.e. the pack will be split and only the outstanding amount issued)

- Whilst this process cannot eradicate wastage entirely, **wastage of sorafenib is generally small and is not believed to be a major issue for the Christie**.

- The number of patients on sorafenib therapy is relatively small (approximately 10 patients/month across all indications) - this does not differ significantly month by month therefore stock going out of date has not been an issue over the past 3-4 years of use in the centre.

University Hospitals Birmingham

Chief Pharmacist: [redacted]

- **Only one month of sorafenib therapy is prescribed at any given time**; treatment is initiated by a cancer specialist and patients are fully informed about appropriate use of their oral anticancer therapy (both verbally and written)

- In cases where patients are determined to be high risk, a decision may be taken to issue only **one week's supply** and make a reassessment after one week of therapy, this however rarely occurs with sorafenib due to long-term clinical practice with the drug.

- Prescribing of sorafenib will occur at patient's monthly follow-up appointment where a clinical decision is made regarding the patient's suitability for treatment in the following month. Based on a standard evaluation of the patient including blood tests, where appropriate a following month's supply of sorafenib will be prescribed.

- The clinician, pharmacist and patient work closely to reconcile what medicines were used in the month (patients are advised to bring their medicines pack and any unused tablets to the appointment). Where the patient has not used some tablets, only the remainder of another month's supply will be issued to reduce wastage. **The supply is actively managed by splitting packs where appropriate** to ensure only the outstanding amount is issued.

- Whilst this process cannot eliminate wastage entirely, wastage of sorafenib is generally uncommon and not considered to be a major issue within the Trust.

4.2.2 Scenario analysis: Inclusion of wastage

Consideration of both trust statements and the clinical assessment required for prescription suggests that progression is not likely to be evenly distributed throughout the month of treatment and in the majority of cases stopping treatment would be aligned with a consultation, resulting in zero wastage.

To meet the Appraisal Committee's request a scenario analysis was conducted that considers incorporation of post-progression wastage. To accurately incorporate wastage into the model a separate Markov trace was built whereby the cost of the drug is either applied daily (if no wastage is assumed) or every 7 days (for up to 7 days of wastage).

Results:

Description of scenario	Scenario details	Incremental cost per QALY
Appraisal Committee's preferred assumptions	• Reduced treatment cost for sorafenib • Log-normal extrapolation of OS • Duration of treatment based on SHARP patient level data (extrapolated from last observed point on KM) • Independent assessment of disease progression • Pooled resource use • Inclusion of wastage	**£49,060**

4.3 Conclusion

There is no evidence base in which to estimate the wastage of sorafenib due to disease progression. This has not previously been raised as an uncertainty by the Committee and is unusual in the consideration of oral oncology treatments. Key trusts when approached communicated that wastage was not an issue and shared steps used to minimise treatment wastage (i.e. pack splitting and clinical assessment prior to prescribing). For the majority of patients treatment discontinuation is aligned with consultation and is non-existent, otherwise estimates are confounded as provision via homecare where pack splitting is established would limit wastage further. Whilst the company believes that wastage should not be included in this Appraisal, after inclusion in the model sorafenib remains a cost-effective use of NHS resources.

5 Results for consideration

5.1 Manufacturer base case

5.1.1 Deterministic result

The following estimates show the base case that was presented in the original resubmission, with the new cost for sorafenib applied. The following model settings were applied:

- Full parametric extrapolation of OS and TTP using the lognormal model
- Investigator assessment of TTP for duration of treatment
- Updated resource use

Table 9 Cost-effectiveness estimates of the original base case (deterministic)

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£35,695

5.1.2 Probabilistic results

Probabilistic sensitivity analysis (PSA) of the model was performed by sampling the value of the same set of parameters as in the original 2016 submission. Table 10 below contains the cost-effectiveness estimates using 1,500 iterations of the model.

Table 10: Cost-effectiveness estimates of the original base case (probabilistic)

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£35,992

5.2 DSU's preferred assumptions

The table below list the estimates for the DSU's preferred assumptions as stated in their report (3). This included:

- Fully parametric extrapolation of OS and TTP using the lognormal model
- Independent review of TTP as a proxy for duration of treatment
- Pooled resource use

Table 11: Cost-effectiveness estimates using DSU preferred assumptions

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£46,863

5.3 Appraisal Committee's preferred assumptions

The following considers a scenario analysis where duration of treatment is modelled based on patient level data from the SHARP trial. These analyses reflect the entirety of the committee's preferred assumptions as presented in the ACD.

5.3.1 Hybrid method of extrapolation

The method was used to estimate the ICER using the Appraisal Committee's preferred assumptions:

- Fully parametric extrapolation of OS and TTP using the lognormal model
- Independent assessment of TTP for utilities
- Pooled resource use
- DoT as a measure of sorafenib duration of treatment using the hybrid method of extrapolation
- Up to 7 days of sorafenib wastage

Table 12: Cost-effectiveness estimates using the AC's preferred assumptions and hybrid method of extrapolation for DoT

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£47,852

5.3.2 Model results using hybrid method for OS and TTP estimates

To maintain consistency in the model, the hybrid method of extrapolation (presented above) was also implemented to the OS and TTP long-term estimates for both the treatment and BSC arms. The survival curves for both the OS and TTP can be shown on Figure 6.

Figure 6: Survival/Progression curves using the hybrid method of extrapolation

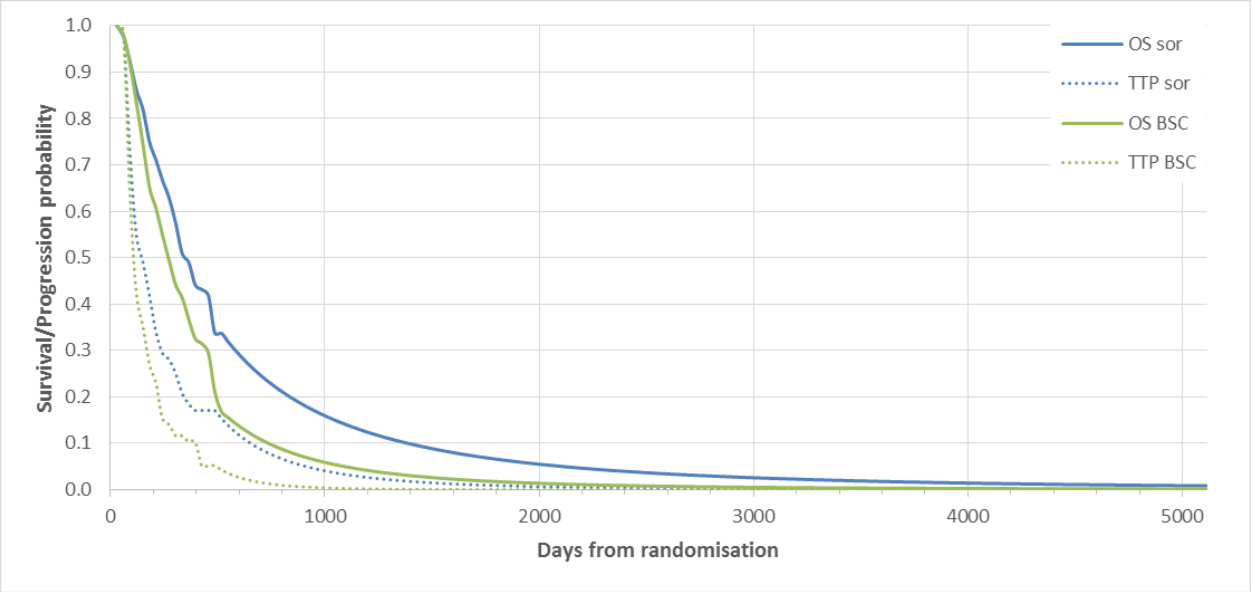


Table 13 shows the cost-effectiveness estimates when this method is applied to the extrapolation of DoT, extrapolation of OS and extrapolation of TTP in addition to the preferred assumptions of the AC. The following assumptions are applied:

- Hybrid method of extrapolation of OS and TTP using the lognormal model
- Independent assessment of TTP for utilities
- Pooled resource use
- DoT as a measure of sorafenib duration of treatment using the hybrid method of extrapolation
- Up to 7 days of wastage

Table 13: Cost-effectiveness estimates using the hybrid method of extrapolation and the AC preferred assumptions

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£43,740

5.3.3 Full parametric method of extrapolation

A scenario was explored in which the DoT estimates were fully parameterised using the lognormal model. The cost-effectiveness estimates of this analysis are presented in Table 14 below using the following assumptions:

- Fully parametric extrapolation of OS and TTP using the lognormal model
- Independent assessment of TTP for utilities
- Pooled resource use
- DoT as a measure of sorafenib duration of treatment, fully parameterised using a lognormal model
- Up to 7 days of sorafenib wastage

Table 14: Cost-effectiveness estimates using the appraisal committee preferred assumptions and fully parametrised model of extrapolation of DoT with up to 7 days wastage

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£49,060

5.4 Scenario analysis: Palmer et al (13)

5.4.1 Equalising utilities across pre and post-progression health states

Updated costs and resource use estimates were used for this analysis, with utility values for the pre-progression state and progression state equalised.

Table 15: Cost-effectiveness estimates using Palmer et al and equalised utility weights

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£18,870

Sensitivity analyses

One-way sensitivity analysis (OWSA) was conducted to determine if results were sensitive to variations in parameter values. ICERs ranged from £16,768 to £21,244 per QALY gained, with the most impactful parameters being the mean and standard deviations for TTP survival estimates in the sorafenib arm and the cost of BSC.

5.4.2 Palmer et al using Appraisal Committee's preferred assumptions

This scenario used the data from Palmer et al as reported above, however using the Committee's preferred assumption i.e. pooled resource use, DoT as a measure of time on treatment, and up to 7 days wastage. Note that a hybrid extrapolation method was used for the long-term DoT calculation and investigator review for TTP utilities.

Table 16: Cost-effectiveness estimates using Palmer et al using the AC preferred assumptions

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£25,965

5.5 Alternative resource use assumptions

5.5.1 Pooled resource use using manufacturer based case assumptions

The original base case assumptions (as detailed in section 5.1) were tested with pooled and hybrid resource use settings, an analysis implemented by the ERG. Table 17 and Table 18 show the results for pooled and hybrid resource use assumptions, respectively.

Table 17: Cost-effectiveness estimates using manufacturer based case assumptions and pooled resource use

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£41,905

Table 18: Cost-effectiveness estimates using manufacturer based case assumptions and hybrid resource use

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£40,596

5.5.2 Appraisal Committee's preferred assumptions using alternative resource use and hybrid method of extrapolation

This analysis considers the AC's previously preferred assumptions (which employ the pooled resource use survey) as considered in section 5.3, using both the updated and hybrid resource use questionnaire.

Table 19: Cost-effectiveness estimates using the ACs preferred assumptions and hybrid method of extrapolation of DoT, updated resource use with up to 7 days wastage.

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£42,406

Table 20: Cost-effectiveness estimates using the appraisal committee preferred assumptions and hybrid method of extrapolation of DoT, hybrid resource use with up to 7 days wastage.

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£46,685

5.5.3 Appraisal Committee's preferred assumptions using updated resource use and fully parametric model for extrapolation

This analysis considers the Appraisal Committee's previously preferred assumptions (which employ the pooled resource use survey) as considered in section 5.3, using both the updated and hybrid resource use questionnaire.

Table 21: Cost-effectiveness estimates using the AC's preferred assumptions and fully parametric model for extrapolation of DoT, updated resource use with up to 7 days wastage

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£43,614

Table 22: Cost-effectiveness estimates using the AC's preferred assumptions and fully parametric model for extrapolation of DoT, hybrid resource use with up to 7 days wastage

	Sorafenib	BSC	Incremental
Cost	[redacted]	[redacted]	[redacted]
PFLYs	[redacted]	[redacted]	[redacted]
LYG	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
Incremental cost per QALY			£47,893

References

(1) Verslype C, Rosmorduc O, Rougier P, ESMO Guidelines Working Group. Hepatocellular carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology 2012;23(suppl 7):vii41-vii48.

(2) National Institute for Health and Care Excellence. Appraisal consultation document: Sorafenib for treating advanced hepatocellular carcinoma. 2016.

(3) Decision Support Unit. Cancer Drugs Fund Rapid Review of NICE guidance TA189: Sorafenib for the treatment of advanced hepatocellular carcinoma. 2016.

(4) Latimer N. NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. Sheffield: Report by the Decision Support Unit 2011.

(5) Bayer Healthcare. TA189 Sorafenib for the treatment of HCC: Manufacturer's submission. 2009.

(6) National Institute for Health and Care Excellence. Hepatocellular carcinoma (advanced) - sorafenib (first line): 1st appraisal consultation document. 2010.

(7) Bayer Healthcare. Bayer Healthcare response to ACD- Sorafenib for advanced hepatocellular carcinoma. 2009.

