NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

CDF RAPID RECONSIDERATION

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

The following documents are made available to the consultees and commentators:

1. **Response to consultee, commentator and public comments on the Appraisal Consultation Document (ACD2)**

2. **Consultee and commentator comments on the Appraisal Consultation Document** from:
   - Bayer
     - Bayer response to ERG queries
     - Response to ERG queries addendum
   - British Liver Trust
   - NCRI-ACP-RCP-RCR
   - NHS England

   Please note notification of no comments was received from the Department of Health

3. **ERG critique** from the Decision Support Unit (DSU)

4. **Commercial Access Arrangement**

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

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Cancer Drugs Fund reconsideration of TA189

Sorafenib for treating advanced hepatocellular carcinoma

Response to consultee, commentator and public comments on the Appraisal Consultation Document (ACD2)

Definitions:

Consultees – Organisations that accept an invitation to participate in the appraisal including the companies, national professional organisations, national patient organisations, the Department of Health and the Welsh Government and relevant NHS organisations in England. Consultees can make a submission and participate in the consultation on the appraisal consultation document (ACD; if produced). All non-company consultees can nominate clinical experts and/or patient experts to verbally present their personal views to the Appraisal Committee. Company consultees can also nominate clinical experts. Representatives from NHS England and clinical commissioning groups invited to participate in the appraisal may also attend the Appraisal Committee as NHS commissioning experts. All consultees have the opportunity to consider an appeal against the final recommendations, or report any factual errors, within the final appraisal determination (FAD).

Clinical and patient experts and NHS commissioning experts – The Chair of the Appraisal Committee and the NICE project team select clinical experts and patient experts from nominations by consultees and commentators. They attend the Appraisal Committee meeting as individuals to answer questions to help clarify issues about the submitted evidence and to provide their views and experiences of the technology and/or condition. Before they attend the meeting, all experts must either submit a written statement (using a template) or indicate they agree with the submission made by their nominating organisation.

Commentators – Commentators can participate in the consultation on the ACD (if produced), but NICE does not ask them to make any submission for the appraisal. Non-company commentator organisations can nominate clinical experts and patient experts to verbally present their personal views to the Appraisal Committee. Commentator organisations representing relevant comparator technology companies can also nominate clinical experts. These organisations receive the FAD and have opportunity to report any factual errors. These organisations include comparator technology companies, Healthcare Improvement Scotland any relevant National Collaborating Centre (a group commissioned by NICE to develop clinical guidelines), other related research groups where appropriate (for example, the Medical Research Council and National Cancer Research Institute); other groups such as the NHS Confederation, the NHS Commercial Medicines Unit, the Scottish Medicines Consortium, the Medicines and Healthcare Products Regulatory Agency, the Department of Health, Social Services and Public Safety for Northern Ireland).

Public – Members of the public have the opportunity to comment on the ACD when it is posted on the Institute's web site 5 days after it is sent to consultees and commentators. These comments are usually presented to the appraisal committee in full, but NICE reserves the right to summarise and edit comments received during consultations, or not to publish them at all, where in the reasonable opinion of NICE, the comments are voluminous, publication would be unlawful or publication would be otherwise inappropriate.

Please note: Comments received in the course of consultations carried out by NICE are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the submissions that NICE has received, and are not endorsed by NICE, its officers or advisory committees.

Comments received from consultees

Consultee	Comment [sic]	Response
Bayer	Given the age of this sorafenib with patent expiry in 5 years, there is great experience and understanding amongst clinicians about its place in the treatment of HCC. We therefore do not think that further data collection through the cancer drugs fund would be the best use of resources. Given Bayer has over the lifetime of this medicine reduced the net price three times and it has the lowest price in Europe, there is little more discount that we can offer. Patients need this medicine which, despite its age, remains innovative as it is the only option for this group of patients. Additionally, this treatment forms the basis of the next set of therapies in second line treatment of HCC. Therefore this document, in addition to presenting new evidence to address the uncertainties raised in the ACD, incorporates a revised Patient Access Scheme (PAS). Under the previous PAS the cost of sorafenib per pack was [redacted]. The new commercial arrangement results in an additional discount of [redacted], resulting in a new cost per pack of [redacted], a [redacted] discount on the NHS list price. The new price would be available via the Commercial Medicines Unit framework agreement and will apply to all indications of sorafenib. All cost-effectiveness results presented in this response incorporate this new price. **Executive summary** Following the most recent appraisal committee meeting, the committee, in the ACD, highlighted three areas of uncertainty; duration of treatment, overall survival, and resource use. As such, the appraisal consultation document recently issued by the National Institute for Health and Care Excellence (NICE) recommends sorafenib for use within the Cancer Drugs Fund for the treatment of patients with advanced hepatocellular carcinoma (HCC) who have failed or are unsuitable for surgical or loco regional therapies. Whilst the company acknowledges that for new drugs there may be potential value in the prospective collection of real-world evidence to address particular uncertainties via the CDF, for sorafenib which has been available for over 10 years, more robust evidence sources already exist. These are highlighted in Table 4. One such source is GIDEON which has the following advantages:	Comments noted. The committee agreed that the most plausible ICER was approximately £54,000 per QALY gained for sorafenib compared with best supportive care, including the new Commercial Medicines Unit price. The committee was aware that the most plausible ICER was higher than ICERs previously accepted for technologies that had met the end-of-life criteria. The committee highlighted consultation comments that sorafenib was the only treatment option available for people with advanced hepatocellular carcinoma. New therapies for second-line treatment would likely need previous treatment with sorafenib, which would exclude patients newly diagnosed with hepatocellular carcinoma if sorafenib were not available in England. The committee also understood that new therapies were being developed for first-line treatment. It acknowledged that sorafenib was innovative, given that it is the only systemic treatment to have been granted a marketing authorisation for advanced hepatocellular carcinoma

	• GIDEON is the most robust data source in which to assess outstanding uncertainties highlighted by the Appraisal Committee, with a population of over 3,200 patients treated with sorafenib. Crucially, through this submission Bayer have matched a large cohort of GIDEON patients, based on patient characteristics to those enrolled in SHARP (n=895), allowing consideration of outcomes in patients with characteristics reflecting those enrolled in SHARP. • Treatment duration and overall survival from the matched GIDEON population allows uncertainty surrounding the relationship between length of treatment and effectiveness in clinical practice to be assessed. With the matched population providing a greater overall survival benefit than that seen in SHARP. For this reason use of matched treatment data from GIDEON and efficacy from SHARP can be seen as conservative. **New evidence** This response presents new evidence of the observed unrestricted mean duration of treatment and dosing intensity from a cohort of the GIDEON study matched via propensity scoring based on the characteristics of patients enrolled in the SHARP trial (in which evidence of overall survival has previously been considered by the Committee). The unrestricted mean duration of treatment in the matched GIDEON population is [redacted] months, with a 95% confidence interval ([redacted] months). This is substantially lower than the Appraisal Committee's preferred extrapolation of duration of treatment, the log-normal, which resulted in an estimated treatment duration of [redacted] months. In addition the response highlights reasons as to why the use of the statistical fit criteria published by Kass et al, used by the Committee to determine the selection of the log-normal extrapolation of treatment duration is not appropriate. **Updated cost-effectiveness results** The result of updating the cost-effectiveness analyses to reflect this new evidence or to validate an alternative extrapolation of duration of treatment both greatly reduces the size and increases the robustness of the ICER's previously considered by the Appraisal Committee: • When the duration and dose intensity from the matched GIDEON population are used in the economic model (instead of estimates obtained via extrapolation of the SHARP RCT), the resulting ICERs are £32,819 when the log normal extrapolation for OS is	in the last 10 years. However, it highlighted that the benefits not captured in the QALY would not substantially decrease the ICER for sorafenib compared with best supportive care as stipulated in NICE's guide to the methods of technology appraisal. After taking a vote, the committee concluded that it could not recommend sorafenib for routine commissioning in the NHS. Comments noted. The committee understood from the company that everyone in GIDEON stopped treatment so the company provided only an unrestricted mean and a Kaplan–Meier analysis (rather than a parametric model). But the committee highlighted that it would have preferred the company to also fit parametric curves to the data because of the differences in the GIDEON and SHARP populations, and the small number of events towards the end of the Kaplan–Meier curves of time to treatment discontinuation, which leads to uncertainty. The committee concluded that data from SHARP should be used to estimate duration of treatment, and the total cost of treatment.

	used and £54,929 when the Weibull is used. (Based on the Committees preferred assumptions other than for duration of treatment). • Alternatively if this new evidence is used to validate an alternative estimate of duration of treatment from SHARP, such as the Weibull or Gompertz, whilst accepting the Committee's guidance that the most plausible ICER lies below the midpoint of the log-normal and Weibull distribution, the plausible ICER is £48,599 per QALY gained. **Conclusion** Sorafenib is an innovative treatment, which upon launch resulted in a step change for the treatment of patients with advanced HCC. In over 10 years following its marketing authorisation there still remains no alternative for patients with advanced disease. Hepatocellular carcinoma is the second most common cause of cancer death worldwide, with future treatments for advanced HCC patients reliant on the continued availability of sorafenib as a first line treatment option. In this response the company present new evidence which addresses key uncertainties highlighted in the ACD and results in robust ICERs which fall within the range normally considered cost effective for medicines which fulfil end of life criteria. **1. Introduction** Following the most recent appraisal committee meeting, the committee, in the ACD, highlighted three areas of uncertainty; duration of treatment, overall survival, and resource use. This document outlines Bayer's response to each of these, and an overview of further analyses conducted by Bayer. **2. Area of uncertainty 1: duration of treatment** ***2.1 Background*** o Following the first Committee meeting in July 2016, the Appraisal Committee did not agree with Bayer's approach for estimating mean duration of treatment based on time to progression (TTP) from the SHARP trial, and recommended that an unrestricted mean be used to estimate treatment duration from the SHARP trial.	

	o In advance of the second Appraisal Committee meeting, Bayer implemented this change and presented an estimate of the unrestricted mean of treatment duration. Based on this extrapolation, using the SHARP trial, the analysis of the statistical fit indicated that the lognormal was statistically the most appropriate fit of the distributions considered. As such, lognormal was presented by the company for estimating duration of treatment. o The DSU, however, argued that based on the empirical data submitted by Bayer and visual inspection of the extrapolations, the Weibull and Gompertz distributions were most plausible. o Based on interpretation of statistical fit using criteria published by Kass et al 1995, the Appraisal Committee stated that the log-normal should be used to extrapolate treatment duration, and ruled out the other extrapolations. ***2.2 Actions/recommendation by AC to address areas of uncertainty following second appraisal*** o The Appraisal Committee expressed concern regarding estimates of treatment duration and a recommendation was made to collect data from the Systemic Anti-Cancer Therapy dataset whilst within the Cancer Drugs Fund to further address this uncertainty. ***2.3 Bayer's response and considerations on the appraisal consultation document*** o Previous estimates of mean treatment duration have been obtained via extrapolation of treatment data from SHARP. Not all patients in the SHARP study discontinued treatment, as a result SHARP can only provide an estimate of the treatment duration, and this varies greatly based on the parametric model selected. o As there was uncertainty around duration of treatment, Bayer felt that it would be informative to conduct further analyses on results from a cohort of the GIDEON study matched via propensity scoring to patients enrolled in SHARP as a means of further understanding the duration of treatment. The GIDEON study followed all patients until treatment discontinuation, resulting in an observed result as opposed to an estimate. o Furthermore, Bayer wishes to highlight supportive evidence, which when considered in addition to new data may validate the approach the company has	

taken to address this area of uncertainty.

- The analyses conducted and observations from the ACD are presented below.

2.4 Further analyses conducted

2.4.1 Overview of analysis

- Following guidance from the first Appraisal Committee meeting, GIDEON a long-term observational study reporting the overall survival of 3,213 advanced HCC patients treated with sorafenib, was used to validate the extrapolation of overall survival applied to the SHARP RCT. For this analysis, patients in the GIDEON study were matched based on the baseline characteristics of patients enrolled in the SHARP trial. The DSU subsequently reviewed these methods and expressed satisfaction with Bayer's analysis, which was noted by the Appraisal Committee.
- As the GIDEON matched population resulted in patients reflective of those enrolled in SHARP and were all followed until treatment discontinuation, Bayer conducted an analysis to obtain the unrestricted mean duration of treatment from the matched GIDEON population (n=895)
- Critically this provides an unrestricted mean duration of treatment in a patient population approximately three times of that considered in SHARP, which could reduce uncertainty around extrapolating from incomplete data obtained from SHARP.

2.4.2 Results

- Results using the unrestricted mean from the GIDEON matched population resulted in a mean duration of treatment of [redacted]). The median duration of treatment was [redacted] days ([redacted]months). The mean dose intensity from the matched population was [redacted], with a 95% confidence interval from [redacted].

Table 1: Duration of treatment (days) in matched GIDEON population

Matched GIDEON population (n=895)	Mean/ 95% CI	Median/ IQR
Duration of treatment	[redacted]	[redacted]

	[redacted]	[redacted]
Mean dose intensity*	[redacted]	[redacted]

* Values are based on the mean daily dose per patient

** No mean daily dose was available for 98 GIDEON patients matched to the SHARP cohort.

2.4.3 Interpretation

- This is the first analysis showing an unrestricted mean (without the need for extrapolation) from a population matched to those enrolled in the SHARP trial. Additionally, the availability of overall survival data for this population may address the Committee's preference for obtaining overall survival and treatment duration from a matched study population (i.e. balance on patient characteristics).
- The mean duration of treatment of [redacted] derived from the analysis is substantially lower than that obtained via the extrapolation of SHARP. The mean dose intensity is also lower at [redacted] than the 710.5mg used in the economic model.
- In the matched GIDEON analysis the mean duration of treatment at a 95% confidence interval are [redacted]. The log-normal extrapolation of the SHARP study favoured by the Appraisal Committee provides an estimate of [redacted]months, exceeding the upper confidence interval of the GIDEON analysis by over 2.5 months.
- This provides robust evidence to support the DSU view that the log-normal may not provide the most plausible extrapolation
- It should also be noted that the mean derived from this analysis lies closer to the estimates from the extrapolations favoured by the DSU, which resulted in an estimated mean duration of treatment of [redacted].
- Findings from the matched analysis suggest that treatment duration is lower than estimates obtained via extrapolation of treatment duration from the SHARP trial. Considering a scenario where the Appraisal Committees preferred assumptions are updated to include a mean duration of treatment from the matched analysis, the resultant ICER would be between £32,819 and £54,929 using the log-normal and Weibull extrapolations of OS respectively (full results are presented in Section 6)
- Overall survival of patients in the matched GIDEON sample exceeded that observed in SHARP. It is significant that a lower duration of treatment (than that estimated from

Comments noted. The committee understood from the company that everyone in GIDEON stopped treatment so the company provided only an unrestricted mean and a Kaplan–Meier analysis (rather than a parametric model). But the committee highlighted that it would have preferred the company to also fit parametric curves to the data because of the differences in the GIDEON and SHARP populations, and the small number of events towards the end of the Kaplan–Meier curves of time to treatment discontinuation, which leads to uncertainty.

