Confidential until published

Atezolizumab with bevacizumab for untreated hepatocellular carcinoma. A Single Technology Appraisal

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Matt Stevenson, Professor of Health Technology Assessment, ScHARR, University of Sheffield, Sheffield, UK Emma Simpson, Senior Research Fellow, ScHARR, University of Sheffield, Sheffield, UK John Stevens, Reader in Decision Science, ScHARR, University of Sheffield, Sheffield, UK Andrew Metry, Research Associate, ScHARR, University of Sheffield, Sheffield, UK Geoff Holmes, Research Associate, ScHARR, University of Sheffield, Sheffield, UK Ruth Wong, Information Specialist, ScHARR, University of Sheffield, Sheffield, UK
Correspondence Author	Matt Stevenson, Professor of Health Technology Assessment, ScHARR, University of Sheffield, Sheffield, UK
Date completed	Date completed (23/06/2020)

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 13/14/17.

Declared competing interests of the authors

None of the authors has any conflicts of interest to declare.

Acknowledgements

We would like to thank Aileen Marshall for clinical advice relating to this project. We would also like to thank Paul Tappenden, ScHARR, for providing comments on the draft report and Andrea Shippam, Programme Manager, ScHARR, for providing administrative support and in preparing and formatting the report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Stevenson M, Simpson E, Stevens J, Metry A, Holmes G, Wong R. Atezolizumab with bevacizumab for untreated hepatocellular carcinoma: A Single Technology Appraisal. School of Health and Related Research (ScHARR), 2020.

Contributions of authors

Ruth Wong critiqued the company's search strategy. Emma Simpson summarised and critiqued the clinical effectiveness data reported within the company's submission. John Stevens and Geoff Holmes critiqued the statistical aspects of the submission. Matt Stevenson and Andrew Metry critiqued the health economic analysis submitted by the company. All authors were involved in drafting and commenting on the final report.

CONTENTS

List of tables

List of figures

List of Boxes

Abbreviations

A+B	Atezolizumab in combination with bevacizumab
AEs	Adverse events
AFP	Alpha-fetoprotein
AIC	Akaike Information Criterion
BCLC	Barcelona Clinic Liver Cancer
BIC	Bayesian Information Criterion
BID	Twice-daily
CCOD	Clinical cut-off date
CEAC	Cost-effectiveness acceptability curve
CI	Confidence interval
CrI	Credible interval
CS	Company's submission
CSR	Clinical Study Report
DOR	Duration of response
ECOG	Eastern Cooperative Oncology Group
ECOG PS	Eastern Cooperative Oncology Group Performance Status
EORTC-QLQ	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire
EQ-5D-3L	EuroQol 5 dimensions 3 level
EQ-5D-5L	EuroQol 5 dimensions 5 level
ERG	Evidence Review Group
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HR	Hazard ratio
HRQoL	Health-related quality of life
ICER	Incremental cost-effectiveness ratio
ITC	Indirect treatment comparison
ITT	Intention to treat
IV	Intravenous
IVRS/IWRS	Interactive voice/web response system
KM	Kaplan-Meier
mRECIST	Modified Response Evaluation Criteria In Solid Tumours
MRU	Medical resource use

NCI-CTCAE	National Cancer Institute – Common Terminology Criteria for Adverse Events
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NR	Not reported
ORR	Objective response rate
OS	Overall survival
PAS	Patient Access Scheme
PD	Progressed disease
PF	Progression-free
PFS	Progression-free survival
PRO	Patient-reported outcomes
PSM	Partitioned survival model
QA	Quality assessment
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
RECIST	Response Evaluation Criteria In Solid Tumours
SAE	Serious adverse event
SLR	Systematic literature review
STA	Single Technology Appraisal
TACE	Transarterial chemoembolisation
TKI	Tyrosine kinase inhibitor
TTD	Time to treatment discontinuation
TTP	Time to progression

1 SUMMARY

1.1 Critique of the decision problem in the company's submission

The NICE scope details the population to be adults with locally advanced or metastatic and/or unresectable hepatocellular carcinoma who have had no previous systemic treatment. The intervention is atezolizumab and bevacizumab (hereafter referred to as "A+B") with comparators being sorafenib, lenvatinib and best supportive care (BSC). The company provided an appropriate description of hepatocellular carcinoma (HCC). Following the clarification process, the company provided an appropriate overview of current practice guidelines regarding lines of treatment and the potential positioning of A+B in the treatment pathway, which is the current recommended position for both sorafenib and lenvatinib. The company did not include BSC in the decision problem as it argued that if A+B could be tolerated then so would either sorafenib or lenvatinib. Clinical advice provided to the ERG supported this view.