(8) West Midlands Health Technology Assessment Collaboration. Sorafenib for advanced hepatocellular carcinoma: Assessment of a manufacturer's proposal for a Patient Access Scheme entitled: "one pack free for responders every 4 months". 2009.

(9) Camma et al. Survival of patients with hepatocellular carcinoma in cirrhosis: a comparison of BCLC, CLIP and GRETCH staging systems. Alimentary pharmacology & therapeutics 2008;28(1):62-75.

(10) Greten TF, Papendorf F, Bleck JS, Kirchhoff T, Wohlberedt T, Kubicka S, et al. Survival rate in patients with hepatocellular carcinoma: a retrospective analysis of 389 patients. British journal of cancer 2005;92(10):1862-8.

(11) National Institute for Health and Care Excellence. Hepatocellular carcinoma (advanced) - sorafenib (first line): 2nd appraisal consultation document. 2010.

(12) Marrero JA, Kudo M, Venook AP, Ye SL, Bronowicki JP, Chen XP, et al. Observational registry of sorafenib use in clinical practice across Child-Pugh subgroups: The GIDEON study. Journal of Hepatology 2016.

(13) Palmer DH, Hussain SA, Smith AJ, Hargreaves S, Ma YT, Hull D, et al. Sorafenib for advanced hepatocellular carcinoma (HCC): impact of rationing in the United Kingdom. British journal of cancer 2013;109(4):888-90.

(14) Chamberlain C, Owen-Smith A, Donovan J, Hollingworth W. A systematic review of geographical variation in access to chemotherapy. BMC cancer 2015;16(1):1.

(15) Stephens P, Thomson D. The Cancer Drug Fund 1 year onΓÇösuccess or failure? The Lancet Oncology 2012;13(8):754-7.

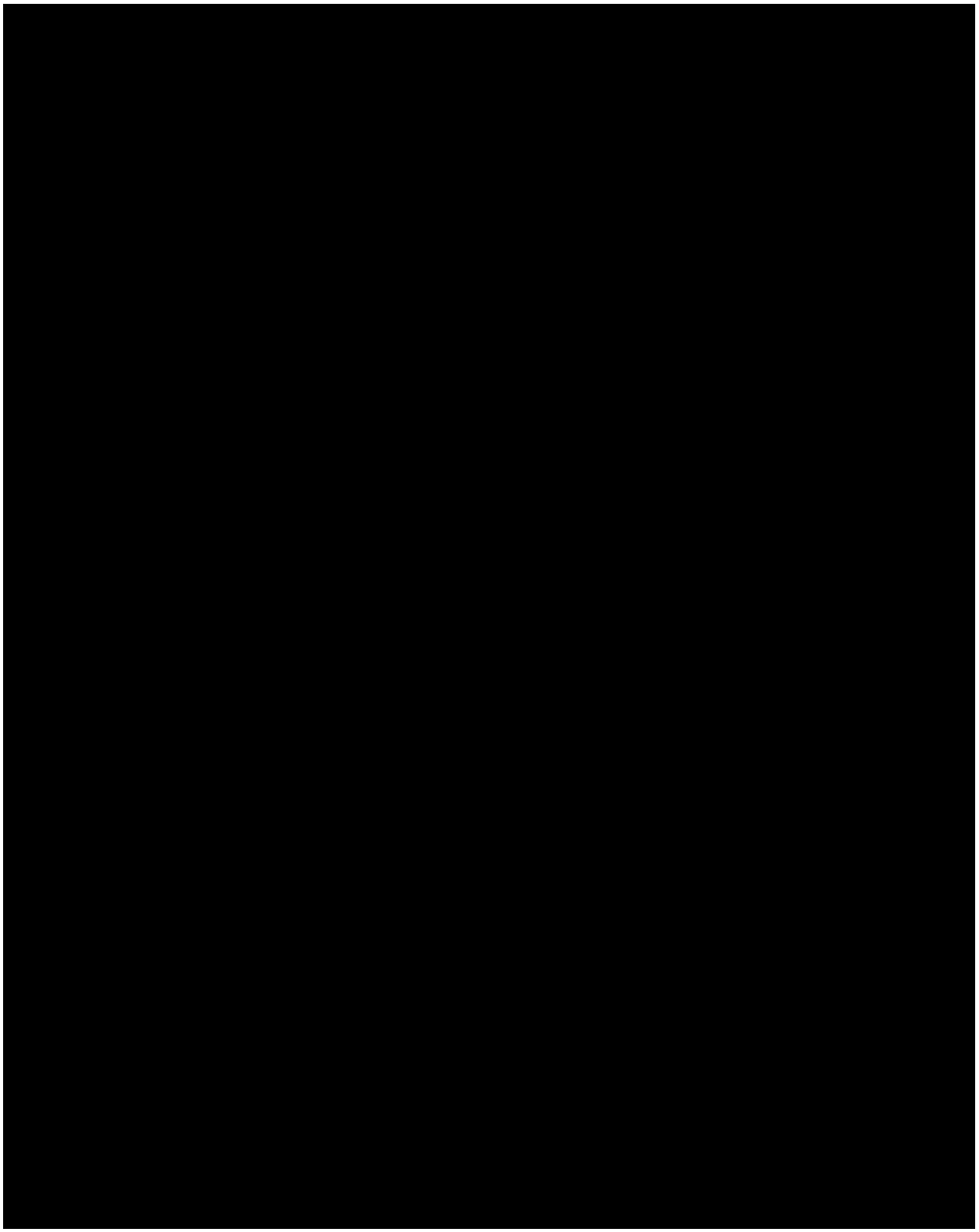
(16) Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JFdr, et al. Sorafenib in advanced hepatocellular carcinoma. New England journal of medicine 2008;359(4):378-90.

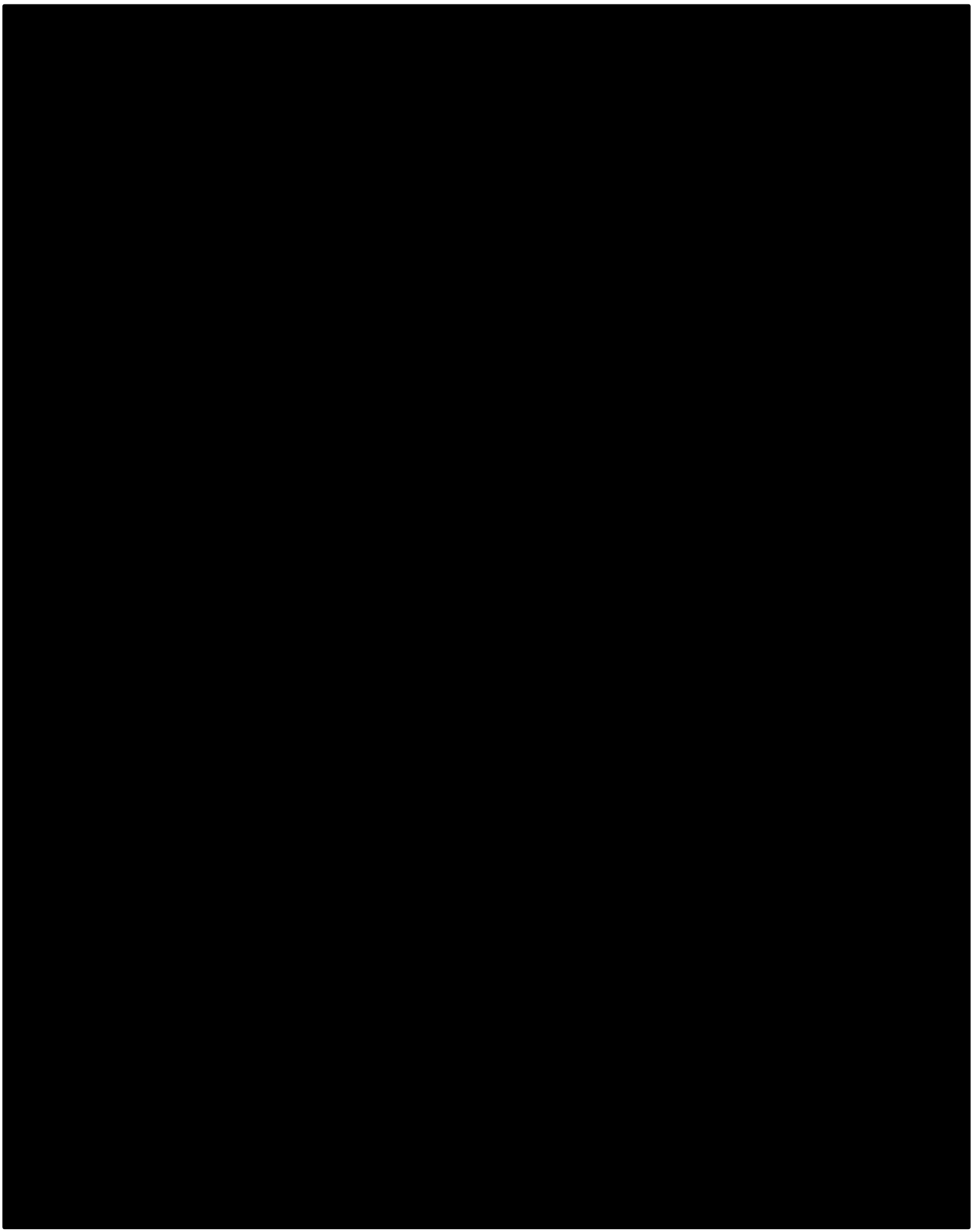
(17) Faria R, Alava MH, Manca A, Wailoo AJ. NICE DSU TECHNICAL SUPPORT DOCUMENT 17: THE USE OF OBSERVATIONAL DATA TO INFORM ESTIMATES OF. 2015.

(18) Brose MS, Nutting C, Jarzab B, Elisei R, Siena S, Bastholt L, et al. Sorafenib in locally advanced or metastatic patients with radioactive iodine-refractory differentiated thyroid cancer: The phase III DECISION trial. 2013 p. 4.

(19) King et al. Outcomes for UK patients with advanced hepatocellular carcinoma. 2013.

Appendix A: Pharmacist statements: treatment wastage





Appendix B: Methodology: SHARP/GIDEON matched analysis

Methodology:

A propensity score (PS) approach was used to match Individual Patient Data (IPD) from the GIDEON prospective non-intervention study to IPD from the treatment arm of the SHARP RCT. The benefit of the study approach is that by matching on clinically relevant baseline characteristics, observed cross-study differences are reduced.

Following Austin (2011), some of the clear benefits of propensity scoring matching are:

1. **Bias reduction:** An adequately specified PS model assumes that the distribution of measured baseline covariates is independent of the study arm the patient belongs to. PS therefore specifically requires examination of the degree of overlap in the distribution of observed/measured baseline covariates between the study arms (Austin, 2011).

2. **Integration in the study design:** PS allows the separation of the selection of the study sample from the analysis (Austin, 2011). Similar to the comparison of treatment groups from a randomized clinical trial, the balance between treatment groups across all potential confounders can be inspected before and after the matching. As such, the balancing of the covariates occurs without reference to the final study outcome (Patorno et al. 2013)

The intention of the PS in the current study is broadly similar to a matching-adjusted indirect comparisons approach (MAIC, Signorovitch et al., 2012) where individual patient data (IPD) from one trial arm or cohort are matched against the baseline summary statistics of the other trial arm or cohorts. The main difference, however, relies on access to IPD from both trials/cohorts which increases accuracy and the precision of the matching process

Whereas Austin 2011 described the PS approach as a general purpose tool, cautiousness about the interpretation of the suggested study is recommended as it cannot exclude:

1. Bias introduced by unmeasured confounding, introduced through unmeasured covariates or resulting from the selection criteria of the phase 3 trial;

2. Bias introduced through non-randomisation: in line with the DSU guidance comments IPD from different sources can be used if the individuals can be assumed to be drawn from the same patient population and be exposed to similar confounders (Faria et al., 2015). Although patients' selection in the proposed approach comes from different patient populations, the PS approach is still highly recommended to make the association between the populations more credible. In addition, the novel proposed approach has the advantage that the comparison group is randomised. As such, it can be assumed that an adequate PS approach reduces bias due to non-randomisation as the distributional characteristics of the study population will be similar to those of the randomized SHARP treatment arm.