The committee heard from NHS England that patients now have treatment for a shorter period of time than was standard in 2007, trading a sizeable decrease in adverse events for a small drop in effectiveness. But taking all the

	SHARP) in combination with a lower dose intensity ([redacted] vs 710.5mg) did not impact health outcomes. Therefore use of the SHARP overall survival data in combination with duration of treatment derived from the GIDEON matched cohort is conservative. ***2.5 Observations on treatment duration from ACD and Bayer's response*** o In response to the first ACD Bayer presented results based on AIC/BIC criteria to inform model selection for extrapolating duration of treatment and presented analysis using a log-normal extrapolation. o Based on both visual inspection and consideration of the external data, the Decision Support Unit (DSU) concluded that the Weibull and Gompertz distributions appeared to provide a more plausible explanation of treatment duration than the log-normal. Both of these extrapolations resulted in duration of treatment of [redacted]. o In the Appraisal Committee meeting guidance from the DSU was not followed on the basis that using criteria outlined by Kass et al (1995) the Bayesian information criterion (BIC) statistics strongly indicated that the Weibull did not fit the data. On this basis the log-normal extrapolation was selected resulting in an estimate mean duration of treatment of [redacted]months. o Having considered the DSU response and guidance in the relevant technical support document Bayer concurs with the DSU that the Weibull and log-normal provide a more plausible extrapolation of the data. ***2.6 AC justification for extrapolation:*** o The decision criteria cited by Kass et al 1995 provides an interpretation of the difference in Bayesian information criterion (BIC) in terms of model specification. o Table 2 presents the results of AIC/BIC tests conducted by the company when evaluating approaches to extrapolation of duration of treatment from the SHARP trial. The difference in BIC score of 10.4 between the log-normal and Weibull was used to determine that log-normal was the conclusive distribution for the extrapolation of duration of treatment. o Bayer would like to highlight potential uncertainties with the use of the Kass et al 1995 criteria for the basis of selecting the most appropriate model fit. These are discussed below following an overview of the criteria used for selection. **Table 2: SHARP OS model extrapolations: AIC/BIC statistics**	observational evidence into account, the committee noted it had concerns about the generalisability of these results to the SHARP randomised controlled trial. See section 4.26 of the FAD. The committee concluded that data from SHARP should be used to estimate duration of treatment, and the total cost of treatment.

	Exponential	Weibull	Loglogistic	Gompertz	Lognormal
AIC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BIC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

1. **<u>Statistical fit should not be used in isolation to decide on model fit alone</u>**

   - The selection of the log-normal was based heavily on the use of criteria outlined by Kass et al (1995) this assessment does not allow for the consideration of evidence from visual inspection, or the availability of external data from the literature to contribute to curve selection.
   - In the ACD guidance it is suggests that choice of extrapolations should not be based on statistical interpretation alone. In using statistical criteria alone this approach fails to acknowledge that the uncertainty lies in the extrapolation of the Kaplan Meir data to which Figure 1 shows that the log-normal overestimates.
   - The DSU have previously noted the following with regard to extrapolation of treatment duration in this appraisal:
     - AIC/BIC criteria should not be used in isolation to determine model fit
     - Differences in AIC/BIC for extrapolation of duration of treatment from SHARP are relatively small
     - Upon visual inspection (of the model fits) the Weibull and Gompertz both appear to match the latter part of the KM curve (where there is uncertainty) better than the log-normal.

2. **<u>Methods outlined by Kass et al (1995) may not be appropriate for assessing the fit of alternative parametric models</u>**

   - The NICE Technical Support Document (TSD) 14 provides guidance on methods for undertaking survival analysis for use in an economic evaluation when patient-level data are available.
   - This TSD provides no guidance in terms of inferring from numerical differences in AIC/BIC statistics to inform selection of a parametric model. However, it does advise that the interpretation outlined by the Appraisal Committee to be the grounds for the selection of the log-normal distribution, may not be an appropriate

Comments noted. The committee stated that the log normal distribution was the best statistical fit of the 5 distributions explored by the company. The committee noted that based on the Kass et al. criteria, the Bayesian information criterion statistics strongly indicated that the lognormal distribution was a better fit to the observed data than the Weibull. The committee also heard from the clinical expert that approximately 10% of patients are still having sorafenib treatment at 3 years, which supported using the log normal distribution. The committee concluded that the company's fully parametric method using the log normal distribution was the most robust estimate of treatment duration. See section 4.28 of the FAD.

Please be aware that whilst the DSU is funded by NICE, these documents do

	criteria for the selection of parametric models. It states that in the past this has been used erroneously, stipulating that *"measures such as the negative 2 log likelihood are only suitable for comparing nested models, whereby one model is nested within another (for example, one model adds an additional covariate compared to another model). Different parametric models which use different probability distributions cannot be nested within one another. Thus the negative 2 log likelihood test is not suitable for assessing the fit of alternative parametric models"* **3. <u>Implications for using criteria published by Kass et al for other areas of uncertainty</u>** o In the circumstance that this test is accepted by the Committee as a definitive method for the selection of parametric models for extrapolation of health outcomes, consideration should be given to the results of applying this interpretation to overall survival fits conducted on the matured overall survival cohort data from the matched GIDEON population. These analyses were conducted to inform the extrapolations previously applied to the SHARP trial (further consideration of this is presented in Section 3) *Interpretation:* o Although the Appraisal Committee concluded that the log-normal was the superior fit, and Bayer had previously presented results using the log-normal, it was clear that there still existed some uncertainty from the DSU regarding the most appropriate model for extrapolation. o If an alternative model was in fact the most appropriate fit or use of these selection criteria was used to address uncertainty in regard to most plausible extrapolation of overall survival, the resultant ICER are likely to be within the cost-effectiveness range for end of life treatments. Results from these analyses are presented in Section 6. ***Supportive evidence*** In light of new evidence provided in the sections above, and the Appraisal Committee's recommendation that duration of treatment data collected from SACT would be helpful in addressing this uncertainty, the company believes that real-world studies reporting on the	not constitute formal NICE guidance or policy. Comments noted. Consideration of the Kass criteria was applied to the selection of parametric models for treatment duration and overall survival. See sections 4.25 and 4.28 of the FAD.

duration of treatment should be used in determining the most plausible source or extrapolation of treatment duration for the economic model.

- Over the past three years over [redacted]_HCC patients[1] have had access to treatment in England via the CDF. Unlike for new treatments, there is therefore published evidence on the duration of treatment derived from use on the NHS.
- The duration and therefore cost of treatment are reflective of those faced by the NHS and as such should be considered at the very minimum to inform an appropriate extrapolation. Extrapolations conducted on the SHARP data with resulting estimates of mean duration of treatment are presented in Figure 1 and Table 3 respectively.

Figure 1: Extrapolation of treatment duration from SHARP

Table 3: Estimated mean duration of treatment obtained via extrapolation

	Exponential	Weibull	Loglogistic	Gompertz	Lognormal
Mean DoT (months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[1] Patient numbers taken from July 2013 - June 2016

- Whilst many studies cannot provide a mean duration of treatment due to length of follow-up, it is possible to compare medians across studies and note the relationship where both a median and unrestricted mean is reported to provide further evidence that the extrapolation of [redacted] months to be unrealistically high.
- Table 4 presents findings from the literature showing UK sources. The median duration of treatment in all sources is lower than the matched GIDEON population. Although it is impossible to be certain, it is likely that this indicates that treatment duration in the UK is less than the unrestricted mean reported in the matched GIDEON population.

Table 4: Empirical estimates of treatment duration

Source	Sample size (n=)	Duration of treatment (months)	
		Median	Mean
J King et al (2013)	379	3.2	NR
GIDEON (total population)	3,202	3.46	5.52
GIDEON (matched population)	895	[redacted]	[redacted]
J King et al (2016)	484	3.6	NR
Ziogas et al (2017)	Age≤ 75: 151 Age>75: 31	Age ≤75: 3.0 95% CI (2.5–3.9) Age >75: 5.1 95% CI (3.1–7.1)	NR

NR: Not reported

2.7 Conclusion

- New evidence considering the unrestricted mean duration of treatment from a cohort of the GIDEON study population matched based on characteristics of those enrolled in SHARP (n=895) provides for the first time a true unrestricted mean duration of treatment (as opposed to extrapolated estimates from SHARP)

	• The mean unrestricted duration of treatment in the matched GIDEON population is [redacted] months, with a 95% confidence interval from [redacted] months to [redacted] months. This provides strong evidence that the log-normal extrapolation applied to SHARP used to derive a treatment duration estimate of [redacted] months is unrealistically high. Further to this the company highlight uncertainties in the criteria published by Kass et al used in which to justify the selection of the log-normal extrapolation, identifying if correct in its use, how this could be applied to other uncertainties such as overall survival • This new analysis reports a median and unrestricted mean that is aligned with previous empirical evidence sources presented to the Committee. Incorporation of this treatment duration in the economic model, or using these results to validate either the Weibull or Gompertz extrapolations of treatment duration lead to ICERs likely to fall into the range of cost-effective estimates which are normally in the acceptable range for end of life medicines. **Area of uncertainty 2: extrapolation of overall survival** ***Background*** o In the first Appraisal Committee meeting, data from GIDEON, a large international prospective study collecting overall survival data from 3,213 sorafenib patients, was presented to validate the parametric model used in determining the long-term extrapolation of overall survival from SHARP. o The analysis presented was deemed by the Appraisal Committee to validate the log-normal extrapolation of the data in SHARP; however, there was uncertainty in the comparability of the GIDEON study population to patients enrolled in SHARP. o In the second Appraisal Committee, Bayer met the requests of the Committee in presenting the population of GIDEON matched via propensity scoring to patients enrolled in the SHARP trial. The DSU expressed satisfaction with the matching performed by the company and this subsequently provided overall survival data from a sample of 895 patients o The Committee noted that in general the log-normal function used by the company to extrapolate survival beyond SHARP fitted the matched GIDEON data better than the Weibull function.	Comments noted. The committee appreciated that clinical experience with sorafenib had improved over time and adverse events may now be managed better, partly by shorter duration of treatment. The committee heard from NHS England that patients now have treatment for a shorter period of time than was standard in 2007, trading a sizeable decrease in adverse events for a small drop in effectiveness. But taking all the observational evidence into account, the committee noted it had concerns about the generalisability of these results to the SHARP randomised controlled trial. See section 4.26 of the FAD. The committee understood that the ERG considered that the estimates of mean and median treatment duration reported by the Cancer Drugs Fund, King et al., GIDEON and Palmer et al. were inconclusive and therefore did not support the company's claim that SHARP overestimated the treatment duration of sorafenib in clinical practice.

	Actions/recommendation by AC to address areas of uncertainty following second appraisal o The Committee considered that uncertainty remained around the most appropriate model to extrapolate overall survival, with the Weibull distribution remaining plausible. The Committee agreed that the most plausible ICER for sorafenib was lower than the midpoint of the log-normal and Weibull distributions. o In light of the uncertainty, a recommendation was made to collect data from the Systemic Anti-Cancer Therapy data set to help resolve some of this uncertainty. ***Bayer's response and considerations on the appraisal consultation document*** o In this section, Bayer wishes to comment on issues identified in the appraisal consultation document, including the consistency in the use of criteria published by Kass et al in determining model selection and the availability of follow-up overall survival data from the SHARP trial, which was highlighted by the Appraisal Committee as a potential for addressing uncertainty in the ACD. ***Consistency in use of statistical criteria in determining model (Kass et al 1995)*** o The justification for the choice of treatment duration extrapolation by the Appraisal Committee was based on criteria for the interpretation of the numerical differences in BIC scores published by Kass et al (1995). o As this approach was used by the Appraisal Committee to select and recommend a preferred method for extrapolation of duration of treatment, it may be reasonable to argue that this approach should also be considered for the selection of an appropriate distribution for the extrapolation of overall survival. o If this approach was applied, the use of the log-normal extrapolation could be validated for extrapolation of overall survival, as interpreted by the difference in BIC scores. o Upon application of the Kass et al decision criteria to the extrapolation of overall survival, the analysis conducted on the matched GIDEON population to assess the correct model to fit to the SHARP data resulted in a difference in the BIC scores between the log-normal and Weibull of 20.68 points. To use the interpretation provided by the Appraisal Committee this would strongly indicate that the Weibull does not fit the data (to a greater extent than the 10.4 point difference observed in the duration of treatment analysis) o Results of the AIC/BIC analyses conducted on the log-normal and Weibull overall	Comments noted. The company fitted log normal and Weibull curves to the Kaplan–Meier data for the matched GIDEON population and stated that the log normal curve provided a better statistical fit to the observed data than the Weibull curve; the committee agreed this based on standard statistical criteria using the Bayesian information criterion described in Kass et al. (1995). The committee considered that beyond about 600 days, the Weibull curve fitted the data better than the log normal curve. However, the committee was aware that the uncertainty was greater in the tail of the curve where limited or no data existed. The committee understood from the ERG that the log normal function would overestimate overall survival whereas the Weibull function would

survival extrapolations are presented in Table 5.

Table 5: Results of AIC/BIC criteria in assessing the OS extrapolation from SHARP

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]
Difference	**20.67**	**3.39**	**24.06**	**20.68**	**3.38**	**24.06**

Interpretation

- Should the criteria outlined by Kass et al be used in the assessment of extrapolation for duration of treatment to conclusively favour the log-normal distribution, this same criteria should be used to evaluate the extrapolation of overall survival to conclusively favour the log-normal distribution
- When considering the population from the GIDEON study that are matched based on patient characteristics to those enrolled in SHARP is considered, a difference of 20.7 points can be seen between the best fitting extrapolation (log-normal) and the Weibull distribution. This difference is approximately twice that seen when results are inferred in a similar way for the extrapolation of treatment duration.