1.2 Summary of clinical effectiveness evidence submitted by the company

The key evidence of the clinical effectiveness of A+B was derived from one randomised controlled trial (RCT), IMbrave150. Safety data were available from IMbrave150 and the Phase 1b study GO30140.

IMbrave150 randomised adults with locally advanced or metastatic and/or unresectable HCC, who had no previous systemic treatment for HCC, to A+B (atezolizumab 1200 mg IV infusions every three weeks, and bevacizumab 15 mg/kg every three weeks, n=336) or sorafenib (400 mg orally twice per day n=165).

OS was statistically significantly higher for A+B, than for sorafenib HR (stratified) 0.58 (95% CI 0.42, 0.79) p=0.0006. Median OS for A+B was not estimable (NE), median OS for sorafenib was 13.2 months (95% confidence interval [CI] 10.4, NE). There was a statistically significant treatment group difference for PFS HR (stratified) 0.59 (95% CI 0.47, 0.76) p<0.0001. Median PFS was 6.8 months (95% CI 5.6, 8.3) in the A+B group, and 4.3 months (95% CI 4.0, 5.6) for the sorafenib group.

The most common National Cancer Institute – Common Terminology Criteria for Adverse Events (NCI-CTCAE) Grade 3 or 4 AEs experienced in the A+B group were hypertension (10.3%), aspartate aminotransferase increased (4.3%) and proteinuria (2.7%). The most common Grade 3 or 4 AEs in the sorafenib group were hypertension (9.0%), palmar-plantar erythrodysaesthesia syndrome (8.3%), diarrhoea (3.8%), decreased appetite (3.8%), hypophosphataemia (3.2%), and fatigue (3.2%).

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

The ERG believes that the only RCT with available data informing on the clinical effectiveness of A+B in adults with previously untreated locally advanced or metastatic and/or unresectable HCC was included in the company submission. The company's submission (CS) study selection criteria for the review were consistent with the decision problem defined in the final NICE scope. Although the CS study selection criteria for comparators were broader than the decision problem, this allowed inclusion of the relevant comparators, and best supportive care was excluded as deemed appropriate by the ERG's clinical advisor.

The quality of the IMbrave150 RCT was assessed using well-established and recognised criteria. IMbrave150 was an open label trial, but was of otherwise good methodological quality. A literature review of A+B and global comparators identified 59 studies of which 23 connected to provide an evidence network. One of the comparator studies was directly relevant to the decision problem, REFLECT, an open label RCT of otherwise good methodological quality, that compared sorafenib and lenvatinib.

1.4 Summary of cost effectiveness submitted evidence by the company

Following the clarification process, the ERG believes the company's model to be generally well programmed and free from major errors. The company submitted a partitioned survival model comprising three health states (progression-free, post progression, and death). Movements between health states were inferred via PFS and OS models fitted to data from IMbrave150 for A+B and sorafenib, with an indirect treatment comparison performed to inform an HR for lenvatinib versus A+B.

Health-related quality of life data (HRQoL) were collected using the EuroQol 5 dimensions 5 level (EQ-5D-5L) questionnaire within IMbrave150 and mapped to the 3L version using a published algorithm. The time horizon in the base case was 20 years, with discounting of both benefits and costs at 3.5% per annum. The company's base case results suggested that A+B compared with sorafenib had a probabilistic incremental cost-effectiveness ratio (ICER) £22,419 per QALY gained, whilst A+B dominated lenvatinib (i.e. provided more QALYs at a lower cost).

However, as recommended by NICE, confidential Patient Access Schemes (PAS) for sorafenib, lenvatinib and regorafenib were not included in the company's analyses; results incorporating these PASs are provided in a confidential appendix.

The company made the case that A+B met NICE's end of life criteria with patients receiving sorafenib estimated to live for 1.50 years, and those receiving lenvatinib estimated to live for 1.54 years. Those receiving A+B were expected to live for [redacted] years.