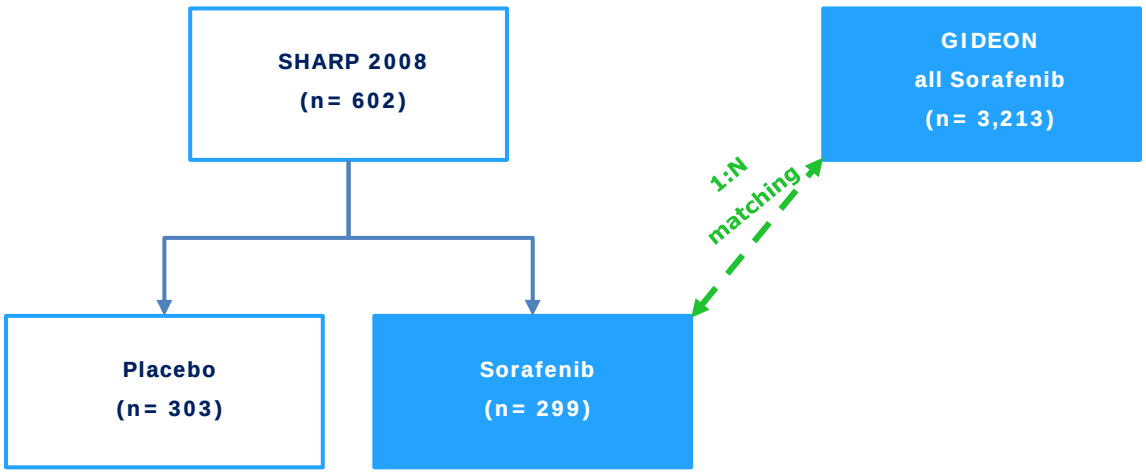
Study population

As shown in the below Figure 7, the main study population will consist of all patients:

- From the treatment arm of the SHARP 2008 study (n=299);
- From the GIDEON study (n= 3,213).

Patients from the placebo arm of the SHARP 2008 study will be excluded from the main study population.

Figure 7: Illustration of the proposed study design



Definition of the matching criteria / Covariates of interest

- Propensity scores were estimated using a multivariable logistic regression model with all available variables thought to be related to the outcome (Brookhart et. al., 2006) or related to the treatment decision for sorafenib initially eligible to be included. However, given the sample size of sorafenib-treated patients (n=299), a consideration was the need to limit the number of covariates for mathematical/ technical considerations as the resulting estimates/model may become unstable with too many covariates (Harrell, 2015). Therefore, univariate analyses on the treatment decision can be conducted first and covariates showing the strongest association will be entered into the multivariate logistic regression model. As selection of covariates occurs when specifying the model for the propensity score, the process is blind to outcome status i.e. study outcomes has no influence on the process of matching study covariates. All categorical covariates were represented in the logistic regression model using a set of binary indicator variables (Austin, 2011).

- Matching criteria of interest were in line with those listed by the ERG and presented in the previous NICE Appraisal Committee meeting:

1. Patients demographics:
    - Age (in years)
    - Gender (male)
2. Outcome related covariates measured at baseline:
    - Child-Pugh status (Class A, Class B)
    - BCLC (stage B, Stage C)
    - ECOG Status (Status 0/1, Status 2)
    - Vascular invasion presence
    - Extrahepatic spread presence
    - Hepatitis B presence
    - Hepatitis C presence

Statistical matching approach

The purpose of the analysis was to derive a comparative study cohort, matched according to key covariates, on which survival analyses would be conducted and alignment between the overall survival explored.

For propensity score matching, a one-to-many matching was applied in which, for every patient from the treatment arm of the SHARP trial (n=299), multiple patients from the GIDEON observational study (n=3,213) were selected (see Figure 7). In order to avoid imbalanced populations, the number of matches were restricted to 3 treated patients for every treated patient from the SHARP treatment arm. When compared to one-to-one matching, it is assumed that the variance of the estimator decreases by using more matches in the control group (Faria et al, 2015 ; Austin, 2011)

Propensity scores were generated using a multivariable logistic regression model. A greedy matching algorithm with nearest neighbour was applied in which a patient from the GIDEON cohort was matched onto a patient from the SHARP treatment group based on the smallest distance between the propensity scores of the patients. The selection process was done without replacement, which implicates that final estimates depended on the order in which the observations were matched. In order to overcome this issue, patients from both the GIDEON and SHARP cohorts were randomly ordered before matching. After the first patient had been matched, the patients with the second and third nearest propensity scores from the GIDEON treatment arm were matched onto the same SHARP patient. Selection without replacement was chosen over with replacement as matching with replacement likely results more in bad matches when the control group is small (Faria et al, 2015).

In order to further ensure high quality matching, caliper adjustment was implemented (Austin, 2011; Harris & Horst, 2016 ; Stuart, 2010). Caliper adjustment allows matching only when propensity scores fall within a designated distance (or caliper). For the initial analysis, a caliper of width of 0.15 of the standard deviation of the logit of the propensity score was taken, which is slightly lower compared to the 0.2 threshold as proposed by Austen (2011). Whilst a lower caliper ensures further bias reduction as a closer match is chosen (Faria et al., 2015), the proposed caliper value still is close enough to the Austin threshold to minimise the mean squared error of the estimated treatment effect (Austin, 2011).

After matching, the covariate balance between GIDEON and SHARP patients was assessed in order to ensure covariate balance between both cohorts (Harris & Horst, 2016). In line with more recent developments within the field, covariate balance was evaluated by comparing the effect size (Cohen's d):

$d = \frac{(\bar{x}_{treatment} - \bar{x}_{control})}{\sqrt{\frac{s^2 treatment + s^2 control}{2}}}$, for continuous variables; and

$d = \frac{(\hat{p}_{treatment} - \hat{p}_{control})}{\sqrt{\frac{\hat{p}_{treatment}(1-\hat{p}_{treatment}) + \hat{p}_{control}(1-\hat{p}_{control})}{2}}}$ for dichotomous variables (Austin, 2011).

where a Cohen's *d* equal to 0 indicated complete balance of covariates between the GIDEON and SHARP cohorts. Whilst there is a lack of agreement on the threshold, a standard difference less than 0.1 has been widely agreed upon to indicate a negligible difference in the mean or prevalence of a covariate between the GIDEON and SHARP cohorts (Austin, 2011). In addition, the variance ratio of treatment over control was calculated with a value of 1 indicating perfect balance.

An iterative approach was taken in case the propensity score model resulted in unbalanced covariates between the GIDEON and SHARP cohorts until an acceptable balance had been achieved (Austin, 2008).

Missing covariate values

In order to conduct a multivariate analysis to calculate the propensity scores, fully observed covariates are often assumed (Stuart, 2010). As missing data leads to incomplete matching, cases with missing data were excluded. In case a large proportion of a covariate was missing, the covariate was excluded from further analysis. As differential missing proportions can be expected between matched cases, there is the potential for the analysis population to be substantially reduced if clinical factors are adjusted for, which is a recognized limitation of this type of analysis.

Matching result

Baseline comparison of the treatment groups

Baseline characteristics are shown in Table 23 for the 299 Sorafenib treated patients from the SHARP study and 3,213 Sorafenib treated patients from the GIDEON study cohort. As indicated in the table, statistically significant differences between both treatment cohorts at baseline can be observed for Age, Gender (male), ECOG (0 or 1), Child-Pugh Status (Class A and B), Advanced BCLC (Stage C). Further evidence of imbalance between the GIDEON and SHARP cohorts is indicated by Cohen's *d* values for the same covariates of lower than -0.1 or larger than 0.1 (Table 23).

Vascular invasion presence, extrahepatic spread presence, hepatitis B presence, and hepatitis C presences showed large proportions of missing values for GIDEON patients (with 77.7%, 60.2%, 63.0%, and 67.5% missing values respectively) and were omitted from the analysis.

Table 23: Baseline characteristics of SHARP and GIDEON patients before propensity score matching

Covariate	Overall Sample (N= 3,512)	SHARP (N = 299)	GIDEON (N= 3,213)	*p value*	Cohen's *d*
Age					
Male					
ECOG					
- ECOG (0 or 1)					
- ECOG (2)					
Child-Pugh Status					
- Class A					
- Class B					
BCLC					
- Intermediate (Stage B)					
- Advanced (Stage C)					

Notes: [1] Continuous variables are represented as Mean ± standard deviation; Dichotomous variables as N (%); [2] Significance testing was carried out by t-test for continuous variables and Chi-square test for dichotomous variables

Propensity score estimation and matching

Due to the exclusion of vascular invasion presence, extrahepatic spread presence, hepatitis B presence, and hepatitis C presence, the logistic regression model was conducted with all other covariates related to the outcome:

1. Patients demographics:
    - Age (in years)
    - Gender (male)
2. Outcome related covariates measured at baseline:

- Child-Pugh status (Class A, Class B)
- BCLC (stage B, Stage C)
- ECOG Status (Status 0/1, Status 2)

After elimination of respondents with missing data, the analytical study set contained 3,311 observations (299 SHARP patients and 3,012 GIDEON patients).

Table 24: Number and proportion of missing data by covariate for SHARP and GIDEON patients before propensity score matching

	SHARP		GIDEON		OVERALL SAMPLE	
Total number of patients, N (%)	**299**	**(100%)**	**3,213**	**(100%)**	**3,512**	**(100%)**
Missing data by covariate, N (% of Total)						
Age	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Male	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ECOG						
- ECOG (0 or 1)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
- ECOG (2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Child-Pugh Status						
- Class A	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
- Class B	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BCLC						
- Intermediate (Stage B)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
- Advanced (Stage C)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total number of patients with missing data, N (% of Total)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Resulting sample size before matching, N (% of Total)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

In a next step, patients from the SHARP and GIDEON study cohorts were matched on propensity scores. For each patient of the SHARP cohort, the attempt was to match three patients from the GIDEON cohort

imposing a caliper of width equal to 0.15 of the standard deviation of the estimated propensity scores. In total, 98% (n= 296) of the SHARP patients were successfully matched to three patients from the GIDEON cohort. Each of the two remaining patients were matched to two distinct patients from the GIDEON study cohort.

Balance diagnostics

The baseline characteristics for matched SHARP and GIDEON patients are presented in Table 25. The absolute values for the standardised differences of the mean (Cohen's *d*) ranged from 0.001 for the intermediate BCLC stage (Stage B) to 0.025 for Child-Pugh Status Class A, indicating negligible difference in the mean or prevalence across all covariate between the GIDEON and SHARP cohorts.

Table 25: Baseline characteristics of SHARP and GIDEON patients in the final propensity scoring matched sample

Variable	SHARP_n (N = 299)	GIDEON_n (N= 895)	Cohen's *d*	Variance ratio
Age	[redacted]	[redacted]	[redacted]	[redacted]
Male	[redacted]	[redacted]	[redacted]	[redacted]
ECOG				
- ecog (0 or 1)	[redacted]	[redacted]	[redacted]	[redacted]
- ecog (2)	[redacted]	[redacted]	[redacted]	[redacted]
Child-Pugh Status				
- Class A	[redacted]	[redacted]	[redacted]	[redacted]
- Class B	[redacted]	[redacted]	[redacted]	[redacted]
BCLC				
- Intermediate (Stage B)	[redacted]	[redacted]	[redacted]	[redacted]
- Advanced (Stage C)	[redacted]	[redacted]	[redacted]	[redacted]

Note. Continuous variables are represented as Mean ± standard deviation; Dichotomous variables as N (%)

When compared to the unmatched cohort, a substantial improvement in Cohen's d can be noticed (Figure 8). Variance ratios for the matched model ranged from 0.903 to 1.036.

Figure 8: Comparison of Cohen's d for the matched against the unmatched sample

In addition, Figure 9 below further reports the mean for each of the covariates against the estimated propensity score, separated by treatment cohort (SHARP vs GIDEON). The nearly identical LOESS (representing the means) for the treatment groups indicates that the distribution of each of the covariates was very similar between the treatment groups.

Figure 9 Mean for each of the covariates against the estimated propensity scores for the SHARP and GIDEON treatment groups

With all absolute values of Cohen's *d* within the range of 0 and 0.1 and the variance ratios close to 1, it can be concluded that a good covariance balance is present between the matched patients from the SHARP and GIDEON cohorts.

Survival analysis of matched patients

Following completion of patient matching, a survival analysis was carried out using OS data and was run in STATA 13. The dataset generated was used to construct Kaplan-Meier curves for the 2 different groups of patients (SHARP sorafenib RCT patients and GIDEON sorafenib patients). Lognormal and Weibull parametric models were assessed using the AIC and BIC statistics produced from the different parametric models. We also present an approximate survival estimate at different time points.

Sensitivity analysis

To explore the potential impact on results of using a 3 to 1 matching approach, revised analyses were also undertaken based on propensity score matching conducted using 1 to 1 matching between the SHARP and GIDEON populations. Results from this analysis are presented below.

Figure 10: Kaplan-Meier graph using 1 to 1 matched dataset

Table 26: AIC and BIC results using 1 to 1 matched dataset

Parametric	AIC			Parametric	BIC		
	GIDEON Matched	*SHARP RCT*	*SUM*		*GIDEON Matched*	*SHARP RCT*	*SUM*
Weibull	[redacted]	[redacted]	[redacted]	Weibull	[redacted]	[redacted]	[redacted]
Lognormal	[redacted]	[redacted]	[redacted]	Lognormal	[redacted]	[redacted]	[redacted]

Table 27: Percentage survivors at different time points (days) based on parametric extrapolation using 1 to 1 matched dataset

Parametric Distribution	**SHARP RCT Survival**			**GIDEON matched survival**		
	50%	**30%**	**20%**	**50%**	**30%**	**20%**
Weibull	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lognormal	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Results in this scenario analysis are relatively closely aligned with those generated using the 3 to 1 matching, with AIC and BIC statistics also suggesting the lognormal curve is the most appropriate. This suggests that the base case findings from this analysis using 3 to 1 matching are robust, and are not driven by the approach taken to patient matching.