Request by Appraisal Committee for follow-up overall survival data from the SHARP trial

- It was noted by the Appraisal Committee that further follow-up survival data from SHARP, as used in the economic model, could clarify the uncertainty surrounding overall survival.
- Bayer would like to note that in the extrapolation of overall survival from SHARP used to inform the economic model, all available survival data from the SHARP trial has been used. Unfortunately, as raised in the original submission process, the SHARP trial did not follow-up overall survival past 19 months.

underestimate it. Therefore, the ERG advised that both curves should be considered when extrapolating overall survival, and to estimate the ICER for sorafenib compared with best supportive care. The committee acknowledged that it would not use statistical goodness of fit alone to choose the most appropriate survival function. It noted that in general the log normal function used by the company to extrapolate survival beyond SHARP fitted GIDEON better than the Weibull function, but that the Weibull function was still plausible. The committee was also aware that the 3 data sets the company had presented (SHARP, GIDEON, and Palmer et al.) for informing the choice of survival distribution did not conclusively favour 1 single distribution. For example, the Bayesian information criterion statistics provided evidence that the log normal function fitted the data better than the Weibull function in the SHARP analysis based on Kass et al., but this was not considered a statistically strong result and therefore the committee considered that the Weibull function remained plausible. The committee reiterated that SHARP was among the most robust source of evidence it had seen for sorafenib during the Cancer Drugs Fund reconsideration committee meetings. Therefore, the committee concluded, as it had done at its second meeting, that

	Area of uncertainty 3: Resource use o In order to reduce uncertainty resource use data for patients not treated with sorafenib would be required. No new evidence sources have been identified that offer comparative data, as patients with advanced HCC no longer receive best supportive care if suitable for sorafenib. o The company proposes that the Appraisal Committee's preferred assumption, to pool results from the original and update resource use surveys to address uncertainty. **Company response to recommendation to enter the Cancer Drugs Fund** • Bayer acknowledges that for treatments that are new to market, there may be value in the prospective collection of real-world evidence to address certain uncertainties. However sorafenib has been used to treat advanced HCC on the CDF for 6 years, and obtained marketing authorisation over 10 years ago. In addition sorafenib is currently recommended for routine use in Scotland and Wales. For this reason data which would routinely not be available for new treatments is available and published for sorafenib. • The GIDEON study is a large multicentre prospective study considering both mean duration of treatment and overall survival in over 3,000 patients treated with sorafenib. Through this appraisal Bayer has matched a large cohort of patients, based on patient characteristics to those enrolled in SHARP (n=895). This type of analysis would not be possible in SACT and as such there would be no method to link observations back to the SHARP population. • Additional publications of real-world evidence of the use of sorafenib are also available providing evidence of use within the NHS. **Results: Exploratory analyses conducted based on new evidence and interpretation of the AC response** • The appraisal consultation document concludes that the true estimate of overall survival with sorafenib was likely to lie between the estimates from the log-normal and Weibull distributions with these two extrapolations informing the range of plausible ICERs (£49,500 - £87,000). • The Committee agreed that the most plausible ICER was likely to be lower than the	the true estimate of life expectancy with sorafenib compared with best supportive care was likely to lie between the estimates from the log normal and the Weibull distributions, but agreed it was closer to the log normal estimates than the Weibull estimates. See section 4.25 of the FAD. Comments noted. At its third meeting, the committee noted that the company had chosen not to submit a proposal for sorafenib to be included in the Cancer Drugs Fund because it considered the GIDEON data were better than those the Cancer Drugs Fund could collect. The committee concluded that sorafenib could not be recommended for use within the Cancer Drugs Fund. See section 4.34 of the FAD.

midpoint of the preferred ICER range (£68,250) but higher than technologies previously accepted that had met the end of life criteria. That is that the most plausible ICER is likely to be closer to the log-normal extrapolation, a conclusion that has been reached by both the DSU and Appraisal Committee.

- Using the above guidance from the Appraisal Committee, in addition to new evidence submitted to address uncertainty in the duration of treatment from the matched GIDEON population, new cost-effectiveness analyses have been conducted which all reflect the Appraisal Committees preferred assumptions of:
    - Independent assessment of progression
    - Wastage (7 days)
    - Pooled resource use estimates
    - Treatment duration using an unrestricted mean

6.1: Cost-effectiveness results using duration of treatment obtained from the matched GIDEON sample and the Appraisal Committee's preferred assumptions

- The ICERs using both the log-normal and Weibull extrapolations for overall survival and calculation of the midpoint*,in addition to duration of treatment derived from the matched GIDEON sample are presented in Table 6.
- The results detail the ICER using the mean dose/patient/day from the GIDEON study ([redacted]) and that reported in the SHARP study (710.5mg).

Table 6: Updated PAS price: Cost-effectiveness results using treatment obtained from the matched GIDEON sample

		Duration of treatment: Matched GIDEON population (n=895) ([redacted])	
		GIDEON matched mean dose ([redacted])	SHARP mean dose (710.5mg)
Overall	Log-normal	£32,819	£36,050

survival	Weibull	£54,929	£61,290
	Midpoint* (log-normal and Weibull ICERs)	**£43,874**	**£48,670**

***** The midpoint between log-normal extrapolation and the Weibull, represents the upper bound of the Appraisal Committee's plausible ICER range

6.2 Cost-effectiveness results assuming plausible ICER is lower than the mid-point of the log-normal and Weibull and the log-normal extrapolation.

- The Committee concluded in the ACD that the most plausible ICER was likely to be lower than the mid-point of the log-normal and Weibull ICERs (calculated in Table 6)
- Using the calculated midpoint as an upper bound (Table 6) and;
- Using the log-normal extrapolation as a lower bound (also presented in Table 6)
- A midpoint of this range which reflects a midpoint of the plausible ICER range is presented in Table 7.
- This analysis uses all Appraisal Committee's preferred assumptions (except the derivation of treatment duration based on log-normal extrapolation from SHARP)

Table 7: Updated PAS price: Cost-effectiveness results following Appraisal Committee guidance that the plausible ICER is likely lower than the mid-point of the log normal and Weibull and the log-normal extrapolation

		Duration of treatment: Matched GIDEON population (n=895) ([redacted])	
		GIDEON matched mean dose ([redacted])	SHARP mean dose (710.5mg)
Overall survival	Lognormal	£32,819	£36,050
	Midpoint (log-normal and Weibull ICERs)	£43,874	£48,670

	ICER	**£38,347**	**£42,360**

6.3 Cost-effectiveness results presenting variation of the mean duration of treatment between that observed in the matched GIDEON sample and the estimates of duration of treatment derived from the Weibull and Gompertz extrapolations from the SHARP trial as favoured by the DSU

- Acknowledging that uncertainty exists around the duration of treatment, Bayer have varied the mean duration of treatment between that observed in the matched GIDEON sample and the estimates of duration of treatment derived from the Weibull and Gompertz extrapolations from the SHARP trial, resulting in a mean duration of treatment of [redacted]. In this analysis both the dosing intensity from the GIDEON sample and the SHARP trial are presented. Results of these analyses are presented in Table 8.

Table 8: Updated PAS price: Cost-effectiveness results varying source of duration of treatment (matched GIDEON population/SHARP)

		Duration of treatment			
		Matched GIDEON population (n=895) ([redacted])		Gompertz/Weibull extrapolation from SHARP: ([redacted])	
		GIDEON matched mean dose ([redacted])	SHARP mean dose 710.5mg	GIDEON matched mean dose ([redacted])	SHARP mean dose 710.5mg
Overall survival	Lognormal	£32,819	£36,050	£37,202	£41,073
	Midpoint (log-normal and Weibull ICERs)	£43,874	£48,670	£50,380	£56,125

	ICER	**£38,347**	**£42,360**	**£43,791**	**£48,599**

Conclusion: Results

- In this section Bayer presented cost-effectiveness analyses varying two of the key areas of uncertainty, the mean duration of treatment and extrapolation of overall survival. In each of the analyses Bayer have reflected the Appraisal Committee's preferred assumptions unless stated.
- As documented in the ACD the most plausible ICER lies between the midpoint of the log-normal/ Weibull ICER as an upper bound, and the log-normal exploration as a lower bound (equivalent to a 25%:75% weighting)

Use of matched GIDEON treatment duration

- Analyses presented show that when the observed mean duration of treatment is used from the matched GIDEON sample, the ICER does not rise above £50,000 per QALY when the extrapolation of overall survival is varied from the log-normal (£32,819 per QALY) to the midpoint of the log-normal and Weibull ICER (£43,873 per QALY). This provides a plausible ICER of £38,347 per QALY and considers the impact of the Weibull distribution in line with the Committee's preferred assumptions.

Validation of the extrapolation of the estimation of treatment duration applied to SHARP

- A final analysis considered the cost-effectiveness of sorafenib when duration of treatment analysis conducted on the matched GIDEON population was used to validate extrapolated estimates from SHARP (using either the Weibull or the Gompertz extrapolation following the preference of the DSU).
    - The midpoint of the plausible cost-effectiveness range outlined by the Appraisal Committee was used to reach a plausible ICER of £48,599.

- Over the course of the appraisal the company has sought to incorporate a range of preferred assumptions into cost-effectiveness estimates. In the results presented in this section the company explored duration of treatment, providing two methods that meet the preferred assumptions. Both show sorafenib to be cost-effective. Despite

	evidence favouring the log-normal as a preference for overall survival extrapolation, the company has sought to incorporate the Committee's preferred assumptions based on guidance provided in the ACD. **Overall conclusion** • GIDEON is the most robust data source in which to assess outstanding uncertainties highlighted by the Appraisal Committee, with a population of over 3,200 patients treated with sorafenib. Crucially, through this submission Bayer have matched a large cohort of GIDEON patients, based on patient characteristics to those enrolled in SHARP (n=895). This allows considerations of outcomes in patients with characteristics reflecting those enrolled in SHARP. • New evidence presented in this response finds the unrestricted mean duration of treatment in the GIDEON population to be [redacted], with a 95% confidence interval ([redacted] months) suggesting that the estimates from the SHARP study to be a gross overestimate in treatment duration. • Matched dosing intensity and overall survival from GIDEON allows uncertainty surrounding the relationship between length of treatment and effectiveness in clinical practice to be assessed. With the matched population providing a greater overall survival benefit than that seen in SHARP, using matched treatment data from GIDEON and efficacy from SHARP could be seen as conservative. • Cost-effectiveness results presented in Section 6, reflect the Appraisal Committee's preferred assumptions (other than from the updated evidence on duration of treatment) and show the range of ICER to vary between £38,346 and £48,599 per QALY. Each of these ICERs reflects the potential plausibility of the Weibull in line with guidance in the appraisal consultation document. Given all evidence considered to date it is with confidence that these ICERs are likely to be in the range normally considered cost effective for medicines which fulfil end of life criteria • Sorafenib is an innovative treatment, which upon launch resulted in a step change for the treatment of patients with advanced HCC. In over 10 years following its marketing authorisation there still remains no alternative for patients with advanced disease.	

	Hepatocellular carcinoma is the second most common cause of cancer death worldwide, with future treatments for advanced HCC patients reliant on the continued availability of sorafenib as a first line treatment option.	
British Liver Trust	As per our previous response the British Liver Trust would want patients with advanced stage hepatocellular carcinoma who have failed or are unsuitable for surgical or loco-regional treatment to have access to treatment with Sorafenib. Sorafenib is the only treatment available for this group of patients, the alternative being specialist palliative care. We highlight, again, the immense benefits of not only prolonging life but also the improved symptom control and quality of life that can be achieved. As Sorafenib is available to patients in Scotland and Wales it would be unfair not to give equal access to patients in England.	Comments noted. The committee agreed that the most plausible ICER was approximately £54,000 per QALY gained for sorafenib compared with best supportive care, including the new Commercial Medicines Unit price. The committee was aware that the most plausible ICER was higher than ICERs previously accepted for technologies that had met the end-of-life criteria. The committee highlighted consultation comments that sorafenib was the only treatment option available for people with advanced hepatocellular carcinoma. New therapies for second-line treatment would likely need previous treatment with sorafenib, which would exclude patients newly diagnosed with hepatocellular carcinoma if sorafenib were not available in England. The committee also understood that new therapies were being developed for first-line treatment. It acknowledged that sorafenib was innovative, given that it is the only systemic treatment to have been granted a marketing authorisation for advanced hepatocellular carcinoma in the last 10 years. However, it highlighted that the benefits not captured in the QALY would not substantially

		decrease the ICER for sorafenib compared with best supportive care as stipulated in NICE's guide to the methods of technology appraisal. After taking a vote, the committee concluded that it could not recommend sorafenib for routine commissioning in the NHS. See section 4.33 of the FAD.
Department of Health	No comments.	Comments noted. No action required.
Royal College of Physicians	**Has all of the relevant evidence been taken into account?** • The UK sorafenib audit including 448 patients has now been published in full and is attached. This provides good multisite data across the UK regarding mean daily dose and median time on treatment for the UK population. Additionally, it provides data on survival outcomes along with prognostic factors associated with survival for sorafenib treated patients. The paper provides support for the efficacy of sorafenib in patients with good PS and preserved liver function. Of note this study was academically driven with no commercial support. • The Gideon study is referred to in the appraisal document but there are multiple publications and it is not clear which data were considered. The most recent publication is attached from Dec 2016. This is global study sponsored by Bayer and we do not believe that the UK contributed to this. Hence the data from the UK audit may be more relevant. That said, there are some consistent findings with regard to poorer outcomes for Child B patients >B7. The daily dose and duration are also broadly similar. **Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?** • Our experts felt they were not qualified to judge this and not clear how this was calculated.	Comments noted. At its first and second meetings, the committee agreed that the effectiveness and costs should ideally come from the same study. The committee noted that in King et al. people with Child-Pugh grade A liver function did not live as long as people in SHARP (9.5 months compared with 10.7 months). The committee considered that this may have been partly explained by the reduction in treatment duration (3.6 months in King et al. compared with 5.3 months in SHARP) and daily dose (590 mg in King et al. compared with 711 mg in SHARP) between the studies. The committee was also aware that people with Child-Pugh grade A liver function in GIDEON had a median overall survival of 13.6 months and a median treatment duration of 4.1 months, which the committee stated seemed counterintuitive when compared with King et al. and SHARP. The committee appreciated that clinical experience with