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

The ERG identified seven limitations within the company's model and reporting of results. These were: i) identification of perceived modelling errors; ii) extrapolation of time-to-event data, in particular the use of an exponential model for overall survival; iii) the assumptions related to the dosage and acquisition costs of each treatment, in using planned dosages rather than actual dosages; iv) the use of utility values for patients with unresectable HCC which are on average higher than those for the general population; v) overestimation of the adverse events associated with lenvatinib; vi) underestimation of the relative efficacy of lenvatinib, and vii) uncertainty relating to subsequent treatments in IMbrave150 that are not recommended in England. The ERG explored the impact of amending some of these limitations; using the list prices of sorafenib, lenvatinib and regorafenib, these only had a moderate impact on the ICER for atezolizumab. In addition, the ERG conducted subgroup analyses as the acquisition price of lenvatinib is dependent on whether a patient weighs under 60kg or not, explored the impact on the ICER of excluding Asian patients (bar Japanese patients) from the analyses, and undertook analyses exploring the impact on the ICER of uncertainty in the acquisition costs of sorafenib and lenvatinib associated with reduced dose intensity (RDI).

1.6 ERG commentary on the robustness of evidence submitted by the company

1.6.1 Strengths

The search for A+B studies was comprehensive and the ERG believes that no relevant RCTs with relevant data for A+B were excluded from the company's review.

The one included RCT of A+B, IMbrave150, was of good methodological quality, apart from its use of an open-label design. IMbrave150 had an active comparator (not placebo).

According to clinical advice, prior treatments used in IMbrave150 were broadly reflective of UK practice, although prior radiotherapy is rare in the UK, whereas 10% of trial patients received prior radiotherapy.

According to the CS, subsequent treatments used in IMbrave150 were not reflective of UK practice but were unlikely to influence results to a great extent. Subsequent treatment in the UK would include regorafenib, or, where possible, interventions being assessed in ongoing trials.

According to clinical advice, the baseline characteristics of the IMbrave150 trial population were broadly representative of the UK population eligible for A+B treatment, although there are a smaller proportion of Asian patients, and patients with an aetiology of hepatitis B in the UK than is represented in the trial. In addition, more patients in the UK would have aetiology of alcohol or non-alcohol related fatty liver disease, than in the study. Baseline characteristics of REFLECT and IMbrave150 were similar although REFLECT had a higher proportion of patients from the Asia-Pacific region, and implied lower alpha-fetoprotein than IMbrave150.

The implementation of the submitted mathematical model was of good quality. The company responded well to the clarification questions raised and provided a revised model and undertook the analyses requested by the ERG.

1.6.2 Weaknesses and areas of uncertainty

Apart from one study providing additional safety data, there was only one trial of A+B. IMbrave150 was open-label, and also permitted the use of subsequent treatments not recommended in England. There were no head-to-head RCTs of A+B compared with lenvatinib.

There were limitations in the company's economic analyses as summarised in Section 1.5

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG modified the company's base case in the model submitted post-clarification to generate an ERG-preferred base case range. Five changes were made in both ERG base case A and ERG base case B. These were: adjusting for perceived modelling errors; use of log-normal distributions to model OS for all treatments; including seven days wastage for oral chemotherapy when discontinuing treatment; capping utilities for people with unresected HCC at the level of the age- and sex-matched general population, and costing subsequent TKIs and nivolumab treatments from IMbrave150, assuming that the resource use for lenvatinib was the same as for sorafenib. For ERG base case A, it was assumed that the RDI for A+B observed in IMbrave150 was used, whilst the RDI for lenvatinib and sorafenib was assumed to be 1.0. For ERG base case B, it was assumed that the RDI for A+B and sorafenib observed in IMbrave150 was used, whilst the RDI for lenvatinib was assumed based on the REFLECT study.

For the full IMbrave150 population assuming costs for patients weighing under 60kg, the probabilistic ICER range (ERG base case A to ERG base case B) for A+B was £16,567 to £21,843 per QALY gained when compared with sorafenib and £83 to £3,962 per QALY gained when compared with lenvatinib. Assuming costs for patients weighing 60kg or more these ranges were £21,427 to £26,653, and A+B dominant to A+B dominant, respectively.

When patients from Asia (bar Japanese patients) were excluded from the analysis, assuming costs for patients weighing under 60kg, the probabilistic ICER range for A+B was £15,387 to £21,488 per QALY gained when compared with sorafenib and A+B dominant to £3381 per QALY gained when compared with lenvatinib. Assuming costs for patients weighing 60kg or more these ranges were £20,837 to £27,017, and A+B dominant to A+B dominant, respectively.

These results do not incorporate the PAS discounts for sorafenib, lenvatinib and regorafenib. A confidential appendix provides results incorporating these PASs.