SHARP/GIDEON analysis references

Austin, PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Statist. Med. 27 (2008): 2037–2049.

Austin, PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research 46(3) (2011): 399–424.

Brookhart, AM., Schneeweiss, Kenneth J, Robert J. Glynn, Robert J, et al. Variable Selection for Propensity Score Models. Am. J. Epidemiol. 163(12) (2006): 1149-1156.

Faria, R., Hernandez Alava, M., Manca, A., Wailoo, A.J. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. 2015. Available from http://www.nicedsu.org.uk

Harris, H & Horst SJ. A brief guide to decisions at each step of the propensity score matching process. Practical Assessment , Research & Evaluation 21(4) (2016).

Harrell, F. (2015). Regression Modeling Strategies with Applications to Linear Models, Logistic and Ordinal Regression and Survival Analysis (2nd Edition). Springer.

Patorno, E., Grotta, A., Bellocco, R., & Schneeweiss, S. (2013). Propensity score methodology for confounding control in health care utilization databases. Epidemiology Biostatistics and Public Health, 10(3).

Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci.2010;25(1):1–21.

Signorovitch, JE. et al. Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research. Value in Health 15(6) (2012): 940 - 947

Response to Appraisal Consultation Document; Sorafenib for treating advanced hepatocellular carcinoma (August 2016)

On behalf of the NCRI Hepatobiliary Subgroup

[redacted]

Members: [redacted]

Professor [redacted]

We are concerned that the Committee has recommended that sorafenib is not recommended for use within the Cancer Drug Fund or for routine commissioning within the NHS.

The committee agreed that sorafenib is a clinically effective therapy for patients with advanced hepatocellular carcinoma but concluded that it is not a cost-effective use of NHS resources. We would like the committee to consider the following points in in their subsequent review.

1. The assumptions in the cost model have been based on data from the SHARP trial and two observational studies; GIDEON and Palmer et al 2013. These provide very limited data on UK practice. We have undertaken and presented an extensive multi-site audit involving 15 centres and including 448 patients (King et al ESMO 2015). Important and relevant findings were:
    a. Median time on treatment in the UK is 15.6 weeks not 24. Therefore most patients are not treated until death and reasons for stopping were; 29% radiological progression, 25% toxicity, 19% clinical progression and 19% death.
    b. Actual daily dose is 590mg rather than 800mg per day.
    c. Patients with Child-Pugh A do much better than B (OS 9.5 vs 4.6m) and those with a performance status of 0 do better than those >0 (12.9 vs 8.0 months)

    These data suggest that the cost of treatment is less that that used in the models and that cost effectiveness can be improved by selecting patients with Child-Pugh A disease and good performance status.

2. Sorafenib is the only approved drug for advanced HCC and, while chemotherapy is used in selected patients, the evidence base is weaker with no placebo or BSC trials to define the survival benefit. Therefore, Sorafenib is defined as the current standard-of-care for advanced HCC in international guidelines agreed in Europe by EASL and EORTC and in USA by AASLD. Removing NHS access to sorafenib would have the following consequences.
    a. A global standard of care of care would cease to available to patients with advanced HCC in England and the use of less evidence based therapy such as chemotherapy may increase with associated increase in toxicity and treatment costs.
    b. The UK has established an excellent reputation for running trials in advanced HCC and now attracts both early and late phase studies including the most exciting new agents available. There is a large portfolio of second-line clinical trials for which sorafenib is required as standard first line therapy (See Appendix 1), and England would be excluded from these studies, further decreasing therapeutic options for patients and diminishing our reputation as an outstanding research environment.
    c. As a consequence of points a and b, the survival for patients with advanced HCC treated in England is likely to fall and compare unfavourably with outcomes in other countries of similar economic status.

3. Globally, HCC is the second leading cause of cancer death but in the UK, the incidence is low at around 3000 cases per year. It is therefore an uncommon cancer in UK and affects an often hard-to-reach population with broad ethnic diversity reflecting the aetiology of chronic liver disease. The UK Sorafenib audit conducted over a 6 year period, identified less than 450 patients in total treated across all the UK major centres. A typical liver centre of which there are 5-6 in the UK, initiates sorafenib in around 20 patients per year. Given that patients with advanced HCC will be left with no evidence based therapeutic options, the overall burden to the NHS in providing sorafenib to a relatively small number of patients should be considered.

Based on these considerations, we ask the committee to reconsider its recommendation and suggest that the provision of sorafenib should be maintained for patients with advanced HCC, Child-Pugh A liver disease and performance status of <2.

Yours Sincerely,

[redacted]

[redacted]

Appendix 1. Second line trials in the UK for which first-line sorafenib is standard.

1. A Study to Evaluate the Effectiveness, Safety and Tolerability of Nivolumab and the Combination Nivolumab Plus Ipilimumab in Patients With Advanced Liver Cancer (CheckMate040)

2. A Randomized, Double-Blind, Multi-Center Phase 3 Study of ADI-PEG 20 Plus Best Supportive Care (BSC) Versus Placebo Plus BSC in Subjects With Advanced Hepatocellular Carcinoma (HCC) Who Have Failed Prior Systemic Therapy

3. A Phase 3, Randomized, Double-blind, Controlled Study of Cabozantinib (XL184) vs Placebo in Subjects with Hepatocellular Carcinoma Who Have Received Prior Sorafenib

4. RESORCE: A Randomized, Double Blind, Placebo Controlled, Multicenter Phase III Study of Regorafenib in Patients With Hepatocellular Carcinoma (HCC) After Sorafenib

5. REACH-2: A Study of Ramucirumab (LY3009806) Versus Placebo in Participants With Hepatocellular Carcinoma and Elevated Baseline Alpha-Fetoprotein

6. Phase 1, open-label, first-in-human (FIH) study designed to evaluate the safety, tolerability, pharmacokinetics (PK), pharmacodynamics (PD), and preliminary antineoplastic activity of BLU- 554 administered orally in patients with hepatocellular carcinoma (HCC)
7. OUTREACH: First-in-Human Safety and Tolerability Study of MTL-CEBPA in Patients With Advanced Hepatocellular Cancer

8. Dose Escalation Trial of Tefinostat for Cancer Associated Inflammation in Hepatocellular Carcinoma (HCC) (CHR-2845)

NHS England submission into the NICE re-appraisal of sorafenib in the treatment of hepatocellular carcinoma

1. The SHARP trial randomised a mainly European population of patients to sorafenib plus best supportive care vs supportive care alone. It demonstrated clinically meaningful increases in independently assessed median time to treatment progression (5.5 vs 2.8 months, Δ 2.7 mo, HR 0.58, 95% CI 0.45-0.74, p=0.000007) and median overall survival (10.7 vs 7.9 mo, Δ 2.8 mo, HR 0.69, 95% CI 0.55-0.87, p=0.00058) at the expense of significant but tolerable toxicity. The trial was stopped after the interim analysis showed this benefit and hence longer term information on the trial patients is not known.

2. NHS England urges the NICE Technology Appraisal Committee to use the same patient source of information on which to base their preferred estimates of both treatment duration and overall survival. The SHARP trial data provides the evidence of the median duration of treatment with sorafenib in the same population of patients which provides the evidence of the median duration of survival of sorafenib (and thus the survival gain of treatment with sorafenib). Separating the source of information of treatment duration from the source that provides the survival data increases uncertainty (and greatly so, is NHS England's view).

3. The point about potential wastage of sorafenib is this: a 28 day supply of sorafenib is given to the patient at each visit. As long as the economic modelling uses whole numbers of 28 day supplies and not the actual number of months of treatment that reflect the day on which the treatment was stopped, then the issue of wastage has been addressed. For example, a patient stopping sorafenib 4.1 months (17.5 weeks) since the start of treatment has been given 5 packs of sorafenib; another patient stopping treatment 4.0 months (17.1 weeks) after starting treatment has still been dispensed 5 packs of 28 day supply of sorafenib. Trusts will not waste sorafenib as the Christie and Birmingham audits supplied by Bayer show. What matters is how the economic modelling has incorporated the everyday use of 28 day packs of sorafenib and that only a whole number of 28 day packs are dispensed.

4. If NICE does not recommend sorafenib in hepatocellular carcinoma for baseline commissioning, there are consequences beyond just the availability of sorafenib for treating patients. Regorafenib has been shown to offer a survival benefit as second line treatment and there are other promising drugs such as cabozantinib and nivolumab in the pipeline. If any marketing authorisations state that any of these new drugs can only be used after previous treatment with sorafenib, then these new drugs will be disqualified from NICE appraisal and any use in England would be off

label and thus subject to the extremely competitive NHS England Specialised Commissioning prioritisation process.

5. NHS England knows (as will the NICE TA Committee) that sorafenib is currently the only proven systemic therapy which is clinically effective in the treatment of heapatocellular carcinoma and thus is in all the national and international treatment guidelines for this disease. In the past, the Cancer Drugs Fund placed a special emphasis on those drugs that were the only proven systemic therapies for a particular cancer. This latter thinking now plays no part in NHS England in the decision making of Individual Funding Requests or in how it regards drugs referred to NICE for appraisal. What matters now is whether sorafenib is cost effective in this indication or not.

[redacted]

[redacted]

28 November 2016

Name	[redacted]
Role	Patient Organisation
Organisation	British Liver Trust
Job Title:	Chief Executive
Location	England
Conflict	Whilst the British Liver Trust does accept support from other pharmaceutical companies we are not currently in receipt of any funding or other support from Bayer

Comments on the ACD:

The British Liver Trust would be very disappointed and concerned if NICE were not to approve the use of Sorafenib for advanced Hepatocellular Carcinoma.

It is crucial NICE recognises that Hepatocellular Carcinoma (HCC) is the 18th most common cancer in the UK and accounts for approximately 85% of liver cancers. If HCC is detected early, potentially curative treatment options are available such as transplant or surgical removal but for advanced HCC there are no specific symptoms, and so less than 30% of patients are diagnosed in the early stages of the disease where potentially curative treatment is available.

Patients with advanced HCC have a very poor prognosis and given the lack of therapeutic alternatives, Sorafenib fulfils a key unmet clinical need for patients with advanced inoperable HCC; the alternative remains as best supportive or palliative care

Sorafenib is the only systemic treatment proven to increase survival in HCC, as well as providing quality of life benefits

Sorafenib does have well-known and manageable side effects. As an oral treatment it is convenient to administer and patients do not have to attend hospital for intravenous therapy

Sorafenib would only be needed for a relatively small patient population so the overall cost to the NHS will be small - there were approximately 3,867 cases of liver cancer in England in 2012, of these 85% will be HCC resulting in approximately 3,287 cases. Of those with HCC, a smaller sub-population (25-30%) will be eligible for Sorafenib. See: https://www.nice.org.uk/guidance/ta189/resources/costing-statement-378010765

Currently patient access in England has been possible through the CDF and from April 2013 to September 2015, 968 patients accessed Sorafenib via the CDF

As we have raised in previous submissions routine patient access in Scotland and Wales where Scottish Medicines Consortium (SMC) and All Wales Medicines Strategy Group (AWMSG) approved Sorafenib as cost-effective to treat patients with advanced HCC who have failed or are unsuitable for surgical or loco-regional therapies so if NICE does not approve there will be variation in access and standard of care across UK with English patients being disadvantaged. The Trust is also concerned about how the potentially negative NICE guidance would apply in Wales and affect current positive AWMSG guidance.

AS I was present, I would also argue that at the committee meeting Bayer gave a very poor representation of the desperate need for Sorafenib but I would hope that NICE would see why when there is no other systemic treatment patients with HCC would want Sorafenib to be an option that would lengthen their life and very

importantly improve their quality of life.

In summary and to answer NICEs main questions:

On behalf of patients with and at risk of HCC the British Liver Trust does not think that enough has been made of the significant benefits for patients of this life lengthening and life improving treatment.

The clinical evidence needs to be read with the evidence from patient organisations so that a full holistic view can be taken of the need and benefits of the use of Sorafenib

The proposed negative recommendation is not a sound or suitable basis for guidance to the NHS as it denies English patients access to the only treatment available for HCC and will lead to inequity in the treatment of HCC between English and Scottish & Welsh patients with advanced HCC.