	Are the recommendations sound and a suitable basis for guidance to the NHS? • We support the recommendation to make sorafenib available within the CDF and the proposal for prospective data collection to clarify the uncertainties that remain regarding efficacy, time on treatment and resource usage. • There are a number of additional benefits from sorafenib being available which include the placement of clinical trials in the UK which provide access to new drugs. During the past few years the UK has gained an excellent reputation for recruitment into HCC trials from which some patients have gained outstanding benefits. In the absence of sorafenib being available, the UK would risks becoming an international outlier and not considered and appropriate environment to conduct clinical research in HCC. **Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of race, gender, disability, religion or belief, sexual orientation, age, gender reassignment, pregnancy and maternity?** • No.	sorafenib had improved over time and adverse events may now be managed better, partly by shorter duration of treatment. The committee heard from NHS England that patients now have treatment for a shorter period of time than was standard in 2007, trading a sizeable decrease in adverse events for a small drop in effectiveness. But taking all the observational evidence into account, the committee noted it had concerns about the generalisability of these results to the SHARP randomised controlled trial. The committee concluded that data from SHARP should be used to estimate duration of treatment, and the total cost of treatment. See section 4.26 of the FAD. The committee agreed that the most plausible ICER was approximately £54,000 per QALY gained for sorafenib compared with best supportive care, including the new Commercial Medicines Unit price. The committee was aware that the most plausible ICER was higher than ICERs previously accepted for technologies that had met the end-of-life criteria. See section 4.33 of the FAD.
NHS England	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 this disease) of 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). These benefits came at the expense of significant but tolerable	

	toxicity. The trial was stopped after the interim analysis showed this survival advantage and hence longer term information on survival in the trial patients is not known. 2. The SHARP population was meant to enrol only Child-Pugh A patients, the eventual entry reflecting 97% Child-Pugh (C-P) A patients and 3% C-P B patients. The trial consisted of 54% performance status (PS) 0 patients, 38% PS 1 and 8% PS 2. The median duration of treatment was 5.3 mo. All patients started at 800mg of sorafenib/day and the mean daily dose was 711mg. 3. There have been at least 4 randomised trials with sorafenib as a treatment arm and the median survival durations in these trials for patients receiving sorafenib have been between 8.5 and 10.2 months. 4. The UK audit of 448 patients published in Clinical Oncology 2016 had 77% known C-P A patients and 16% known C-P B patients. The median survival duration was 9.5 months for C-P A patients and 4.6 mo for C-P B. The audit comprised 26% PS 0 patients, 49% PS 1 and 21% PS 2. The mean daily dose was 590mg and 62% started at 800mg/day and 33% at 400mg/day. The median duration of treatment was 3.6 months. 5. The GIDEON audit of 3202 patients published in J Hepatology 2016 had 61% C-P A patients and 21% C-P B patients. The median survival durations were 13.6 months for C-P A patients and 5.2 mo for C-P B. The audit did not split PS 0 and 1 patients in this publication. The mean daily dose was not reported in the analysis but 72% started at 800mg/day. The median duration of treatment was 4.1 mo for C-P A and 2.3 mo for C-P B. Sorafenib had been discontinued at 8 weeks in 26% of C-P A patients and 42% of C-P B; at 28 weeks, 33% of C-P A patients and 20% of C-P B were still on treatment, respectively. 6. Bayer's matched GIDEON to SHARP analysis of 895 patients which was done to obtain longer term survival information had a median treatment duration of 3.9 mo with a mean daily dosage of 620mg and a mean treatment duration of 6 mo. 7. The SPC for the use of sorafenib in hepatocellular carcinoma is based on the results of the SHARP trial and states that 'there are limited data from this study in patients with	

	Child-Pugh B liver impairment'. There is therefore no robust randomised controlled trial data to demonstrate the survival benefit of sorafenib in C-P B patients	
	8. NHS England would normally wish the NICE Technology Appraisal Committee to use the same patient source of information on which to base its preferred estimates of both treatment duration and overall survival. The SHARP trial data provides evidence of the median duration of treatment with sorafenib (5.3 mo) in the same population of patients which provides the evidence of gain in median duration of survival with sorafenib over best supportive care. Separating the source of information of treatment duration from the source that provides the survival data usually increases uncertainty.	Comments noted. At its first and second meetings, the committee agreed that the effectiveness and costs should ideally come from the same study; this approach was supported by the ERG and by NHS England. See 4.26 of the FAD.
	9. The information from the 448 patient UK audit and the 3202 patient GIDEON audit suggests a lower median duration of treatment than in the SHARP trial (3.6 and 4.1 mo vs 5.3 mo) and the UK audit points to a lower mean daily dose of sorafenib (590mg vs 711mg) than the SHARP data. The other clear message from both the UK and GIDEON audits is that patients with C-P B do much worse than C-P A (A vs B is 9.5 vs 4.6 mo in the UK audit and 13.6 vs 5.2 mo in GIDEON).	
	10. The Bayer analysis of 895 patients matched from GIDEON to SHARP has a median treatment duration of 3.9 mo which is similar to the audit figures above in paragraph 9. The mean daily sorafenib dose in the matched analysis (620mg) is close to that in the UK audit (590mg), the difference being potentially explained by the different case mixes in the two populations.	
	11. The issue therefore is whether it is reasonable to retain the initial survival data from SHARP and its consequent modelling with GIDEON to SHARP analaysis but use the treatment duration and dosage information from the matched GIDEON to SHARP analysis. Oncologists have learned how to use sorafenib better than at the time at which the SHARP trial was performed, the trial having mandated the starting dose to be 800mg daily. As a consequence of this learning, sorafenib prescription in the clinic is often at a starting dose of less than 800mg daily (38% of patients in the UK audit and 28% in GIDEON as a whole) and the greater experience with continued sorafenib administration has led to earlier dose reductions than may have taken place in SHARP. There is thus a rationale for using the matched GIDEON to SHARP figure for mean daily dose of sorafenib. With this consideration also comes the issue of duration of	Comments noted. The committee appreciated that clinical experience with sorafenib had improved over time and adverse events may now be managed better, partly by shorter duration of treatment. The committee heard from NHS England that patients now have treatment for a shorter period of time than was standard in 2007, trading a sizeable decrease in adverse events for a small drop in effectiveness. But taking

	treatment and the same issues apply in terms of a rationale for using the median figure of 3.9 mo (and mean 6 mo figure) in the GIDEON to SHARP analysis. The advantage of using these GIDEON to SHARP figures is that they are more likely to represent what happens in the clinic currently; the disadvantage is that the greater duration of treatment and higher daily dose in SHARP may have translated into the degree of the survival gain with sorafenib, though if so, NHS England considers this effect is likely to be small.	all the observational evidence into account, the committee noted it had concerns about the generalisability of these results to the SHARP randomised controlled trial. See section 4.26 of the FAD.
	12. If the NICE TA Committee chooses to use the matched GIDEON to SHARP analysis to provide the treatment duration, sorafenib dosage and longer term survival in the economic model and these assumptions are key to the determination of cost effectiveness, then NHSE would state that any recommendation by NICE for the use of sorafenib should directly follow the inclusion criteria for patients in the SHARP trial: - histologically confirmed diagnosis of hepatocellular carcinoma - metastatic disease or advanced local disease that is ineligible for or failed surgical or locoregional therapies - no previous systemic therapy for hepatocellular carcinoma - Child-Pugh liver function class A - performance status of 0-2.	Comments noted. The committee heard from NHS England that Child-Pugh grade A liver function and an ECOG performance status of 0 to 2 would be the relevant population in UK practice. The committee concluded that this was the appropriate population for any recommendations for sorafenib. See section 4.21 of the FAD.
	13. If NICE recommends sorafenib for just Child-Pugh A patients, NHS England accepts that the present CDF inclusion of what amounts to C-P B7 patients would end. Given the survival figures in the UK and GIDEON audits for C-P B vs A patients (and the median survival duration in the whole GIDEON audit for B7 patients was 6.2 mo), NHS England regards such a conclusion of a NICE recommendation for only C-P A patients to be based on the best current data. This is because there were only 3% C-P B patients in SHARP (despite being excluded in the trial design) and there were very low survival durations for C-P B patients in the 2016 publications of these UK and GIDEON audits. In addition, consideration of the toxicity of sorafenib in patients with such a short life expectancy would also support this conclusion.	
	14. NHS England believes that sorafenib's patent in hepatocellular cancer will expire in 2020 or thereabouts.	Comments noted. The committee agreed that the most plausible ICER was approximately £54,000 per QALY gained

	15. 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, lenvatinib 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 competitive NHS England Specialised Commissioning prioritisation process. 16. 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; and if cost-effective, in what group of patients.	for sorafenib compared with best supportive care, including the new Commercial Medicines Unit price. The committee was aware that the most plausible ICER was higher than ICERs previously accepted for technologies that had met the end-of-life criteria. The committee highlighted consultation comments that sorafenib was the only treatment option available for people with advanced hepatocellular carcinoma. New therapies for second-line treatment would likely need previous treatment with sorafenib, which would exclude patients newly diagnosed with hepatocellular carcinoma if sorafenib were not available in England. The committee also understood that new therapies were being developed for first-line treatment. It acknowledged that sorafenib was innovative, given that it is the only systemic treatment to have been granted a marketing authorisation for advanced hepatocellular carcinoma in the last 10 years. However, it highlighted that the benefits not captured in the QALY would not substantially decrease the ICER for sorafenib compared with best supportive care as stipulated in NICE's guide to the methods of technology appraisal. After taking a vote, the committee concluded that it could not recommend sorafenib for routine commissioning in the NHS. See

		section 4.33 of the FAD.

Response to the Appraisal Consultation Document (ACD)
Sorafenib for treating advanced hepatocellular carcinoma
Updated Patient Access Scheme
Bayer plc, 27th January 2017

Updated: New Patient Access Scheme

Given the age of this sorafenib with patent expiry in 5 years, there is great experience and understanding amongst clinicians about its place in the treatment of HCC. We therefore do not think that further data collection through the cancer drugs fund would be the best use of resources. [redacted]. Patients need this medicine which, despite its age, remains innovative as it is the only option for this group of patients. Additionally, this treatment forms the basis of the next set of therapies in second line treatment of HCC. Therefore this document, in addition to presenting new evidence to address the uncertainties raised in the ACD, incorporates a revised Patient Access Scheme (PAS). Under the previous PAS the cost of sorafenib per pack was [redacted]. The new commercial arrangement results in an additional discount of [redacted], resulting in a new cost per pack of [redacted], a [redacted] discount on the NHS list price. The new price would be available via the Commercial Medicines Unit framework agreement and will apply to all indications of sorafenib. All cost-effectiveness results presented in this response incorporate this new price.

Executive summary

Following the most recent appraisal committee meeting, the committee, in the ACD, highlighted three areas of uncertainty; duration of treatment, overall survival, and resource use. As such, the appraisal consultation document recently issued by the National Institute for Health and Care Excellence (NICE) recommends sorafenib for use within the Cancer Drugs Fund for the treatment of patients with advanced hepatocellular carcinoma (HCC) who have failed or are unsuitable for surgical or loco regional therapies.

Whilst the company acknowledges that for new drugs there may be potential value in the prospective collection of real-world evidence to address particular uncertainties via the CDF, for sorafenib which has been available for over 10 years, more robust evidence sources already exist. These are highlighted in Table 4. One such source is GIDEON which has the following advantages:

- GIDEON is the most robust data source in which to assess outstanding uncertainties highlighted by the Appraisal Committee, with a population of over 3,200 patients treated with sorafenib. Crucially, through this submission Bayer have matched a large cohort of GIDEON patients, based on patient characteristics to those enrolled in SHARP (n=895),

allowing consideration of outcomes in patients with characteristics reflecting those enrolled in SHARP.

- Treatment duration and overall survival from the matched GIDEON population allows uncertainty surrounding the relationship between length of treatment and effectiveness in clinical practice to be assessed. With the matched population providing a greater overall survival benefit than that seen in SHARP. For this reason use of matched treatment data from GIDEON and efficacy from SHARP can be seen as conservative.

New evidence

This response presents new evidence of the observed unrestricted mean duration of treatment and dosing intensity from a cohort of the GIDEON study matched via propensity scoring based on the characteristics of patients enrolled in the SHARP trial (in which evidence of overall survival has previously been considered by the Committee).

The unrestricted mean duration of treatment in the matched GIDEON population is [redacted] months, with a 95% confidence interval ([redacted] months). This is substantially lower than the Appraisal Committee's preferred extrapolation of duration of treatment, the log-normal, which resulted in an estimated treatment duration of [redacted] months.

In addition the response highlights reasons as to why the use of the statistical fit criteria published by Kass et al, used by the Committee to determine the selection of the log-normal extrapolation of treatment duration is not appropriate.

Updated cost-effectiveness results

The result of updating the cost-effectiveness analyses to reflect this new evidence or to validate an alternative extrapolation of duration of treatment both greatly reduces the size and increases the robustness of the ICER's previously considered by the Appraisal Committee:

- When the duration and dose intensity from the matched GIDEON population are used in the economic model (instead of estimates obtained via extrapolation of the SHARP RCT), the resulting ICERs are £32,819 when the log normal extrapolation for OS is used and £54,929 when the Weibull is used. (Based on the Committees preferred assumptions other than for duration of treatment).
- Alternatively if this new evidence is used to validate an alternative estimate of duration of treatment from SHARP, such as the Weibull or Gompertz, whilst accepting the Committee's guidance that the most plausible ICER lies below the midpoint of the log-normal and Weibull distribution, the plausible ICER is £48,599 per QALY gained.

Conclusion

Sorafenib is an innovative treatment, which upon launch resulted in a step change for the treatment of patients with advanced HCC. In over 10 years following its marketing authorisation

there still remains no alternative for patients with advanced disease. Hepatocellular carcinoma is the second most common cause of cancer death worldwide, with future treatments for advanced HCC patients reliant on the continued availability of sorafenib as a first line treatment option.

In this response the company present new evidence which addresses key uncertainties highlighted in the ACD and results in robust ICERs which fall within the range normally considered cost effective for medicines which fulfil end of life criteria.

1. Introduction

Following the most recent appraisal committee meeting, the committee, in the ACD, highlighted three areas of uncertainty; duration of treatment, overall survival, and resource use. This document outlines Bayer's response to each of these, and an overview of further analyses conducted by Bayer.

2. Area of uncertainty 1: duration of treatment

- ***2.1 Background***
    - Following the first Committee meeting in July 2016, the Appraisal Committee did not agree with Bayer's approach for estimating mean duration of treatment based on time to progression (TTP) from the SHARP trial, and recommended that an unrestricted mean be used to estimate treatment duration from the SHARP trial.
    - In advance of the second Appraisal Committee meeting, Bayer implemented this change and presented an estimate of the unrestricted mean of treatment duration. Based on this extrapolation, using the SHARP trial, the analysis of the statistical fit indicated that the lognormal was statistically the most appropriate fit of the distributions considered. As such, lognormal was presented by the company for estimating duration of treatment.
    - The DSU, however, argued that based on the empirical data submitted by Bayer and visual inspection of the extrapolations, the Weibull and Gompertz distributions were most plausible.
    - Based on interpretation of statistical fit using criteria published by Kass et al 1995, the Appraisal Committee stated that the log-normal should be used to extrapolate treatment duration, and ruled out the other extrapolations.

- ***2.2 Actions/recommendation by AC to address areas of uncertainty following second appraisal***
    - The Appraisal Committee expressed concern regarding estimates of treatment duration and a recommendation was made to collect data from the Systemic Anti-Cancer Therapy dataset whilst within the Cancer Drugs Fund to further address this uncertainty.

- ***2.3 Bayer's response and considerations on the appraisal consultation document***
    - Previous estimates of mean treatment duration have been obtained via extrapolation of treatment data from SHARP. Not all patients in the SHARP study discontinued treatment, as a result SHARP can only provide an estimate of the treatment duration, and this varies greatly based on the parametric model selected.
    - As there was uncertainty around duration of treatment, Bayer felt that it would be informative to conduct further analyses on results from a cohort of the GIDEON study matched via propensity scoring to patients enrolled in SHARP as a means of further understanding the duration of treatment. The GIDEON study followed all patients until treatment discontinuation, resulting in an observed result as opposed to an estimate.
    - Furthermore, Bayer wishes to highlight supportive evidence, which when considered in addition to new data may validate the approach the company has taken to address this area of uncertainty.
    - The analyses conducted and observations from the ACD are presented below.

- ***2.4 Further analyses conducted***

2.4.1 Overview of analysis

- Following guidance from the first Appraisal Committee meeting, GIDEON a long-term observational study reporting the overall survival of 3,213 advanced HCC patients treated with sorafenib, was used to validate the extrapolation of overall survival applied to the SHARP RCT. For this analysis, patients in the GIDEON study were matched based on the baseline characteristics of patients enrolled in the SHARP trial. The DSU subsequently reviewed these methods and expressed satisfaction with Bayer's analysis, which was noted by the Appraisal Committee.

- As the GIDEON matched population resulted in patients reflective of those enrolled in SHARP and were all followed until treatment discontinuation, Bayer conducted an analysis to obtain the unrestricted mean duration of treatment from the matched GIDEON population (n=895)

- Critically this provides an unrestricted mean duration of treatment in a patient population approximately three times of that considered in SHARP, which could reduce uncertainty around extrapolating from incomplete data obtained from SHARP.

2.4.2 Results

- Results using the unrestricted mean from the GIDEON matched population resulted in a mean duration of treatment of [redacted]). The median duration of treatment was [redacted]). The mean dose intensity from the matched population was [redacted], with a 95% confidence interval from [redacted].

Table 1: Duration of treatment (days) in matched GIDEON population

Matched GIDEON population (n=895)	Mean/ 95% CI	Median/ IQR
Duration of treatment	[redacted]	[redacted]
	[redacted]	[redacted]
Mean dose intensity*	[redacted]	[redacted]

* Values are based on the mean daily dose per patient
** No mean daily dose was available for [redacted] GIDEON patients matched to the SHARP cohort.

2.4.3 Interpretation

- This is the first analysis showing an unrestricted mean (without the need for extrapolation) from a population matched to those enrolled in the SHARP trial. Additionally, the availability of overall survival data for this population may address the Committee's preference for obtaining overall survival and treatment duration from a matched study population (i.e. balance on patient characteristics).
- The mean duration of treatment of [redacted] derived from the analysis is substantially lower than that obtained via the extrapolation of SHARP. The mean dose intensity is also lower at [redacted], than the 710.5mg used in the economic model.
- In the matched GIDEON analysis the mean duration of treatment at a 95% confidence interval are [redacted]. The log-normal extrapolation of the SHARP study favoured by the Appraisal Committee provides an estimate of [redacted] months, exceeding the upper confidence interval of the GIDEON analysis by over 2.5 months.
- This provides robust evidence to support the DSU view that the log-normal may not provide the most plausible extrapolation
- It should also be noted that the mean derived from this analysis lies closer to the estimates from the extrapolations favoured by the DSU, which resulted in an estimated mean duration of treatment of [redacted].
- Findings from the matched analysis suggest that treatment duration is lower than estimates obtained via extrapolation of treatment duration from the SHARP trial. Considering a scenario where the Appraisal Committees preferred assumptions are updated to include a mean duration of treatment from the matched analysis, the resultant ICER would be between £32,819 and £54,929 using the log-normal and Weibull extrapolations of OS respectively (full results are presented in Section 6)
- Overall survival of patients in the matched GIDEON sample exceeded that observed in SHARP. It is significant that a lower duration of treatment (than that estimated from SHARP) in combination with a lower dose intensity ([redacted] vs 710.5mg) did not impact health outcomes. Therefore use of the SHARP overall survival data in

combination with duration of treatment derived from the GIDEON matched cohort is conservative.

- ***2.5 Observations on treatment duration from ACD and Bayer's response***
  - In response to the first ACD Bayer presented results based on AIC/BIC criteria to inform model selection for extrapolating duration of treatment and presented analysis using a log-normal extrapolation.
  - Based on both visual inspection and consideration of the external data, the Decision Support Unit (DSU) concluded that the Weibull and Gompertz distributions appeared to provide a more plausible explanation of treatment duration than the log-normal. Both of these extrapolations resulted in duration of treatment of [redacted].
  - In the Appraisal Committee meeting guidance from the DSU was not followed on the basis that using criteria outlined by Kass et al (1995) the Bayesian information criterion (BIC) statistics strongly indicated that the Weibull did not fit the data. On this basis the log-normal extrapolation was selected resulting in an estimate mean duration of treatment of [redacted] months.
  - Having considered the DSU response and guidance in the relevant technical support document Bayer concurs with the DSU that the Weibull and log-normal provide a more plausible extrapolation of the data.

- ***2.6 AC justification for extrapolation:***
  - The decision criteria cited by Kass et al 1995 provides an interpretation of the difference in Bayesian information criterion (BIC) in terms of model specification.
  - Table 2 presents the results of AIC/BIC tests conducted by the company when evaluating approaches to extrapolation of duration of treatment from the SHARP trial. The difference in BIC score of 10.4 between the log-normal and Weibull was used to determine that log-normal was the conclusive distribution for the extrapolation of duration of treatment.
  - Bayer would like to highlight potential uncertainties with the use of the Kass et al 1995 criteria for the basis of selecting the most appropriate model fit. These are discussed below following an overview of the criteria used for selection.

Table 2: SHARP OS model extrapolations: AIC/BIC statistics

	Exponential	Weibull	Loglogistic	Gompertz	Lognormal
AIC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BIC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

1. ***<u>Statistical fit should not be used in isolation to decide on model fit alone</u>***

   - The selection of the log-normal was based heavily on the use of criteria outlined by Kass et al (1995) this assessment does not allow for the consideration of evidence from visual inspection, or the availability of external data from the literature to contribute to curve selection.

- In the ACD guidance it is suggests that choice of extrapolations should not be based on statistical interpretation alone. In using statistical criteria alone this approach fails to acknowledge that the uncertainty lies in the extrapolation of the Kaplan Meir data to which Figure 1 shows that the log-normal overestimates.
- The DSU have previously noted the following with regard to extrapolation of treatment duration in this appraisal:
    - AIC/BIC criteria should not be used in isolation to determine model fit
    - Differences in AIC/BIC for extrapolation of duration of treatment from SHARP are relatively small
    - Upon visual inspection (of the model fits) the Weibull and Gompertz both appear to match the latter part of the KM curve (where there is uncertainty) better than the log-normal.

2. Methods outlined by Kass et al (1995) may not be appropriate for assessing the fit of alternative parametric models

- The NICE Technical Support Document (TSD) 14 provides guidance on methods for undertaking survival analysis for use in an economic evaluation when patient-level data are available.
- This TSD provides no guidance in terms of inferring from numerical differences in AIC/BIC statistics to inform selection of a parametric model. However, it does advise that the interpretation outlined by the Appraisal Committee to be the grounds for the selection of the log-normal distribution, may not be an appropriate criteria for the selection of parametric models. It states that in the past this has been used erroneously, stipulating that *"measures such as the negative 2 log likelihood are only suitable for comparing nested models, whereby one model is nested within another (for example, one model adds an additional covariate compared to another model). Different parametric models which use different probability distributions cannot be nested within one another. Thus the negative 2 log likelihood test is not suitable for assessing the fit of alternative parametric models"*

3. Implications for using criteria published by Kass et al for other areas of uncertainty

- In the circumstance that this test is accepted by the Committee as a definitive method for the selection of parametric models for extrapolation of health outcomes, consideration should be given to the results of applying this interpretation to overall survival fits conducted on the matured overall survival cohort data from the matched GIDEON population. These analyses were conducted to inform the extrapolations previously applied to the SHARP trial (further consideration of this is presented in Section 3)

Interpretation:

- Although the Appraisal Committee concluded that the log-normal was the superior fit, and Bayer had previously presented results using the log-normal, it was clear that there

still existed some uncertainty from the DSU regarding the most appropriate model for extrapolation.

- If an alternative model was in fact the most appropriate fit or use of these selection criteria was used to address uncertainty in regard to most plausible extrapolation of overall survival, the resultant ICER are likely to be within the cost-effectiveness range for end of life treatments. Results from these analyses are presented in Section 6.

Supportive evidence

In light of new evidence provided in the sections above, and the Appraisal Committee's recommendation that duration of treatment data collected from SACT would be helpful in addressing this uncertainty, the company believes that real-world studies reporting on the duration of treatment should be used in determining the most plausible source or extrapolation of treatment duration for the economic model.

- Over the past three years over [redacted] HCC patients[1] have had access to treatment in England via the CDF. Unlike for new treatments, there is therefore published evidence on the duration of treatment derived from use on the NHS.
- The duration and therefore cost of treatment are reflective of those faced by the NHS and as such should be considered at the very minimum to inform an appropriate extrapolation. Extrapolations conducted on the SHARP data with resulting estimates of mean duration of treatment are presented in Figure 1 and Table 3 respectively.

Figure 1: Extrapolation of treatment duration from SHARP

[1] Patient numbers taken from July 2013 - June 2016

Table 3: Estimated mean duration of treatment obtained via extrapolation

	Exponential	Weibull	Loglogistic	Gompertz	Lognormal
Mean DoT (months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

- Whilst many studies cannot provide a mean duration of treatment due to length of follow-up, it is possible to compare medians across studies and note the relationship where both a median and unrestricted mean is reported to provide further evidence that the extrapolation of [redacted] months to be unrealistically high.

- Table 4 presents findings from the literature showing UK sources. The median duration of treatment in all sources is lower than the matched GIDEON population. Although it is impossible to be certain, it is likely that this indicates that treatment duration in the UK is less than the unrestricted mean reported in the matched GIDEON population.

Table 4: Empirical estimates of treatment duration

Source	Sample size (n=)	Duration of treatment (months)	
		Median	Mean
J King et al (2013)	379	3.2	NR
GIDEON (total population)	3,202	3.46	5.52
GIDEON (matched population)	895	[redacted]	[redacted]
J King et al (2016)	484	3.6	NR
Ziogas et al (2017)	Age≤ 75: 151 Age>75: 31	Age ≤75: 3.0 95% CI (2.5–3.9) Age >75: 5.1 95% CI (3.1–7.1)	NR

NR: Not reported

- ***2.7 Conclusion***

  - New evidence considering the unrestricted mean duration of treatment from a cohort of the GIDEON study population matched based on characteristics of those enrolled in SHARP (n=895) provides for the first time a true unrestricted mean duration of treatment (as opposed to extrapolated estimates from SHARP)
  - The mean unrestricted duration of treatment in the matched GIDEON population is [redacted] months, with a 95% confidence interval from [redacted] months to [redacted] months. This provides strong evidence that the log-normal extrapolation applied to SHARP used to derive a treatment duration estimate of [redacted] months is unrealistically high. Further to this the company highlight uncertainties in the criteria published by Kass et al used in which to justify the selection of the log-normal extrapolation, identifying if correct in its use, how this could be applied to other uncertainties such as overall survival
  - This new analysis reports a median and unrestricted mean that is aligned with previous empirical evidence sources presented to the Committee. Incorporation of this treatment duration in the economic model, or using these results to validate either the Weibull or

Gompertz extrapolations of treatment duration lead to ICERs likely to fall into the range of cost-effective estimates which are normally in the acceptable range for end of life medicines.

3. Area of uncertainty 2: extrapolation of overall survival

- ***Background***
  - In the first Appraisal Committee meeting, data from GIDEON, a large international prospective study collecting overall survival data from 3,213 sorafenib patients, was presented to validate the parametric model used in determining the long-term extrapolation of overall survival from SHARP.
  - The analysis presented was deemed by the Appraisal Committee to validate the log-normal extrapolation of the data in SHARP; however, there was uncertainty in the comparability of the GIDEON study population to patients enrolled in SHARP.
  - In the second Appraisal Committee, Bayer met the requests of the Committee in presenting the population of GIDEON matched via propensity scoring to patients enrolled in the SHARP trial. The DSU expressed satisfaction with the matching performed by the company and this subsequently provided overall survival data from a sample of 895 patients
  - The Committee noted that in general the log-normal function used by the company to extrapolate survival beyond SHARP fitted the matched GIDEON data better than the Weibull function.

- ***Actions/recommendation by AC to address areas of uncertainty following second appraisal***
  - The Committee considered that uncertainty remained around the most appropriate model to extrapolate overall survival, with the Weibull distribution remaining plausible. The Committee agreed that the most plausible ICER for sorafenib was lower than the midpoint of the log-normal and Weibull distributions.
  - In light of the uncertainty, a recommendation was made to collect data from the Systemic Anti-Cancer Therapy data set to help resolve some of this uncertainty.

- ***Bayer's response and considerations on the appraisal consultation document***
  - In this section, Bayer wishes to comment on issues identified in the appraisal consultation document, including the consistency in the use of criteria published by Kass et al in determining model selection and the availability of follow-up overall survival data from the SHARP trial, which was highlighted by the Appraisal Committee as a potential for addressing uncertainty in the ACD.