Comments on the ACD Received from the Public through the NICE Website

Name	[redacted]
Role	NHS Professional
Organisation	University of Liverpool
Job Title:	Professor in Translational Oncology
Location	England
Conflict	No

Comments on the ACD:

I have a long standing interest in the management of hepatocellular carcinoma (HCC) and in trials of chemotherapy and targeted therapies. We have recently undertaken an individual patient data (IPD) meta-analysis of three prospective randomised controlled trials of Sorafenib in HCC. The results support previous contentions that sorafenib may have more impact in patients who are hepatitis C positive i.e. the viral status of a patient with HCC is predictive of benefit from sorafenib. Other data supporting such a contention is discussed and referenced in the paper referred to below.

The implications for NICE are:

a) There is little evidence of impact in other aetiologically-defined sub-groups.

b) Within hepatitis C virus sub-group it is possible that the drug may, according to NICEs guidelines, be cost-effective.

Please note that the relevant paper (abstract below) is not yet published. However, the paper has been very positively reviewed by the Journal of Clinical Oncology. An amended manuscript has been re-submitted (9 September 2016) to the Journal and there is a good chance that it will be accepted before the second appraisal committee meeting. I would be happy to provide a copy of the paper and the reviewers comments if requested to by the committee.

Impact of viral status on survival in patients receiving sorafenib for advanced hepatocellular cancer: A meta-analysis of randomised phase III trials

R. Jackson, E. Psarelli, S. Berhane, H. Khan, P. Johnson

Study published as abstract International Liver Congress 2016 Barcelona

Abstract

Purpose: Following the SHARP trial, sorafenib has become the standard of care for patients with advanced unresectable hepatocellular carcinoma (aHCC) but the relation between survival advantage and disease etiology remains unclear. To address this issue we undertook an individual patient data (IPD) meta-analysis of three large prospective randomised trials in which sorafenib was the control arm.

Patients and Methods: 3256 patients, 1643 (50%) of whom received sorafenib were available. The primary endpoint was overall survival (OS). A Bayesian hierarchical approach for IPD meta-analyses was applied using a piecewise exponential model. Results are presented in terms of hazard ratios (HRs) comparing sorafenib to alternative therapies according to hepatitis C virus (HCV) or hepatitis B virus (HBV) status.

Results: HRs show improved OS for sorafenib in patients who are both HBV negative and HCV positive (log (HR) (95%CI): -0.27 (-0.46, -0.06)). Median unadjusted survival is 12.6 (11.15, 13.8) months for sorafenib and 10.2 (8.88, 12.2) months for "other" treatments in this sub-group. There was no evidence of improvement in OS for any other patient sub-groups defined by HBV and HCV. Results were consistent across all trials with heterogeneity assessed using Cochrane's Q statistic.

Conclusions: There is consistent evidence that the affect of sorafenib on OS is dependent on patients' hepatitis status. There is an improved overall survival for HBV negative, HCV positive patients when treated with sorafenib. There was no evidence of any improvement in OS for HBV positive, HCV negative patients attributable to sorafenib.

Name	[redacted]
Role	Healthcare Other
Organisation	BTG
Job Title:	Senior Manager Market Access
Location	England
Conflict	No

Comments on the ACD:

BTG would like to provide comments to the committee in relation to the appraisal consultation document entitled Sorafenib for treating advanced hepatocellular carcinoma.

Since Sorafenib has been funded through the cancer drugs fund, it has been one of the very few treatment options available to patients with advanced hepatocellular carcinoma.

Given the high societal burden of liver cancer (i.e. death, disability and associated economic factors), there appears to be a disproportionately low level of funding for the available treatment options when compared with other common cancers, for example breast cancer.

Considering changing migration patterns, increasing alcohol consumption and an obesity epidemic, liver disease is becoming a growing problem in England, with liver cancer itself increasing by more than 50% in men and women in the last 10 years.

In addition to the increasing incidence of the disease, deaths from liver cancer in England are more prevalent in people living in the most deprived areas in the country. Although treatment options are currently limited, there is certainly more than can be done to ensure equitable access to care and funding for patients.

Historically, this patient group have often been overlooked when funding decisions are being made and there remains little commitment from the NHS to financially address this inequality. Raising the awareness, removing the stigma associated with liver disease and improving education is essential. Identifying high risk populations and intervening earlier in these populations could improve outcomes and reduce societal burden. Increased screening carries its own costs and risks and therefore the problem is complex and more research is required.

With Sorafenibs funding currently under review, patients and physicians are left with very few treatment options in intermediate and advanced HCC.

Areas of emergent practice such as radiopaque drug eluting bead trans arterial chemo embolisation (radiopaque DEB TACE) and ongoing research into Dendritic cells have the potential to offer better care, improve outcomes, and offer more choice to this patient group.

Selective Internal Radiation Therapy (SIRT) has been considered by NICE through both IPAC and the MIB process; both concluded that this therapy was safe and efficacious and had a place in therapy for patients yet funding for this therapy is still not available in the UK.

Providing this relatively small number of patients, treatment options is important. The funding of new technologies and treatments has potentially often been limited in the UK because of the lack of broader support. BTG feel that further consideration should be given to this patient group and emergent therapies such as these.

CANCER DRUGS FUND RAPID REVIEW OF NICE GUIDANCE TA189:

SORAFENIB FOR THE TREATMENT OF ADVANCED HEPATOCELLULAR CARCINOMA

CRITIQUE OF THE COMPANY'S RESPONSE TO THE ACD

BY THE DECISION SUPPORT UNIT

24 October 2016

Sabine Grimm and Iñigo Bermejo

School of Health and Related Research, University of Sheffield

Decision Support Unit, ScHARR, University of Sheffield, Regent Court, 30 Regent Street
Sheffield, S1 4DA

Tel (+44) (0)114 222 0734
E-mail dsuadmin@sheffield.ac.uk
Website www.nicedsu.org.uk
Twitter @NICE_DSU

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) is a collaboration between the Universities of Sheffield, York and Leicester. We also have members at the University of Bristol, London School of Hygiene and Tropical Medicine and Brunel University. The DSU is commissioned by The National Institute for Health and Care Excellence (NICE) to provide a research and training resource to support the Institute's Technology Appraisal Programme. Please see our website for further information www.nicedsu.org.uk.

The production of this document was funded by the National Institute for Health and Care Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

Acknowledgements

We wish to thank Sarah Davis and Paul Tappenden, ScHARR, for providing comments on the draft report, Nick Latimer, for his helpful comments on survival analysis and Jenny Dunn, ScHARR, for providing administrative support and preparing the report.

This report should be referenced as follows:

Grimm S, Bermejo I. Cancer drugs fund rapid review of NICE guidance TA189: Sorafenib for the treatment of advanced hepatocellular carcinoma. Critique of the company's response to the ACD. School of Health and Related Research (ScHARR), 2016.

Use of confidential data

Any 'commercial in confidence' data provided by the company, and specified as such, is highlighted in blue and underlined in the review. Any 'academic in confidence' data provided by the company, and specified as such, is highlighted in yellow and underlined in the review.

EXECUTIVE SUMMARY

In response to the Appraisal Consultation Document (ACD), the company offered a reduced Commercial Medicines Unit (CMU) price of [redacted] per pack of sorafenib (representing a further [redacted] discount and a total discount of [redacted] to the list price). The company also presented new evidence to address uncertainties identified during the Appraisal Committee meeting: (1) treatment duration estimates based on actual individual patient-level data (IPD) from SHARP; (2) an analysis based on the GIDEON study using propensity scores to match the baseline characteristics of patients enrolled in the SHARP trial in order to further inform the choice of curve for extrapolating overall survival (OS); and (3) feedback from two hospitals on sorafenib wastage. Furthermore, the company argued against two of the AC's preferred analyses, notably: (1) using treatment duration estimates based on treatment duration in SHARP, stating that treatment was longer in SHARP than it is in clinical practice, and; (2) the use of resource use estimates based on pooling the estimates from the original appraisal and those of the rapid reconsideration.

Analysis using IPD from SHARP of time to treatment discontinuation

The company fitted parametric curves to the IPD of time to treatment discontinuation and concluded that, based on AIC/BIC scores, the log normal curve provided the best fit. The DSU considers that the Weibull provides a more plausible extrapolation for treatment duration based on visual inspection and external data provided by the company. The DSU believes that the extrapolation based on the log normal model overestimates treatment duration.

The company's preference for using time to progression (TTP) as a proxy to treatment duration

The company argued for the use of TTP as a proxy for treatment duration instead of directly using treatment duration based on the SHARP, claiming that treatment duration was greater in SHARP than in clinical practice. The estimates of mean and median treatment duration reported from external sources were considered by the DSU to be inconclusive in supporting this claim. The company furthermore justified the apparently longer treatment duration estimated in SHARP suggesting that post-progression treatment occurred in SHARP due to the uncertainty about the post-progression benefits of sorafenib. The company argued that

patients would no longer be treated post-progression in clinical practice and therefore the TTP based on investigator assessment was more reflective of clinical practice. The DSU believes that: (1) it is unlikely that the alleged reduced benefit of post-progression treatment would have an important impact, considering the similarity between the survival curves of TTP based on the independent assessment and treatment duration based on IPD; (2) that both efficacy estimates and treatment duration estimates should be based on the SHARP trial; (3) and that it is consistent with the original AC's preference to include treatment costs for patients who had been treated after progression.

Analysis of matched GIDEON population to inform choice of overall survival model

The DSU considered the company's approach to matching the GIDEON population to SHARP to be satisfactory. The DSU notes that the resulting matched GIDEON population exhibited a longer OS than that observed in SHARP. The company fitted log normal and Weibull curves to the Kaplan-Meier (KM) data for the adjusted population and concluded that, based on AIC/BIC scores, the log normal provided a better fit than the Weibull curve. Based upon visual inspection of the fitted curves plotted against the KM curve, the DSU believes that the log normal function would overestimate OS whilst the Weibull would underestimate it. Therefore, both curves should be considered in the extrapolation of OS to estimate the most plausible incremental cost-effectiveness ratio (ICER) for sorafenib compared with best supportive care (BSC).

Hospital feedback on sorafenib wastage

The hospitals' statements coincided in that one month's supply of sorafenib is typically prescribed at a time and wastage is avoided as much as possible by pack splitting and weekly reviews in some patients. The company presented results for analyses including the wastage of up to seven days' worth of treatment. The DSU considers it unlikely that absolutely no wastage would be incurred but also considers the company's approach to incorporating wastage to be acceptable.

The company's preference for using only updated resource use estimates

The company claimed that resource estimates from the original appraisal were no longer accurate because of significant changes in clinical practice in recent years, and that the AC's preferred pooled resource use estimates would therefore introduce bias. The DSU notes that

the company has not provided any new evidence to support their claims and that it could have provided further new estimates that would have served to provide a more reliable resource use estimate. Furthermore, the updated resource use estimates exhibited considerable variation and a small number of experts.

The company's analyses

The company provided new analyses using the newly proposed price for sorafenib and using a series of alternative assumptions. The ICER for sorafenib compared with BSC was estimated to be £35,695 per quality-adjusted life year (QALY) based on the company's preferred assumptions, which included extrapolating OS using the log normal model, the investigator assessment of TTP as a proxy for treatment duration, only new resource use estimates and no wastage. Using the AC's preferred assumptions, the ICER presented by the company was £49,060 per QALY gained, based on using the log normal curve for extrapolating OS, the independent reviewer assessment for TTP and treatment duration based on SHARP IPD (fully parametric curve, log normal), pooled resource use estimates and up to seven days of wastage.

The DSU's exploratory analyses

The DSU undertook exploratory analyses using the new proposed price and considering the new evidence presented by the company. Using the AC's preferred assumptions as its own base-case and correcting for a minor error in the company's model, the DSU's base case resulted in an ICER for sorafenib compared with BSC of £49,299 per QALY gained. However, the DSU notes that the ICER of sorafenib compared with BSC is estimated to be £72,596 per QALY when the Weibull distribution is used for the extrapolation of both OS and treatment duration. The DSU believes that even if the log normal fits the data slightly better, it is likely to lead to an overestimation of OS and treatment duration whilst the Weibull is likely to underestimate them. Therefore, the DSU believes that the most likely ICER lies between £49,299 and £72,596 per QALY, and that it is likely to be closer to the lower end of the range.