- ***Consistency in use of statistical criteria in determining model (Kass et al 1995)***
  - The justification for the choice of treatment duration extrapolation by the Appraisal Committee was based on criteria for the interpretation of the numerical differences in BIC scores published by Kass et al (1995).

- As this approach was used by the Appraisal Committee to select and recommend a preferred method for extrapolation of duration of treatment, it may be reasonable to argue that this approach should also be considered for the selection of an appropriate distribution for the extrapolation of overall survival.
- If this approach was applied, the use of the log-normal extrapolation could be validated for extrapolation of overall survival, as interpreted by the difference in BIC scores.
- Upon application of the Kass et al decision criteria to the extrapolation of overall survival, the analysis conducted on the matched GIDEON population to assess the correct model to fit to the SHARP data resulted in a difference in the BIC scores between the log-normal and Weibull of 20.68 points. To use the interpretation provided by the Appraisal Committee this would strongly indicate that the Weibull does not fit the data (to a greater extent than the 10.4 point difference observed in the duration of treatment analysis)
- Results of the AIC/BIC analyses conducted on the log-normal and Weibull overall survival extrapolations are presented in Table 5.

Table 5: Results of AIC/BIC criteria in assessing the OS extrapolation from SHARP

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]
Difference	**20.67**	**3.39**	**24.06**	**20.68**	**3.38**	**24.06**

Interpretation

- Should the criteria outlined by Kass et al be used in the assessment of extrapolation for duration of treatment to conclusively favour the log-normal distribution, this same criteria should be used to evaluate the extrapolation of overall survival to conclusively favour the log-normal distribution
- When considering the population from the GIDEON study that are matched based on patient characteristics to those enrolled in SHARP is considered, a difference of 20.7 points can be seen between the best fitting extrapolation (log-normal) and the Weibull distribution. This difference is approximately twice that seen when results are inferred in a similar way for the extrapolation of treatment duration.

- ***Request by Appraisal Committee for follow-up overall survival data from the SHARP trial***
  - It was noted by the Appraisal Committee that further follow-up survival data from SHARP, as used in the economic model, could clarify the uncertainty surrounding overall survival.

- Bayer would like to note that in the extrapolation of overall survival from SHARP used to inform the economic model, all available survival data from the SHARP trial has been used. Unfortunately, as raised in the original submission process, the SHARP trial did not follow-up overall survival past 19 months.

4. Area of uncertainty 3: Resource use

- In order to reduce uncertainty resource use data for patients not treated with sorafenib would be required. No new evidence sources have been identified that offer comparative data, as patients with advanced HCC no longer receive best supportive care if suitable for sorafenib.
- The company proposes that the Appraisal Committee's preferred assumption, to pool results from the original and update resource use surveys to address uncertainty.

5. Company response to recommendation to enter the Cancer Drugs Fund

- Bayer acknowledges that for treatments that are new to market, there may be value in the prospective collection of real-world evidence to address certain uncertainties. However sorafenib has been used to treat advanced HCC on the CDF for 6 years, and obtained marketing authorisation over 10 years ago. In addition sorafenib is currently recommended for routine use in Scotland and Wales. For this reason data which would routinely not be available for new treatments is available and published for sorafenib.
- The GIDEON study is a large multicentre prospective study considering both mean duration of treatment and overall survival in over 3,000 patients treated with sorafenib. Through this appraisal Bayer has matched a large cohort of patients, based on patient characteristics to those enrolled in SHARP (n=895). This type of analysis would not be possible in SACT and as such there would be no method to link observations back to the SHARP population.
- Additional publications of real-world evidence of the use of sorafenib are also available providing evidence of use within the NHS.

6. Results: Exploratory analyses conducted based on new evidence and interpretation of the AC response

- The appraisal consultation document concludes that the true estimate of overall survival with sorafenib was likely to lie between the estimates from the log-normal and Weibull distributions with these two extrapolations informing the range of plausible ICERs (£49,500 - £87,000).
- The Committee agreed that the most plausible ICER was likely to be lower than the midpoint of the preferred ICER range (£68,250) but higher than technologies previously accepted that had met the end of life criteria. That is that the most plausible ICER is likely to be closer to the log-normal extrapolation, a conclusion that has been reached by both the DSU and Appraisal Committee.

- Using the above guidance from the Appraisal Committee, in addition to new evidence submitted to address uncertainty in the duration of treatment from the matched GIDEON population, new cost-effectiveness analyses have been conducted which all reflect the Appraisal Committees preferred assumptions of:
    - Independent assessment of progression
    - Wastage (7 days)
    - Pooled resource use estimates
    - Treatment duration using an unrestricted mean

- ***6.1: Cost-effectiveness results using duration of treatment obtained from the matched GIDEON sample and the Appraisal Committee's preferred assumptions***
    - The ICERs using both the log-normal and Weibull extrapolations for overall survival and calculation of the midpoint*,in addition to duration of treatment derived from the matched GIDEON sample are presented in Table 6.
    - The results detail the ICER using the mean dose/patient/day from the GIDEON study ([redacted]) and that reported in the SHARP study (710.5mg).

Table 6: Updated PAS price: Cost-effectiveness results using treatment obtained from the matched GIDEON sample

		Duration of treatment: Matched GIDEON population (n=895) ([redacted])	
		GIDEON matched mean dose ([redacted])	SHARP mean dose (710.5mg)
Overall survival	Log-normal	£32,819	£36,050
	Weibull	£54,929	£61,290
	Midpoint* (log-normal and Weibull ICERs)	**£43,874**	**£48,670**

***** The midpoint between log-normal extrapolation and the Weibull, represents the upper bound of the Appraisal Committee's plausible ICER range

- ***6.2 Cost-effectiveness results assuming plausible ICER is lower than the mid-point of the log-normal and Weibull and the log-normal extrapolation.***
    - The Committee concluded in the ACD that the most plausible ICER was likely to be lower than the mid-point of the log-normal and Weibull ICERs (calculated in Table 6)
    - Using the calculated midpoint as an upper bound (Table 6) and;
    - Using the log-normal extrapolation as a lower bound (also presented in Table 6)
    - A midpoint of this range which reflects a midpoint of the plausible ICER range is presented in Table 7.
    - This analysis uses all Appraisal Committee's preferred assumptions (except the derivation of treatment duration based on log-normal extrapolation from SHARP)

Table 7: Updated PAS price: Cost-effectiveness results following Appraisal Committee guidance that the plausible ICER is likely lower than the mid-point of the log normal and Weibull and the log-normal extrapolation

		Duration of treatment: Matched GIDEON population (n=895) ([redacted])	
		GIDEON matched mean dose ([redacted])	SHARP mean dose (710.5mg)
Overall survival	Lognormal	£32,819	£36,050
	Midpoint (log-normal and Weibull ICERs)	£43,874	£48,670
	ICER	**£38,347**	**£42,360**

- ***6.3 Cost-effectiveness results presenting variation of the mean duration of treatment between that observed in the matched GIDEON sample and the estimates of duration of treatment derived from the Weibull and Gompertz extrapolations from the SHARP trial as favoured by the DSU***

Acknowledging that uncertainty exists around the duration of treatment, Bayer have varied the mean duration of treatment between that observed in the matched GIDEON sample and the estimates of duration of treatment derived from the Weibull and Gompertz extrapolations from the SHARP trial, resulting in a mean duration of treatment of [redacted]. In this analysis both the dosing intensity from the GIDEON sample and the SHARP trial are presented. Results of these analyses are presented in Table 8.

Table 8: Updated PAS price: Cost-effectiveness results varying source of duration of treatment (matched GIDEON population/SHARP)

		Duration of treatment			
		Matched GIDEON population (n=895) [redacted]		Gompertz/Weibull extrapolation from SHARP: ([redacted])	
		GIDEON matched mean dose [redacted]	SHARP mean dose 710.5mg	GIDEON matched mean dose [redacted]	SHARP mean dose 710.5mg
Overall survival	Lognormal	£32,819	£36,050	£37,202	£41,073
	Midpoint (log-normal and Weibull ICERs)	£43,874	£48,670	£50,380	£56,125
	ICER	**£38,347**	**£42,360**	**£43,791**	**£48,599**

Conclusion: Results

- In this section Bayer presented cost-effectiveness analyses varying two of the key areas of uncertainty, the mean duration of treatment and extrapolation of overall survival. In

each of the analyses Bayer have reflected the Appraisal Committee's preferred assumptions unless stated.

- As documented in the ACD the most plausible ICER lies between the midpoint of the log-normal/ Weibull ICER as an upper bound, and the log-normal exploration as a lower bound (equivalent to a 25%:75% weighting)

Use of matched GIDEON treatment duration

- Analyses presented show that when the observed mean duration of treatment is used from the matched GIDEON sample, the ICER does not rise above £50,000 per QALY when the extrapolation of overall survival is varied from the log-normal (£32,819 per QALY) to the midpoint of the log-normal and Weibull ICER (£43,873 per QALY). This provides a plausible ICER of £38,347 per QALY and considers the impact of the Weibull distribution in line with the Committee's preferred assumptions.

Validation of the extrapolation of the estimation of treatment duration applied to SHARP

- A final analysis considered the cost-effectiveness of sorafenib when duration of treatment analysis conducted on the matched GIDEON population was used to validate extrapolated estimates from SHARP (using either the Weibull or the Gompertz extrapolation following the preference of the DSU).
    - The midpoint of the plausible cost-effectiveness range outlined by the Appraisal Committee was used to reach a plausible ICER of £48,599.

- Over the course of the appraisal the company has sought to incorporate a range of preferred assumptions into cost-effectiveness estimates. In the results presented in this section the company explored duration of treatment, providing two methods that meet the preferred assumptions. Both show sorafenib to be cost-effective. Despite evidence favouring the log-normal as a preference for overall survival extrapolation, the company has sought to incorporate the Committee's preferred assumptions based on guidance provided in the ACD.

7. Overall conclusion

- GIDEON is the most robust data source in which to assess outstanding uncertainties highlighted by the Appraisal Committee, with a population of over 3,200 patients treated with sorafenib. Crucially, through this submission Bayer have matched a large cohort of GIDEON patients, based on patient characteristics to those enrolled in SHARP (n=895). This allows considerations of outcomes in patients with characteristics reflecting those enrolled in SHARP.

- New evidence presented in this response finds the unrestricted mean duration of treatment in the GIDEON population to be [redacted], with a 95% confidence interval ([redacted] months) suggesting that the estimates from the SHARP study to be a gross overestimate in treatment duration.

- Matched dosing intensity and overall survival from GIDEON allows uncertainty surrounding the relationship between length of treatment and effectiveness in clinical practice to be assessed. With the matched population providing a greater overall survival benefit than that seen in SHARP, using matched treatment data from GIDEON and efficacy from SHARP could be seen as conservative.

- Cost-effectiveness results presented in Section 6, reflect the Appraisal Committee's preferred assumptions (other than from the updated evidence on duration of treatment) and show the range of ICER to vary between £38,346 and £48,599 per QALY. Each of these ICERs reflects the potential plausibility of the Weibull in line with guidance in the appraisal consultation document. Given all evidence considered to date it is with confidence that these ICERs are likely to be in the range normally considered cost effective for medicines which fulfil end of life criteria

- Sorafenib is an innovative treatment, which upon launch resulted in a step change for the treatment of patients with advanced HCC. In over 10 years following its marketing authorisation there still remains no alternative for patients with advanced disease. Hepatocellular carcinoma is the second most common cause of cancer death worldwide, with future treatments for advanced HCC patients reliant on the continued availability of sorafenib as a first line treatment option.

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

Response to queries raised by the Evidence Review Group 30th January 2017

In response to questions raised by the ERG received on the 26th January 2016. Bayer is pleased to provide the following responses:

1. **Please can Bayer provide a KM analysis of the GIDEON time to discontinuation data?**

The Kaplan Meier graph presented in figure 1 details treatment discontinuation using the full analysis set of [redacted] subjects treated with sorafenib from the GIDEON study.

Figure 1: Treatment discontinuation (weeks) – descriptive statistics (full analysis set)

Figure 2 provides a table extracted from the CSR, confirming that all patients discontinued treatment during the study.

Figure 2: Reasons for discontinuation from treatment or from study – enrolled patients

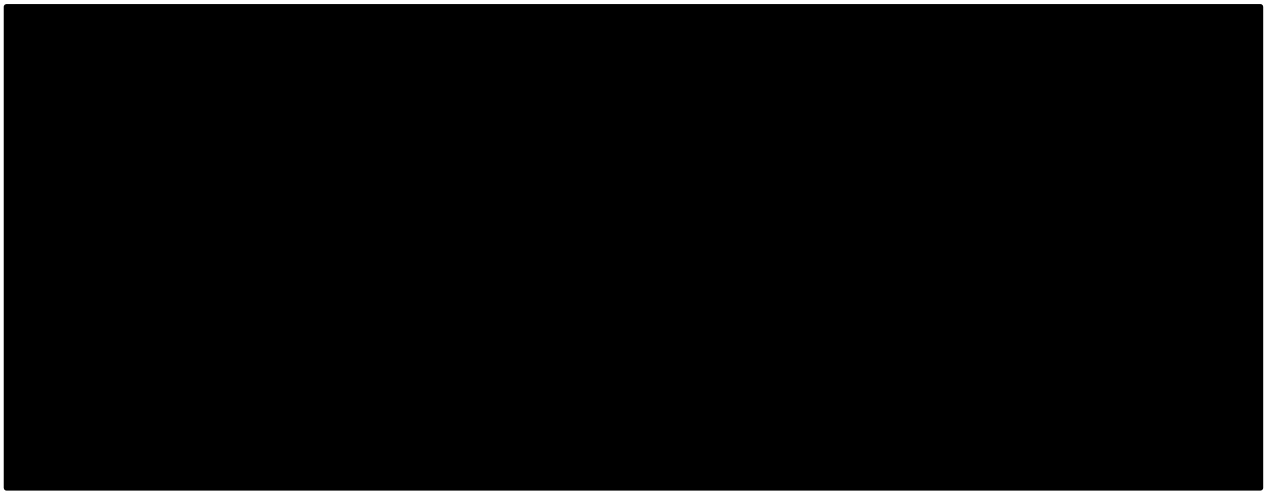
[1] includes only patients treated with Nexavar® for the column "Reasons for discontinuation from treatment"
[2] After the discontinuation of treatment with Nexavar®, patients were further followed up if possible and information including data on tumor assessment, measurement of body weight, blood pressure, and evaluation of ECOG performance scale and Child-Pugh were documented

Figure 3 provides descriptive statistics from the Kaplan Meier analysis. In regard to the censored patients:

- For [redacted] patients no reason was given for treatment discontinuation, these patients were censored at their last recorded visit.
- [redacted] patients were lost to follow-up, these patients were censored at their last recorded visit.

In consideration of the above censoring events, given the number of patients enrolled in the study Bayer believe that event data from [redacted] patients treated with sorafenib to be the strongest data source available to estimate the mean unrestricted duration of treatment.

Figure 3: Treatment discontinuation (weeks) - descriptive statistics (full analysis set)



2. **For Ziogas et al (2017), the abstract the ERG have located seems to report different treatment durations (4.2 and 5.6) to those reported by the company in Table 4 (3.0 and 5.1). Please could Bayer provide the relevant reference, and any other references in ACD2 not already provided?**

In Bayer's response to the ACD the median duration of treatment was cited from a study by Ziogas et al. The values noted by the ERG in the query relate to time to treatment failure (TTF), which also provide evidence that the duration of treatment estimate of [redacted] favoured by the committee is unrealistic. The company has provided this publication as an attachment to the response.

The median duration of sorafenib treatment (95% CI) was 3.0 months (2.5–3.9) for the patients in group A (n=151) and 5.1 (3.1–7.1) months for the patients in group B (n=39) (P=0.075). The two groups relate to patient age at commencement of therapy: (A) those that were up to 75 years old before starting sorafenib and (B) those that were older than 75 years old.

The company present these values for two reasons.

1. **95% confidence intervals suggest [redacted] months treatment duration is unrealistic**

Whilst it is the mean duration of treatment that is of interest for the purpose of the economic evaluation, the median treatment duration presented in this study has a 95% confidence interval suggesting that a treatment duration estimate of [redacted] months to be unrealistic. This exceeds the upper bound of Group A's (n=151) treatment by [redacted] months and the smaller Group B's (n=39) by [redacted] months. It is important to note that these estimates are from UK clinical practice and represent all patients treated in a large London hospital between 2005 and 2015.

2. **Alignment with treatment duration findings from the matched GIDEON population**

When median treatment duration from the two groups is weighted by sample size, a median of 3.43 months is obtained. Comparing this to the median obtained from the matched GIDEON population ([redacted] months) we can infer that the mean duration of treatment, and therefore costs to the NHS, are likely to be lower than that seen in matched GIDEON population ([redacted] months).

The company also note that confidence intervals provided around the time to treatment failure (TTF) results, also give reason to believe the committee estimate of [redacted] months of treatment duration to be unrealistic.

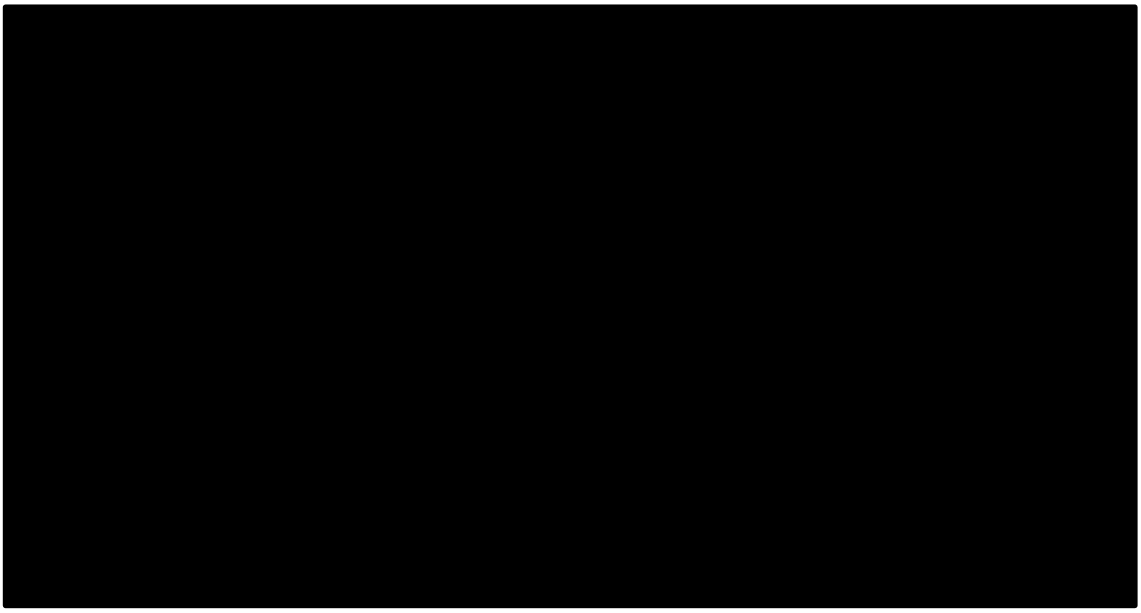
Sorafenib for advanced hepatocellular carcinoma (review of TA189) [ID1012]

Response to queries raised by the Evidence Review Group: Addendum 30th January 2017

Further to Bayer's response to questions raised by the ERG, please find below details of the Kaplan-Meier analysis for treatment discontinuation in the cohort of GIDEON patients matched, based on patient characteristics, to those enrolled in SHARP.

Kaplan Meier analysis of discontinuation in matched GIDEON population

The Kaplan Meier plot presented in figure 1 details treatment discontinuation using the matched GIDEON population resulting in [redacted] subjects treated with sorafenib, matched based on patient characteristics to patients enrolled in the SHARP study.



In this analysis [redacted] patients were censored at their last recorded visit. In consideration of these events, the matched GIDEON data provides information on the unrestricted mean duration of treatment in a sorafenib treated population nearly three times that considered in SHARP. The benefit of this evidence source is that as all events have occurred there is no need for extrapolation, as such this results in an actual as opposed to estimated duration of treatment.

These results are further supported by the overall GIDEON population. The company believe the matched GIDEON population to be the strongest evidence source in which to determine the actual as opposed to estimated unrestricted mean duration of treatment in population matched to SHARP.

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

Submission from the British Liver Trust

As per our previous response the British Liver Trust would want patients with advanced stage hepatocellular carcinoma who have failed or are unsuitable for surgical or loco-regional treatment to have access to treatment with Sorafenib.

Sorafenib is the only treatment available for this group of patients, the alternative being specialist palliative care. We highlight, again, the immense benefits of not only prolonging life but also the improved symptom control and quality of life that can be achieved.

As Sorafenib is available to patients in Scotland and Wales it would be unfair not to give equal access to patients in England.

Royal College of Physicians
11 St Andrews Place
Regent's Park
London NW1 4LE
Tel: +44 (0)20 3075 1560

www.rcplondon.ac.uk

National Institute for Health and Care Excellence
10 Spring Gardens
London
SW1A 2BU
jenna.dilkes@nice.org.uk

From The Registrar
[redacted]

10 January 2017

Dear Jenna

Re: ACD2 - Consultees & Commentators: Hepatocellular carcinoma (advanced and metastatic) - sorafenib (first line) (review of TA189) [1012]

The Royal College of Physicians (RCP) plays a leading role in the delivery of high quality patient care by setting standards of medical practice and promoting clinical excellence. We provide physicians in the United Kingdom and overseas with education, training and support throughout their careers. As an independent body representing over 33,000 Fellows and Members worldwide, we advise and work with government, the public, patients and other professions to improve health and healthcare.

The NCRI-ACP-RCP-RCR are grateful for the opportunity to respond to the above consultation. We would like to make the following comments.

1. Has all of the relevant evidence been taken into account?

- The UK sorafenib audit including 448 patients has now been published in full and is attached. This provides good multisite data across the UK regarding mean daily dose and median time on treatment for the UK population. Additionally, it provides data on survival outcomes along with prognostic factors associated with survival for sorafenib treated patients. The paper provides support for the efficacy of sorafenib in patients with good PS and preserved liver function. Of note this study was academically driven with no commercial support.
- The Gideon study is referred to in the appraisal document but there are multiple publications and it is not clear which data were considered. The most recent publication is attached from Dec 2016. This is global study sponsored by Bayer and we do not believe that the UK contributed to this. Hence the data from the UK audit may be more relevant. That said, there are some consistent findings with regard to poorer outcomes for Child B patients >B7. The daily dose and duration are also broadly similar.

2. Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?

- Our experts felt they were not qualified to judge this and not clear how this was calculated.

3. Are the recommendations sound and a suitable basis for guidance to the NHS?

- We support the recommendation to make sorafenib available within the CDF and the proposal for prospective data collection to clarify the uncertainties that remain regarding efficacy, time on treatment and resource usage.
- There are a number of additional benefits from sorafenib being available which include the placement of clinical trials in the UK which provide access to new drugs. During the past few years the UK has gained an excellent reputation for recruitment into HCC trials from which some patients have gained outstanding benefits. In the absence of sorafenib being available, the UK would risks becoming an international outlier and not considered and appropriate environment to conduct clinical research in HCC.

4. Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of race, gender, disability, religion or belief, sexual orientation, age, gender reassignment, pregnancy and maternity?

- No

Yours sincerely

[redacted]
Registrar

NHS England submission into the NICE re-appraisal of sorafenib in the treatment of hepatocellular carcinoma: January 2017

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 this disease) of 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). These benefits came at the expense of significant but tolerable toxicity. The trial was stopped after the interim analysis showed this survival advantage and hence longer term information on survival in the trial patients is not known.

2. The SHARP population was meant to enrol only Child-Pugh A patients, the eventual entry reflecting 97% Child-Pugh (C-P) A patients and 3% C-P B patients. The trial consisted of 54% performance status (PS) 0 patients, 38% PS 1 and 8% PS 2. The median duration of treatment was 5.3 mo. All patients started at 800mg of sorafenib/day and the mean daily dose was 711mg.

3. There have been at least 4 randomised trials with sorafenib as a treatment arm and the median survival durations in these trials for patients receiving sorafenib have been between 8.5 and 10.2 months.

4. The UK audit of 448 patients published in Clinical Oncology 2016 had 77% known C-P A patients and 16% known C-P B patients. The median survival duration was 9.5 months for C-P A patients and 4.6 mo for C-P B. The audit comprised 26% PS 0 patients, 49% PS 1 and 21% PS 2. The mean daily dose was 590mg and 62% started at 800mg/day and 33% at 400mg/day. The median duration of treatment was 3.6 months.

5. The GIDEON audit of 3202 patients published in J Hepatology 2016 had 61% C-P A patients and 21% C-P B patients. The median survival durations were 13.6 months for C-P A patients and 5.2 mo for C-P B. The audit did not split PS 0 and 1 patients in this publication. The mean daily dose was not reported in the analysis but 72% started at 800mg/day. The median duration of treatment was 4.1 mo for C-P A and 2.3 mo for C-P B. Sorafenib had been discontinued at 8 weeks in 26% of C-P A patients and 42% of C-P B; at 28 weeks, 33% of C-P A patients and 20% of C-P B were still on treatment, respectively.

6. Bayer's matched GIDEON to SHARP analysis of 895 patients which was done to obtain longer term survival information had a median treatment duration of 3.9 mo with a mean daily dosage of 620mg and a mean treatment duration of 6 mo.

7. The SPC for the use of sorafenib in hepatocellular carcinoma is based on the results of the SHARP trial and states that 'there are limited data from this study in patients with Child-Pugh B liver impairment'. There is therefore no robust randomised controlled trial data to demonstrate the survival benefit of sorafenib in C-P B patients

8. NHS England would normally wish the NICE Technology Appraisal Committee to use the same patient source of information on which to base its preferred estimates of both treatment duration and overall survival. The SHARP trial data provides evidence of the median duration of treatment with sorafenib (5.3 mo) in the same population of patients which provides the evidence of gain in median duration of survival with sorafenib over best supportive care. Separating the source of information of treatment duration from the source that provides the survival data usually increases uncertainty.

9. The information from the 448 patient UK audit and the 3202 patient GIDEON audit suggests a lower median duration of treatment than in the SHARP trial (3.6 and 4.1 mo vs 5.3 mo) and the UK audit points to a lower mean daily dose of sorafenib (590mg vs 711mg) than the SHARP data. The other clear message from both the UK and GIDEON audits is that patients with C-P B do much worse than C-P A (A vs B is 9.5 vs 4.6 mo in the UK audit and 13.6 vs 5.2 mo in GIDEON).

10. The Bayer analysis of 895 patients matched from GIDEON to SHARP has a median treatment duration of 3.9 mo which is similar to the audit figures above in paragraph 9. The mean daily sorafenib dose in the matched analysis (620mg) is close to that in the UK audit (590mg), the difference being potentially explained by the different case mixes in the two populations.

11. The issue therefore is whether it is reasonable to retain the initial survival data from SHARP and its consequent modelling with GIDEON to SHARP analaysis but use the treatment duration and dosage information from the matched GIDEON to SHARP analysis. Oncologists have learned how to use sorafenib better than at the time at which the SHARP trial was performed, the trial having mandated the starting dose to be 800mg daily. As a consequence of this learning, sorafenib prescription in the clinic is often at a starting dose of less than 800mg daily (38% of patients in the UK audit and 28% in GIDEON as a whole) and the greater experience with continued sorafenib administration has led to earlier dose reductions than may have taken place in

SHARP. There is thus a rationale for using the matched GIDEON to SHARP figure for mean daily dose of sorafenib. With this consideration also comes the issue of duration of treatment and the same issues apply in terms of a rationale for using the median figure of 3.9 mo (and mean 6 mo figure) in the GIDEON to SHARP analysis. The advantage of using these GIDEON to SHARP figures is that they are more likely to represent what happens in the clinic currently; the disadvantage is that the greater duration of treatment and higher daily dose in SHARP may have translated into the degree of the survival gain with sorafenib, though if so, NHS England considers this effect is likely to be small.

12. If the NICE TA Committee chooses to use the matched GIDEON to SHARP analysis to provide the treatment duration, sorafenib dosage and longer term survival in the economic model and these assumptions are key to the determination of cost effectiveness, then NHSE would state that any recommendation by NICE for the use of sorafenib should directly follow the inclusion criteria for patients in the SHARP trial:

    - histologically confirmed diagnosis of hepatocellular carcinoma
    - metastatic disease or advanced local disease that is ineligible for or failed surgical or locoregional therapies
    - no previous systemic therapy for hepatocellular carcinoma
    - Child-Pugh liver function class A
    - performance status of 0-2.