ABBREVIATIONS

ACD	Appraisal consultation document
AE	Adverse event
AIC	Akaike Information Criterion
BCLC	Barcelona Clinic Liver Cancer
BIC	Bayesian Information Criterion
BSC	Best supportive care
CDF	Cancer Drugs Fund
DSU	Decision Support Unit
ECOG	Eastern Cooperative Ongology Group
GIDEON	Global Investigation of therapeutic DEcisions in hepatocellular carcinoma and Of its treatment with sorafeNib
ICER	Incremental cost-effectiveness ratio
IPD	Individual patient-level data
KM	Kaplan-Meier
LRIG	Liverpool Reviews and Implementation Group
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OS	Overall survival
PFS	Progression-free survival
PS	Performance status
QALY	Quality-adjusted life years
SHARP	Sorafenib Hepatocellular Carcinoma Assessment Randomized Protocol
TSD	Technical Support Document
TTP	Time to progression

1. CRITIQUE OF THE NEW EVIDENCE PRESENTED BY THE COMPANY

1.1. EXTRAPOLATION OF OVERALL SURVIVAL

In the Appraisal Consultation Document (ACD), the Appraisal Committee (AC) concluded that using the Weibull function to extrapolate OS retained plausibility. The company's response to the ACD[1] argues that only the log normal distribution should be used for OS extrapolation and that the Weibull distribution should not be taken into account in any consideration of uncertainty. The company's submission for the rapid reconsideration of sorafenib presented data from the GIDEON study[2] and Palmer *et al.* [3] to support the log normal extrapolation for OS. In their response to the ACD[1], the company argued against the AC's conclusion that Palmer *et al.*[3] was likely to be confounded and presented an analysis matching the baseline characteristics of patients in the GIDEON study[2] to those in the SHARP trial using propensity scoring[4].

1.1.1. Palmer et al.

Palmer *et al.*[3] is a retrospective study that compares the outcomes of patients whose funding applications for sorafenib were approved with those whose applications were refused. The AC concluded that Palmer *et al.*[3] was not suitable to validate the extrapolation of OS beyond SHARP because the results were likely to be confounded. The company argues against this conclusion and claims that potential biases were controlled for in Palmer *et al.*[3]: the same inclusion criteria were applied for funded and non-funded patients; Palmer *et al.*[3] claimed that the sorafenib funding decision was "not apparently based on clinical variables", and; that there was no statistically significant difference in patient characteristics at baseline. The company also claims that the study "provides reasons to doubt that the provision of funding was a confounder." The DSU notes that: (1) it is likely that the decision to approve or deny funding for a patient was based on the expected benefit for the patient and that therefore it cannot be assumed that patients were randomly allocated to funded and non-funded groups, and; (2) the balance between treatment arms in non-randomised studies cannot simply be confirmed by looking for differences in known confounders because there may be unknown confounders that could only have been balanced through randomisation. In fact, it is plausible that the considerably higher efficacy of sorafenib in Palmer *et al.*[3] compared with SHARP (hazard ratios: 0.48 vs 0.69) could be explained by selection bias. The company does not present any new relevant evidence in this response. The DSU refers to its original report[5]

for arguments on why Palmer *et al.*[3] should not be considered suitable for decision-making. The DSU also refers to the recently published study by Grieve *et al.*[6], which cautions that reliance on "real world" observational data undermines the evidence base for clinical practice, with observational studies typically not appropriately measuring key characteristics leading to biased estimates of effectiveness due to residual confounding.

Regarding the choice of the parametric curve to extrapolate OS, the company argues that the log normal curve predicts the plateau at the end of the KM in Palmer *et al.*[3] better than the Weibull model (Figure 2 in company's response to the ACD[1]). The company points out the four censoring events after day 600 to note that there are still patients at risk at this stage. However, the DSU notes that: (1) the plateau is made up of a small number of events; (2) that the tail is likely to fall well within both curves' confidence intervals; (3) that the apparent plateau generated by plotting censoring after the last event does not provide much information and should therefore only be considered with caution, and; (4) that the Weibull seems to fit the data better based on visual inspection. Therefore, the DSU still believes that Palmer *et al.*[3] provides no evidence to favour one curve over the other.

The company also provides a scenario analysis using the Palmer *et al.*[3] data. Given the potential for confounding in Palmer *et al.*[3], the ICERs presented by the company based on the data from Palmer *et al.*[3] were not considered by the DSU to be robust and are subsequently not reported here.

1.1.2. GIDEON

The AC considered that the baseline characteristics of the patient population in the GIDEON study[2] differed significantly from the patient population in the SHARP trial[4]. The committee stated that "it would have been appropriate for the company to modify the GIDEON population to reflect the characteristics of the population in SHARP"[7]. In their response to the ACD[1], the company presents an analysis in which the patient characteristics of GIDEON were matched to the treatment arm in the SHARP trial using propensity score matching based on individual patient-level data[1]. The company presents results from two different analyses. In one analysis, a 3:1 matched dataset was used (that is, for each patient in SHARP, the three closest matches from GIDEON were used in the analysis) resulting in a matched GIDEON population of n=895. In the other analysis, a 1:1 matched dataset was used

(where only one GIDEON patient was matched to each SHARP patient). The 3:1 matching approach was used in the company's preferred analysis based on the matched population. In both approaches, patients were matched based on age, gender, Child-Pugh status (Class A, Class B), Barcelona Clinic Liver Cancer (BCLC) and Eastern Cooperative Ongology Group (ECOG) status. Propensity scores were generated using a multivariable logistic regression model. The DSU reviewed the description of the methods and was satisfied with the matching performed by the company.

The company presents the resulting KM estimates for the matched GIDEON dataset plotted against the KM estimates from the SHARP trial. The DSU notes that the mean survival of patients in GIDEON is still higher than that in SHARP. The company fitted log normal and Weibull curves to the matched GIDEON data and presented their respective Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) scores. The company notes how both AIC and BIC scores indicate that the log normal model fits the KM data slightly better. However, the goodness of fit based on AIC/BIC should not be used in isolation to inform the choice of survival functions[8]. The DSU plotted the log normal and Weibull curves fitted by the company against the KM of the matched GIDEON population (Figure 1).

Figure 1:

The DSU notes that the log normal fits the data slightly better than the Weibull during the first 600 days (in fact, the Weibull overestimates survival during most of this period), whilst the Weibull fits the data better than the log normal after day 600 and until the plateau (around day 726). The DSU notes that this is likely the reason for the better AIC/BIC scores assigned to the log normal function. However, upon examination of the plot, the DSU is not convinced that the log normal provides the most plausible extrapolation of the OS. Based on visual inspection of Figure 1, the DSU believes that the log normal curve overestimates long-term survival whilst the Weibull underestimates it.

1.1.3. Conclusions

In conclusion, the log normal curve appears to fit the KM data slightly better than the Weibull both for SHARP[4] and matched GIDEON populations. However, the DSU considers that the use of the log normal model for the extrapolation of OS would lead to its overestimation and therefore considers that the Weibull model, which would underestimate OS, should also be used to address this uncertainty.

1.2. DURATION OF TREATMENT

1.2.1. Modelling of treatment duration based on patient level data from SHARP

Treatment duration is an important driver of the ICER of sorafenib compared with BSC. In the company submission for the rapid reconsideration of sorafenib, treatment duration was determined by TTP, which was reported both based on investigator assessment and independent reviewer assessment. The company argued that investigator assessment was more appropriate, since investigators assessed when a patient had progressed during the treatment. However, treatment was not necessarily discontinued when a patient progressed, if the investigator deemed the patient to potentially benefit from further treatment. The DSU argued that using the independent reviewer assessment for estimating TTP was more appropriate as a proxy for treatment duration, since the median TTP based on independent assessment (24 weeks) was closer to the median treatment duration (23 weeks) than that

based on investigator assessment (17 weeks). It is noteworthy that within the model, TTP also informs the health-related quality of life gain a patient experiences.

In line with the Appraisal Committee's request, the company presents an analysis estimating treatment duration directly based on IPD for time to treatment discontinuation from SHARP[4]. The DSU favours this approach because it avoids the use of surrogates (such as TTP) for treatment duration: instead, this analysis used data on patient discontinuation based on events from the trial. The company performed survival analysis by applying five parametric models for the extrapolation of treatment duration (exponential, Gompertz, log logistic, log normal and Weibull functions). Based on goodness of fit as measured by AIC and BIC, the company chose the log normal model. The estimated mean treatment duration based on the log normal is [redacted]. The DSU notes that (1) AIC/BIC criteria should not be used in isolation to decide on model fit[8]; (2) differences in AIC/BIC criteria between Weibull, Gompertz and log normal model are relatively small, and; (3) upon visual inspection of Figure 4 of the company's response to the ACD[1], the Weibull and Gompertz both appear to match the latter part of the KM curve better than the log normal. The use of the Gompertz and Weibull curves results in a mean treatment duration of [redacted] months. The company unfortunately did not provide analyses based on these curves.

The company also presented an analysis based on a "hybrid" approach, which used the actual KM data when available, and the hazard from the fitted curve from the last observed event onwards. The DSU notes that the company's approach is significantly different to the "hybrid" approach the company referred to, which was called the Liverpool Reviews and Implementation Group (LRIG) exponential in the Technical Support Document (TSD) for survival analysis[8]. The purpose of the LRIG exponential method was to address the peculiarities of patient survival in lung cancer (TA181). Because all parametric extrapolation methods resulted in a poor fit between extrapolated curves and the KM estimates, the LRIG used KM estimates where available and assumed that the hazard was linear at the end of the KM curve and fitted an exponential curve beyond that time point. The company's hybrid approach differs from this approach by using the log normal fitted to the whole KM function but then only using the hazard from the last event onwards without indicating a rationale for why the whole extrapolated curve should not be used. More importantly, the fitted log normal curve does not pass through the end of the KM but instead lies above it. Therefore, applying the hazard of the log normal model starting from the end of the KM underestimates the area

under the curve, potentially underestimating the ICER of sorafenib versus BSC. As pointed out by Davies *et al.* [9], the KM estimates in the tail are subject to particular uncertainty as the number of patients at risk decreases. For this reason, Davies *et al.* [9] conclude that methods of survival estimation based on fitting parametric distributions to the full range of the data are often preferred. The DSU therefore favours the fully parametric approach over the hybrid approach.

1.2.2. Use of time to progression as a proxy for treatment duration

The company continues to use progression free survival (PFS) resulting from TTP based on the investigator assessment as a proxy for treatment duration. The company argues that the analysis based on the time to treatment discontinuation IPD from SHARP results in a mean treatment duration which is higher than that observed in clinical practice as reported by or calculated from other sources. However, the DSU believes that the evidence provided is inconclusive: (1) the treatment duration estimate based on patient notifications for sorafenib within the Cancer Drugs Fund (CDF) is confounded by not all notifications resulting in treatment and by sorafenib being used in another indication; (2) the results of King *et al.*[10] include only median treatment duration (3.2 months), not the mean, and the baseline characteristics of patients show worse prognosis than those of SHARP (e.g. Child-Pugh A 84% vs 97%, ECOG performance status (PS) 0: 30 vs 54); (3) Palmer *et al.*[3] also reported only median treatment duration and this was close to that observed in SHARP (5.1 vs 5.3 months), and; (4), the mean treatment duration reported for GIDEON ([redacted] months) is based on a population with a worse prognosis to that of SHARP.

The company also argues that mean treatment duration in SHARP was longer than in clinical practice because at the time of the study, post-progression treatment benefit was unknown and "investigators would have had an ethical consideration to treat as long as there may have been a benefit and treatment could be tolerated."[1] Therefore, the company believes that post-progression treatment costs from SHARP should be ignored. The DSU notes that it is unlikely that the alleged reduced benefit of post-progression treatment would have an important impact, considering the similarity between the survival curves of TTP based on the independent assessment and treatment duration based on IPD as shown in Figure 2.

Figure 2: Comparison between the survival curves for treatment duration and TTP based on independent assessment

The DSU also notes that estimating treatment duration based on actual IPD from SHARP : (1) was part of the AC's preferred assumptions; (2) makes better use of existing evidence; and (3) ensures consistency between treatment duration estimates and efficacy estimates, which are also informed by the SHARP study. The DSU therefore considers estimating treatment duration from SHARP IPD as appropriate. This is also consistent with the original AC's conclusion that the treatment costs for the [redacted] of patients who were treated post-progression should also be accounted for. However, the DSU notes that the company could have used the reported external data on mean treatment duration to inform its choice of parametric curve for the extrapolation. Using a Gompertz or a Weibull instead of a log normal distribution results in a mean treatment duration which is closer to those estimates reported in the external references provided by the company and discussed above.

In conclusion, the DSU believes that: (1) treatment duration should be estimated based on SHARP IPD; (2) a fully parametric approach should be used in preference over a "hybrid" approach, and; (3) the Weibull and Gompertz distributions appear to provide more plausible extrapolation of treatment duration than the log normal based upon visual inspection and consideration of external data.

1.3. RESOURCE USE ESTIMATES

The company argues against the AC's conclusion that pooling the original and new resource use estimates was appropriate. The company argues that the introduction of sorafenib as standard of care and the nine years since the original survey has implications for resource use in the BSC arm. However, the DSU notes that the company fails to provide evidence for any such changes since the original survey. The DSU does not consider the resource use estimates based on the new survey as sufficient justification for these changes because the two parameters that changed the most and had considerable impact on the ICER of sorafenib compared to BSC were only informed by two clinicians and there was considerable variation between their estimates. The DSU notes that the company has not provided any new evidence in their response to support only using the new estimates and ignoring those of the original submission. The DSU also notes that, in their response, the company could have provided additional estimates from experts to increase the sample size.

In their response to the ACD[1], the company also mentioned a "hybrid" approach implemented by the DSU in their amended version of the model, whereby the pooled resource use estimates were used for the BSC group and the updated estimates were used for the sorafenib group. The DSU discarded this approach while writing their original report because using a different set of experts to estimate the resource use in each arm could bias the results. In the absence of evidence for a change in the use of resources, the DSU continues to favour the use of pooled resource use estimates to address the variation between the consulted experts.

1.4. WASTAGE

The AC expressed its preference for considering wastage in the model. The company contacted hospitals to inform its wastage estimates and presented their statements. The DSU notes that these statements are almost identical in some places. The DSU notes that both statements coincide in that one month's worth of sorafenib is prescribed at a time. The company argued that, given that patients would have monthly consultations, when a decision about the continuation of the treatment would be taken, the termination of treatment would match the date of prescription. However, the company presented results for analyses including the wastage of up to seven days' worth of treatment.

The DSU considers it unlikely that absolutely no wastage would be incurred. For example, a small proportion of patients will die while they are still on treatment, or others might discontinue treatment in the middle of the month due to multiple possible reasons such as a sudden deterioration of their health. The DSU is satisfied with the company's inclusion of up to seven days of wastage but, in order to assess the existing uncertainty, the DSU presented a pessimistic scenario analysis whereby it is assumed that half a package of sorafenib on average will be wasted by each patient at the end of the treatment.

2. SUMMARY OF THE ANALYSES PRESENTED BY THE COMPANY

The company undertook new analyses using the newly proposed price of sorafenib. Table 1 shows a summary of the most important analyses as presented by the company in their executive summary. These scenario analyses include one based on the assumptions preferred by the company and two based on the assumptions the company believed best reflected (1) the AC's preferences as expressed in the ACD and (2) the DSU's preferences as reflected in its original report. The DSU notes that the company did not present results for exploratory analyses using the Weibull distribution for the extrapolation of OS and duration of treatment.

Table 1: Summary of analyses presented by the company based on new price (deterministic)

Description of scenario	Scenario details	Cost per QALY
Company's base case:	• Extrapolation of OS: Log normal • TTP: investigator assessment • Treatment duration: based on investigator TTP • Resource use: only new estimates • No wastage.	£35,695
DSU's preferred base case in original report (originally £51,208):	• Extrapolation of OS: Log normal • TTP: independent assessment • Treatment duration: based on independent TTP • Resource use: pooled estimates • No wastage.	£46,863
AC's preferred assumptions	• Extrapolation of OS: Log normal • TTP: independent assessment • Treatment duration: based on SHARP patient level data (fully parametric curve, log normal) • Resource use: pooled estimates • Up to seven days of wastage	£49,060

In addition, the company presented a number of exploratory analyses combining alternative assumptions for treatment duration, resource use, and wastage. The lowest ICER for sorafenib versus BSC in these scenario analyses corresponded to the company's base case (apart from that using OS estimated from Palmer *et al.*[3]); the highest ICER corresponded to that based on the AC's preferred assumptions.

The DSU notes that the analyses presented by the company using a hybrid approach for extrapolation are flawed (Sections 5.3.1 and 5.3.2 of the company's response to the ACD[1]): hybrid extrapolation of treatment duration, TTP and OS as implemented by the company suffer from the issues described in Section 1.2.1. In addition, for TTP extrapolation, the company used KM data based on the hybrid assessment of TTP (not to be confused with the hybrid approach to extrapolation): the hybrid assessment of TTP involved using independent assessment of progression when it was available and using investigator assessment elsewhere. The DSU considers that the company should have used KM data for TTP based on independent reviewer assessment instead of hybrid assessment. Appendix A elaborates on the description of this issue.

The company also presented results for a scenario analysis based on Palmer *et al.*[3], which resulted in much lower ICERs for sorafenib vs BSC (£18,870 per QALY gained based on the company's preferred assumptions, £25,965 per QALY gained based on the AC's preferred assumptions) than those calculated from SHARP. The DSU does not believe these results should be considered to inform decision-making as explained in Section 1.1.1.

3. ADDITIONAL ANALYSES UNDERTAKEN BY THE DSU

The DSU undertook some exploratory analyses based on those presented by the company. The DSU adopted the scenario denoted "AC's preferred assumptions" by the company as its own base case. This included:

- Extrapolation of OS based on the log normal distribution
- TTP based on independent reviewer assessment
- Treatment duration extrapolation based on patient level data for treatment duration from SHARP (fully parametric curve, log normal)
- Resource use: pooled estimates from the original appraisal and the new submission

- Up to seven days of wastage

While reviewing the changes applied by the company to the model, the DSU identified a minor error in the calculation of the costs of adverse events (AEs) in the sorafenib arm: the costs of AEs of the patients who discontinued treatment were not being taken into account. Appendix B contains more details on this error and the correction applied by the DSU. After rectifying this error, the ICER of sorafenib compared with BSC in the base case increased slightly from £49,060 to £49,299 per QALY gained.

The DSU undertook the following scenario analyses based on its base case (the AE costing error was rectified, where applicable, that is, whenever duration of treatment was based on SHARP IPD):

1. Extrapolation of OS using the Weibull distribution.
2. Extrapolation of treatment duration using the Weibull distribution. The DSU did not have access to the curve fitted by the company and undertook this analysis by subtracting the treatment cost for the difference in the mean treatment durations estimated by the company for the log normal and the Weibull ([redacted] vs [redacted] months respectively).
3. Combination of scenarios 1 and 2: extrapolation of OS and treatment duration using the respective Weibull distributions.
4. Wastage: on average, half a pack of wastage of sorafenib per patient. The DSU acknowledges this is a worst-case scenario.

The results for the DSU's exploratory analyses are described in Table 2.

Table 2: Results of DSU exploratory analyses, based on the AC's preferred assumptions

Scenario			Total QALYs	Inc. QALYs	Total costs	Inc. costs	ICER
	Base case*	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£49,299
	Base case* (probabilistic)	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£49,239
1	Extrapolation of OS: Weibull	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£87,091

2	Extrapolation of treatment duration: Weibull	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£41,935
3	Weibull OS and DoT	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£72,596
4	Wastage: half a pack	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£50,884

*Equivalent to the "AC's preferred assumptions" scenario company's response to the ACD, but with the AE costing error fixed.

4. CONCLUSIONS

In response to the ACD, the company proposed a new price per sorafenib pack of [redacted] (representing a further [redacted] discount), provided new evidence to address key uncertainties identified during the rapid reconsideration and argued against some of the AC's conclusions.

The key new evidence presented by the company was twofold. First, in order to inform the choice of parametric curve for the extrapolation of OS, the company presented an analysis where it adjusted the population of the GIDEON study to match the baseline characteristics of the population enrolled in the SHARP trial using propensity score matching. The company then fitted log normal and Weibull curves to the KM curve of the adjusted population and concluded that, based on AIC/BIC scores, the log normal provided a better fit. The DSU, upon visual inspection of the fitted curves plotted against the KM curve, believes that the log normal function would overestimate OS whilst the Weibull would underestimate it. Therefore, the DSU concludes that both curves should be considered in the extrapolation of OS to estimate the most plausible ICER for sorafenib compared with BSC.

Second, following the AC's recommendation, the company provided an analysis using IPD from SHARP of time to treatment discontinuation to estimate treatment duration. The company fitted parametric curves to the IPD and concluded, based on AIC/BIC scores, that the log normal curve provided the best fit. The DSU believes that the Weibull provides a more plausible extrapolation for treatment duration based on visual inspection and on external data provided by the company. However, the company argued against using treatment duration estimates based on SHARP, claiming that treatment duration was longer in SHARP than in clinical practice. The company presented estimates of mean and median

treatment durations reported elsewhere to support their claim, but the DSU considers this evidence to be inconclusive. The company justified the longer treatment duration estimated in SHARP suggesting that patients had been treated post-progression due to the uncertainty at the time of the study on the post-progression benefits of sorafenib. The company argued that patients would no longer be treated post-progression in clinical practice and therefore the TTP based on investigator assessment was more reflective of clinical practice. The DSU notes that for the sake of consistency with the efficacy estimates, treatment duration should be based on the SHARP trial. However, the DSU believes that the extrapolation based on the log normal model overestimates treatment duration.

The company argues against the use of resource use estimates based on pooling the estimates from the original appraisal and those of the rapid reconsideration. The company claims that there had been a significant change in clinical practice in the past years and therefore the estimates from the original appraisal were no longer accurate. The DSU notes that the company has not provided any new evidence to support their claims and that it could have provided further new estimates that would have served to provide a more reliable resource use estimate.

The company provided new analyses using the newly proposed price for sorafenib and using some alternative assumptions. The ICER for sorafenib compared with BSC was estimated to be £35,695 per QALY gained based on the company's preferred assumptions and £49,060 per QALY gained based on the AC's preferred assumptions.

The DSU undertook exploratory analyses using the new proposed price and considering the new evidence presented by the company. The DSU believes that the most likely ICER for sorafenib compared with BSC lies between the DSU's base case of £49,299 per QALY gained and £72,596 per QALY gained, but is likely to be closer to the lower end of the range. The latter was estimated based on the Weibull distribution for the extrapolation of both OS and treatment duration.

5. REFERENCES

1. Bayer plc., *Sorafenib for treating advanced hepatocellular carcinoma. Response to the Appraisal Consultation Document (ACD)*. 2016.
2. Bayer CSR, *GIDEON: Global Investigation of therapeutic DEcisions in hepatocellular carcinoma and Of its treatment with sorafeNib (CONFIDENTIAL)*. 2013.
3. Palmer, D.H., et al., *Sorafenib for advanced hepatocellular carcinoma (HCC): impact of rationing in the United Kingdom.* Br J Cancer, 2013. **109**(4): p. 888-890.
4. Llovet , J.M., et al., *Sorafenib in Advanced Hepatocellular Carcinoma.* New England Journal of Medicine, 2008. **359**(4): p. 378-390.
5. Grimm, S., et al., *Cancer drugs fund rapid review of NICE guidance TA189: sorafenib for the treatment of advanced hepatocellular carcinoma*. 2016, School of Health and Related Research (ScHARR).
6. Grieve, R., et al., *Cancer Drugs Fund requires further reform.* BMJ, 2016. **354**.
7. National Institute for Health and Care Excellence, *Sorafenib for treating advanced hepatocellular carcinoma. Cancer Drugs Fund reconsideration of TA189. Appraisal consultation document.* 2016.
8. Latimer, N., *NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data*. 2011.
9. Davies, A., et al., *The Ends Justify the Mean: Outcome Measures for Estimating the Value of New Cancer Therapies.* Health Outcomes Research in Medicine, 2012. **3**(1): p. e25-e36.
10. King, J., et al., *Outcomes for UK patients with advanced hepatocellular carcinoma*, in *NCRI Cancer Conference*. 2013: Liverpool.

APPENDIX A

The company presented an analysis in Section 5.3.2 of their response to the ACD[1], in which a hybrid approach for the extrapolation of treatment duration, TTP and OS was used. In this approach, KM data was used for the timespan where it was available and a hazard based on a survival curve fitted to the whole KM was used from the last observed event in the KM onwards. A critique of this approach has already been outlined in Section 1.2.1.

In addition, the DSU notes that for the extrapolation of TTP, a KM based on the hybrid assessment of TTP is used by the company instead of the KM based on the independent assessment, favoured by the AC. The hybrid assessment of TTP used the independent assessment when available and the investigator assessment elsewhere. The hybrid assessment resulted in higher mean TTP compared with the independent assessment. As shown in Figure 3 and Figure 4, the curve based on the independent assessment does not fit the KM data based on the hybrid assessment, neither in the sorafenib nor in the BSC arm. The DSU notes that it is inconsistent to use KM based on the hybrid assessment and the hazard from the lognormal fitted to independent assessment, and that the use of the KM function resulting from the hybrid assessment of TTP does not reflect the AC's preferences.