13. If NICE recommends sorafenib for just Child-Pugh A patients, NHS England accepts that the present CDF inclusion of what amounts to C-P B7 patients would end. Given the survival figures in the UK and GIDEON audits for C-P B vs A patients (and the median survival duration in the whole GIDEON audit for B7 patients was 6.2 mo), NHS England regards such a conclusion of a NICE recommendation for only C-P A patients to be based on the best current data. This is because there were only 3% C-P B patients in SHARP (despite being excluded in the trial design) and there were very low survival durations for C-P B patients in the 2016 publications of these UK and GIDEON audits. In addition, consideration of the toxicity of sorafenib in patients with such a short life expectancy would also support this conclusion.

14. NHS England believes that sorafenib's patent in hepatocellular cancer will expire in 2020 or thereabouts.

15. 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, lenvatinib 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 competitive NHS England Specialised Commissioning prioritisation process.

16. 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; and if cost-effective, in what group of patients.

Prof Peter Clark

Chair NHS England Chemotherapy Clinical Reference Group and National Clinical Lead for the Cancer Drugs Fund

January 2017

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 ACD2

BY THE DECISION SUPPORT UNIT

30th January 2017

Iñigo Bermejo and Sabine Grimm

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.

This report should be referenced as follows:

Bermejo I., Grimm S. 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 ACD2. 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 their response to the second Appraisal Consultation Document (ACD2), the company submitted a new Patient Access Scheme (PAS) that includes an additional discount of [redacted] off the previous reduced price. The company also expressed that existing evidence for sorafenib was more robust than evidence potentially collectable through the Cancer Drugs Fund (CDF). In particular, the company highlighted the retrospective study GIDEON as one such evidence source: the company provided the unrestricted mean treatment duration and the mean daily dose for the GIDEON population matched using propensity scores to that of the pivotal trial SHARP. The company presented analyses using the unrestricted mean treatment duration and mean dose from the matched GIDEON population. The company also argued against the committee's preference of using a lognormal distribution to extrapolate treatment duration and presented analyses using the Weibull instead. The company presented incremental cost-effectiveness ratios (ICERs) of sorafenib versus best supportive care (BSC) including the revised PAS price and using the alternative sources for treatment duration and mean daily dose. These ICERs ranged from £38,825 to £49,241 per QALY gained, using a weighted average of the ICERs calculated with the lognormal and the Weibull (75% and 25% respectively) to extrapolate overall survival.

The DSU notes that both treatment effectiveness and cost data from the same source should be used and that the committee already concluded that data from SHARP should be used to estimate duration and total cost of treatment. Therefore, the DSU notes that using mean dose and treatment duration from the matched GIDEON population in the analysis whilst keeping the comparative efficacy of SHARP is potentially misleading, as it breaks the link between the quantity received by patients and the resulting treatment benefit. The DSU also believes that the results from the matched GIDEON population might be biased due to uncontrolled variables in the matching. Regarding the extrapolation of treatment duration, the DSU agrees that measures of statistical fit should not be used in isolation to justify the selection of a single curve. The DSU still slightly favours the Weibull distribution to extrapolate treatment duration based on the visual inspection of the curve fit and the external evidence presented by the company but believes that the lognormal curve should also be used in the analyses. In light of the limitations associated with drawing cost and effectiveness data from different sources, and considering the new arguments on most plausible extrapolation methods for treatment duration, the DSU only updated its analyses with the new PAS price. The DSU

estimates that the most likely ICER of sorafenib compared with BSC based on the revised PAS price lies between £48,657 and £71,575 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
BSC	Best supportive care
CDF	Cancer Drugs Fund
DSU	Decision Support Unit
GIDEON	Global Investigation of therapeutic DEcisions in hepatocellular carcinoma and Of its treatment with sorafeNib
ICER	Incremental cost-effectiveness ratio
KM	Kaplan-Meier
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OS	Overall survival
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

In their response to the second Appraisal Consultation Document (ACD2)[1], the company incorporated a revised Patient Access Scheme (PAS). The revised PAS includes an additional discount of [redacted] off the previous price (a [redacted] discount on the NHS list price) resulting in a new cost per pack of [redacted]. The new price would be available via the Commercial Medicines Unit framework agreement and would apply to all indications of sorafenib.

1.1. DURATION OF TREATMENT AND MEAN DOSE

In the ACD2[1], the Appraisal Committee (AC) concluded that sorafenib should be recommended for use within the Cancer Drugs Fund (CDF) because several uncertainties remained that could be addressed through collecting outcome data from patients treated in the National Health Service (NHS). However, in response to the ACD2, the company claimed that there already existed evidence on sorafenib that was more robust than the evidence that could be collected through its use in the CDF. In particular, the company referred to GIDEON,[2] a long-term observational study considering the overall survival of 3,213 advanced hepatocellular carcinoma (HCC) patients treated with sorafenib. In response to the first ACD (ACD1)[3], the company undertook an analysis matching the population of GIDEON to that of the pivotal trial SHARP.[4] This analysis was requested by the AC during the first meeting to inform the choice of parametric curve for the extrapolation of overall survival (OS), but the result of the analysis was inconclusive. The DSU notes that whereas this analysis was useful to try to inform the choice of parametric curve, it should not be used to replace data that is already available from a randomised controlled trial (RCT) such as SHARP. [4] In its critique to the company's response to the ACD1)[3], the DSU did not raise any issues regarding the matching but was aware of its limitations, such as the exclusion of important variables like vascular invasion presence, extrahepatic spread presence, hepatitis B presence, and hepatitis C presences because they showed large proportions of missing values for GIDEON patients. In particular, vascular invasion was found to be a statistically significant factor for mortality in King et al 2016[5]. The DSU notes that the population of GIDEON did not include any patients from the UK and that only one third of patients were European, which may be relevant in terms of the etiology and prognosis of patients and raises further questions over its suitability as a basis for decision-making. These limitations,

together with the potential for unknown confounders that may be present in non-randomised studies, lead the DSU to recommend against the use of estimates from the matched GIDEON population directly in the analyses.

The company reported that the unrestricted mean treatment duration in the matched population of GIDEON was [redacted] (95% confidence interval [redacted]) months claiming that it followed all patients until treatment discontinuation. The DSU notes that this seems to imply that no censoring happened in any of the followed-up patients. The company referred to external evidence (Table 4 in the company's response to the ACD) to show that treatment duration in SHARP was longer than in clinical practice: it referred to King et al. 2013[6] and King et al. 2016[5], but the former appears to be a preliminary publication of the latter, which is a UK audit on the use of sorafenib for HCC; and, it referred to Ziogas et al.[7] which was also based on data from the NHS. Median treatment duration in these sources was generally lower than that observed in SHARP. However, the median overall survival in these studies was also lower than in the SHARP trial. Figure 1 shows the median treatment duration and overall survivals of the studies included in Table 4 of the company's response to the ACD2. As shown in Figure 1, there seems to be a strong linear correlation between treatment duration and overall survival if we exclude the GIDEON study. The results from the GIDEON study show a low median treatment duration and a high overall survival. For example, patients with Child-Pugh A enrolled in GIDEON had a median OS of 13.6 months compared with 9.5 months in King et al. 2016[5]. Given the time constraint the DSU was unable to properly research the causes of such a difference, but it might be due to differences in populations or clinical practice in the countries were GIDEON was conducted compared with more UK or European studies. Being UK-based studies, King et al. 2016[5] and Ziogas et al.[7] are potentially more representative to UK clinical practice than GIDEON. In light of these observations, treatment duration as observed in SHARP retains plausibility, given the better prognosis of patients enrolled in SHARP.

Figure 1: Treatment duration versus overall survival of sorafenib studies referred to by the company

The company furthermore argued that the mean treatment dose used in the analyses should also be based on the matched GIDEON population ([redacted]). This is considerably lower than the mean of 710.5mg calculated based on the SHARP trial. The DSU notes that it is not clear why the dose in SHARP was higher than in the matched GIDEON population but that it might be because patients in trials are less likely to reduce treatment dose due to mild adverse events or toxicity than in clinical practice. The DSU remains unconvinced of the company's claim that a lower mean dose does not affect the efficacy of sorafenib based on the better median overall survival in the matched GIDEON compared with SHARP. The mean dose reported by King et al. 2016 report is lower than that in SHARP (590mg) but the median overall survival seems to be lower too (Child-Pugh A subgroup had a median OS of 9.5 months compared with the 10.7 months in SHARP, where 95% were Child-Pugh A).

Most importantly, the DSU recalls the point raised in its critique to the company's response to the ACD that treatment effectiveness and cost data should both come from the same source. The DSU also notes that the AC concluded that treatment effectiveness and cost data should be based in SHARP. In conclusion, the DSU believes that using the treatment cost data (mean treatment duration and mean daily dose) from the matched GIDEON population introduces high uncertainty in the analysis and could be potentially misleading.

1.2. EXTRAPOLATION OF TREATMENT DURATION

The DSU, as stated in the Technical Support Document (TSD) 14[8], believes that measures of statistical fit should not be used in isolation to favour one parametric curve over others. In this case, the DSU believes that the external evidence and visual inspection work in favour of the Weibull distribution. However, the DSU notes that there is high uncertainty around which curve is the most appropriate and therefore considers that analyses with both curves should be undertaken. In this case, the DSU slightly favours analyses were the same parametric curve (the lognormal or the Weibull) is used for OS, TTP and duration of treatment.

2. SUMMARY OF THE ANALYSES PRESENTED BY THE COMPANY

The company undertook new analyses using the unrestricted mean duration of treatment and mean daily dose of sorafenib calculated from the matched GIDEON population, as well as using the treatment duration estimated using the Weibull or Gompertz curves fitted to data from SHARP. The company believed that the AC's most likely ICER was best reflected by the weighted average of the ICERs resulting from attaching a weight of 75% to the ICER calculated using a lognormal to extrapolate OS and a weight of 25% to the ICER calculated using a Weibull.. As shown in Table 1, the ICERs of sorafenib versus best supportive care (BSC) using the alternative sources for treatment duration and mean daily dose and based on the described weighted average ranged from £38,825 to £49,241 per QALY gained.

Table 1: Summary of the results of the analyses presented by the company in terms of ICERs (£ per QALY) of sorafenib versus BSC (deterministic), including revised PAS price

Duration of treatment (mean)		GIDEON matched ([redacted])		Gompertz/Weibull from SHARP: ([redacted])	
Daily dose (mean)		GIDEON matched ([redacted])	SHARP (710.5mg)	GIDEON matched ([redacted])	SHARP (710.5mg)
OS	Lognormal	£32,819	£36,050	£37,202	£41,073
	Midpoint (lognormal-Weibull)	£43,874	£48,670	£50,380	£56,125
	75% lognormal, 25% Weibull	**£38,347**	**£42,360**	**£43,791**	**£48,599**

OS: Overall survival

3. ADDITIONAL ANALYSES UNDERTAKEN BY THE DSU

The DSU only updated the main analyses included in its critique to company's response to the ACD with the revised PAS price. These results are shown in Table 2 with the AC's preferred assumptions in bold. Table 3 shows midpoint and weighted average results of the scenarios in Table 2.

Table 2: Results of DSU exploratory analyses with old and new PAS prices (AC's preferred assumptions in bold)

Overall survival	DoT		Total QALYs	Inc. QALYs	Total costs	Inc. costs	ICER
Lognormal	**Log normal**	**BSC**	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	**£48,657**
Weibull		**BSC**	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	**£85,827**
Lognormal	Weibull	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£41,417
Weibull		BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£71,575

DoT: Duration of treatment, BSC: best supportive care

Table 3: Midpoint and weighted averaged results for sorafenib vs BSC with different parametric curves for OS

Overall survival	DoT		Total QALYs	Inc. QALYs	Total costs	Inc. costs	ICER
50% lognormal 50% Weibull	Log normal	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£61,179
	Weibull	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£51,575
75% lognormal 25% Weibull	Log normal	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£54,040
	Weibull	BSC	[redacted]		[redacted]		
		Sorafenib	[redacted]	[redacted]	[redacted]	[redacted]	£45,782

DoT: Duration of treatment, BSC: best supportive care

4. CONCLUSIONS

In response to the ACD2, the company proposed a revised PAS price representing a discount of [redacted] off the previous reduced price and submitted new analyses using estimates based on the matched GIDEON population for treatment duration and mean daily sorafenib dose. The DSU notes that it is preferable to use treatment effectiveness and cost evidence from the same source and favours the data collected during the pivotal trial, SHARP, as established by the AC in the last meeting. The company also argued in favour of using a Weibull to extrapolate time to treatment discontinuation. The DSU agrees with the company that a parametric curve should not be favoured based only on measures of goodness of statistical fit. However, the DSU disagrees with the company's exclusive use of the Weibull, especially in combination with the lognormal to extrapolate OS and TTP. In conclusion, the DSU believes that the most likely ICER of sorafenib compared with BSC based on the revised PAS price lies between £48,657 and £71,575 per QALY, and that it is likely to be closer to the lower end of the range.

5. REFERENCES

1. National Institute for Health and Care Excellence, *Sorafenib for treating advanced hepatocellular carcinoma. Cancer Drugs Fund reconsideration of TA189. Appraisal consultation document 2*. 2017.
2. Bayer CSR, *GIDEON: Global Investigation of therapeutic DEcisions in hepatocellular carcinoma and Of its treatment with sorafeNib (CONFIDENTIAL).* 2013.
3. National Institute for Health and Care Excellence, *Sorafenib for treating advanced hepatocellular carcinoma. Cancer Drugs Fund reconsideration of TA189. Appraisal consultation document.* 2016.
4. Llovet , J.M., et al., *Sorafenib in Advanced Hepatocellular Carcinoma.* New England Journal of Medicine, 2008. **359**(4): p. 378-390.
5. King, J., et al., *Sorafenib for the Treatment of Advanced Hepatocellular Cancer: a UK Audit.* Clinical Oncology, 2016.
6. King, J., et al., *Outcomes for UK patients with advanced hepatocellular carcinoma*, in *NCRI Cancer Conference*. 2013: Liverpool.
7. Ziogas, D.C., et al., *Efficacy and safety of sorafenib in patients with advanced hepatocellular carcinoma: age is not a problem.* European Journal of Gastroenterology & Hepatology, 2017. **29**(1).
8. Latimer, N., *NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data*. 2011.

Commercial Access Agreement

Sorafenib for advanced hepatocellular carcinoma (review of TA189)

The contents of this document have been redacted as they are confidential