Figure 3: KM data of hybrid assessment of TTP plotted against curve fitted to independent assessment data in the sorafenib arm

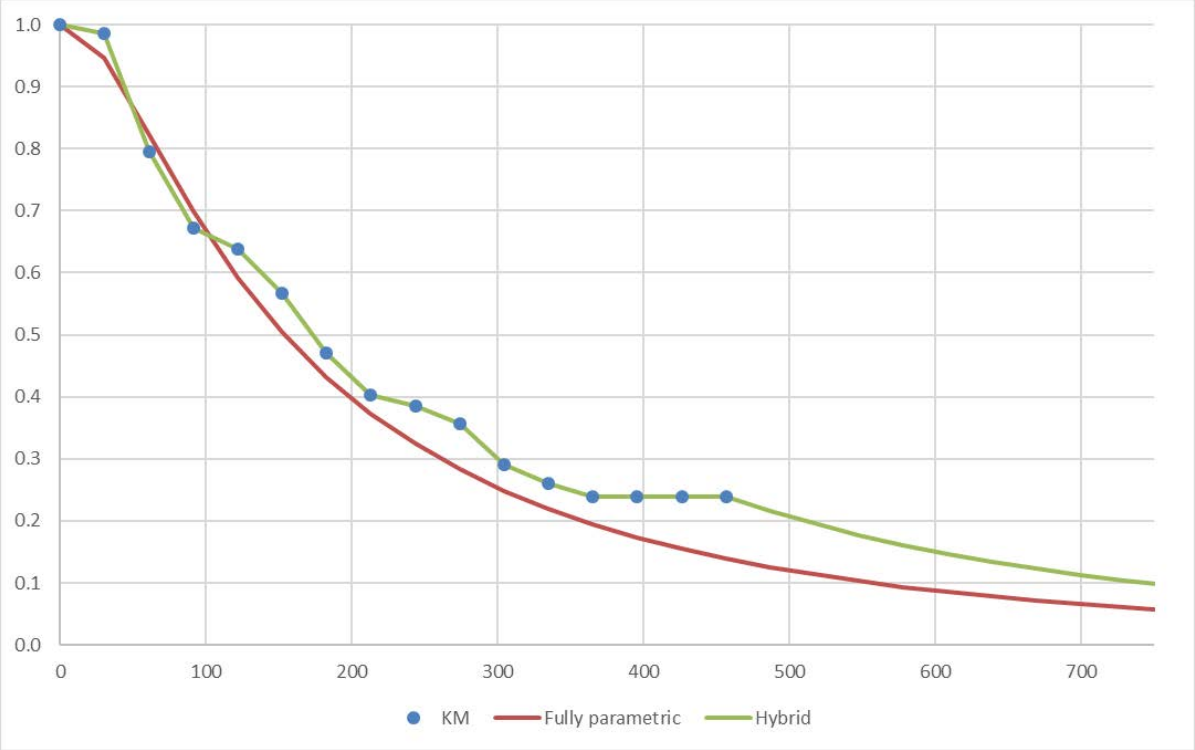
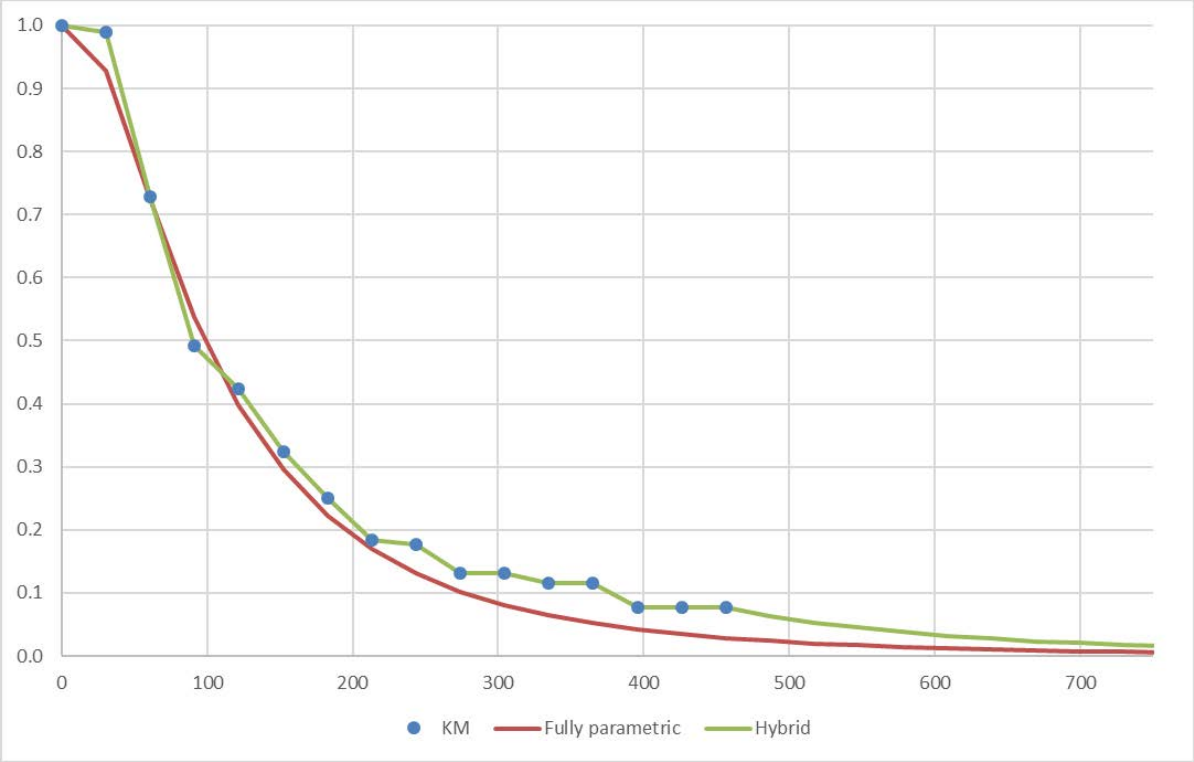


Figure 4: KM data of hybrid assessment of TTP plotted against curve fitted to independent assessment data in the sorafenib arm



The DSU plotted the KM and the hybrid extrapolation along the fully parametric approach for OS to illustrate one of the shortcomings of the hybrid approach applied by the company. The hybrid approach fits a curve to the time point where the last event occurred, when usually only a reduced number of patients remain at risk. This implies that, using the hybrid approach, a single event or a small number of events at the end of the curve, such as the sudden plunge before the last event in the KM of the BSC arm (Figure 5) can have a significant impact on the area under the curve. This can lead to significant inconsistencies between the resulting trends in the different curves. In Figure 6, for instance, the KM of the sorafenib arm shows a smaller plunge but followed by a short plateau with no events, which significantly alters the starting point of the extrapolation curve.

Figure 5: KM, hybrid and fully parametric approaches of overall survival in the BSC arm

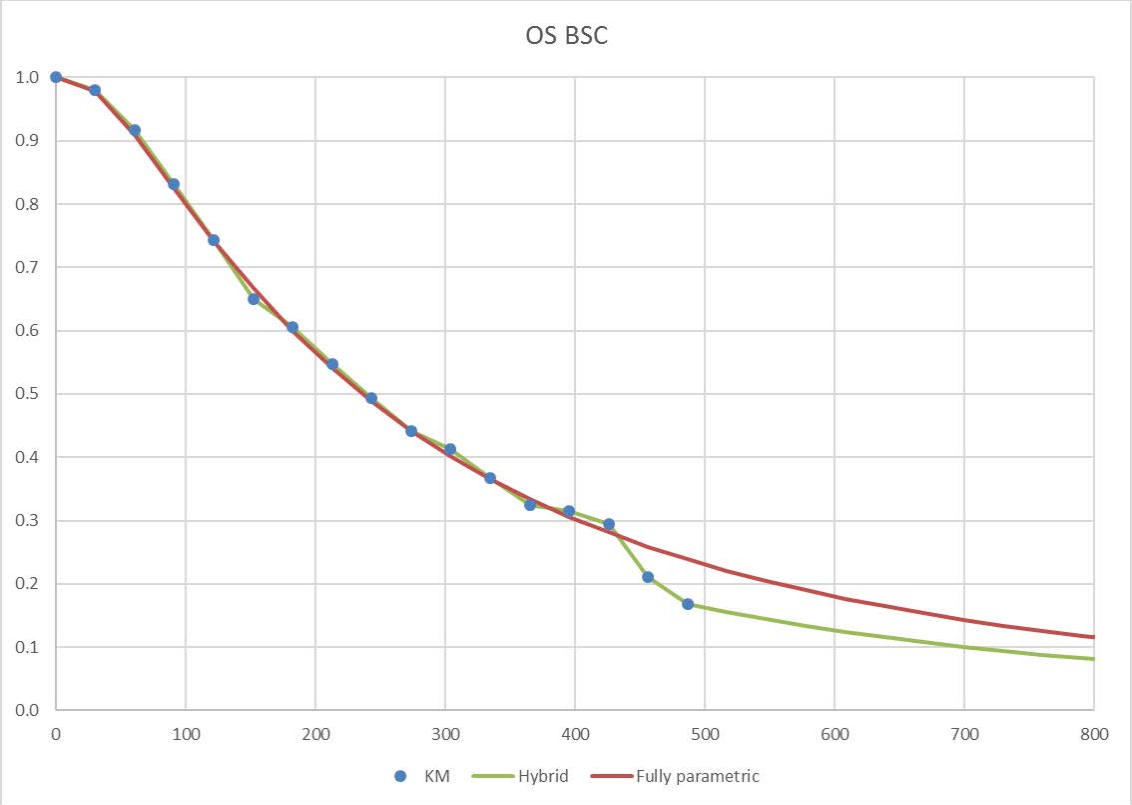
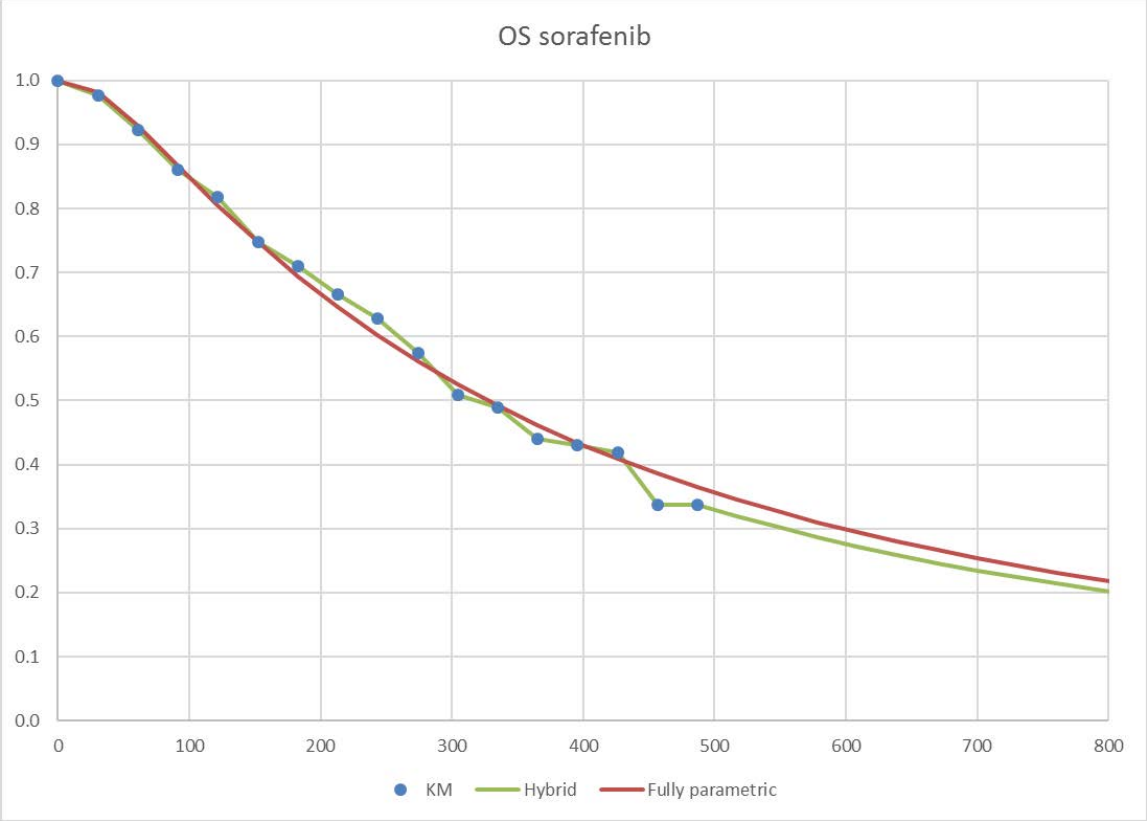


Figure 6: KM, hybrid and fully parametric approaches of overall survival in the sorafenib arm



APPENDIX B

The DSU believes that the new version of the model contained a minor error in the calculation of the cost of adverse events in the sorafenib arm. When treatment duration was extrapolated using duration of treatment data from SHARP instead of using TTP as a proxy, the cost associated with AEs for those patients who had discontinued sorafenib treatment were omitted. In the additional analyses undertaken by the DSU, these costs were added as explained below, resulting in a small increase in the ICER.

In the "Model" tab, in cell AV25, the DSU replaced the following formula:

=CHOOSE(PFS_or_DoT,(AB25*c_AEso+AC25*c_AEbsc),(BC25*c_AEso))

With:

=CHOOSE(PFS_or_DoT,(AB25*c_AEso+AC25*c_AEbsc),(BC25*c_AEso**+(1-BC25-AE25)*c_AEbsc)**)

Where BC25 contained the patients on treatment at the respective cycle and AE25 contained the cumulative probability of death at that cycle.

A similar correction was applied to cells AV25 to AV264.