2 BACKGROUND

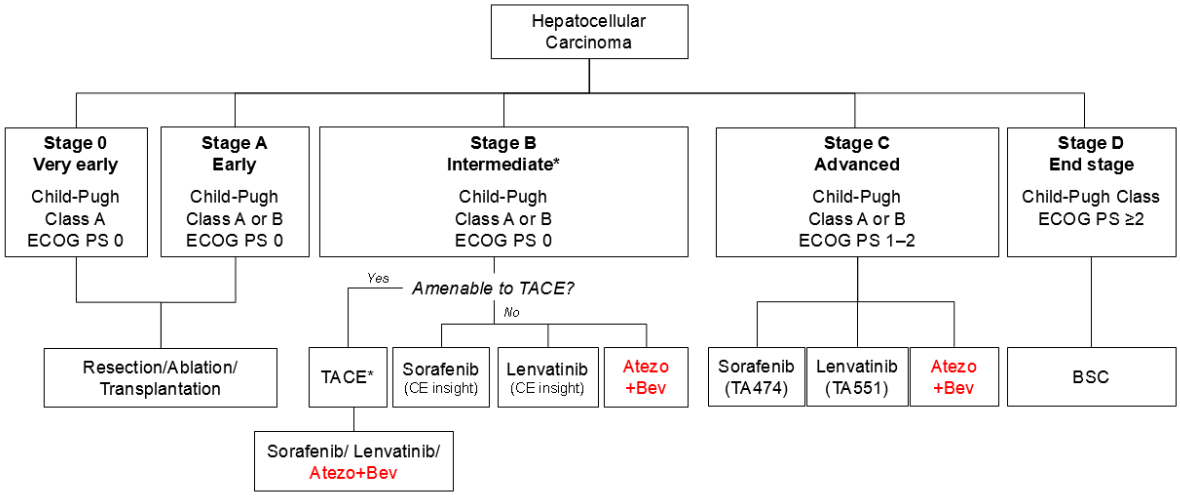
2.1 Critique of company's description of underlying health problem

The CS provide an acceptable description of hepatocellular carcinoma (HCC) in terms of prevalence, symptoms, staging and prognosis.

2.2 Critique of company's overview of current service provision

Following the clarification process, the company revised their diagram depicting the proposed positioning of atezolizumab and bevacizumab (hereafter referred to as "A+B") in the treatment pathway for adults with unresectable HCC (see Figure 1). The revised figure is more aligned to clinical advice provided to the ERG. The company states that "*Due to the highly heterogeneous nature of patients with [Barcelona Clinic Liver Cancer] BCLC Stage B intermediate disease, clinical experts have advised Roche that it is clinically difficult to distinguish those Stage B patients who are not amenable to or who progress on TACE/locoregional therapies from Stage C patients. Therefore, it is not currently possible to separate these populations into distinct groups.*" The positioning of A+B corresponds to the current positioning of sorafenib and of lenvatinib. However, Figure 1 does not show that regorafenib has a positive NICE recommendation for use after sorafenib, but cannot be used after lenvatinib, or A+B, were this to receive a positive recommendation.

Figure 1: Proposed positioning of A+B in treatment pathway for adult patients with unresectable HCC (reproduced from the company's clarification response to question B1)



2.3 Critique of company's definition of the decision problem

2.3.1 Population

The population chosen by the company appears appropriate and in line with NICE's final scope. This is "[redacted]" and covers the full marketing authorisation for A+B for this indication.

2.3.2 Intervention

The intervention is appropriate and matches that in the NICE scope and is A+B.

Atezolizumab (Tecentriq®) is a humanised IgG monoclonal antibody that is administered via an intravenous (IV) infusion of 1200mg every three weeks until the loss of clinical benefit or unmanageable toxicity.

Bevacizumab (Avastin®) is a vascular endothelial growth factor inhibitor which is administered via an IV infusion at a dose of 15mg/kg every three weeks until disease progression or unacceptable toxicity.

2.3.3 Comparators

The comparators included in the CS are sorafenib and lenvatinib. This deviates from the NICE scope which also included best supportive care. The CS states that "*Best supportive care is not a relevant comparator as patients considered eligible for Atezo+Bev would be eligible for alternative active treatment*." (Table 1) Clinical advice to the ERG concurred, stating that if a patient were considered for A+B then they would also be considered for sorafenib or lenvatinib.

2.3.4 Outcomes

The outcomes considered in the CS were consistent those listed in the NICE scope, namely overall survival (OS), progression-free survival (PFS), response rate, adverse effects of treatment and health-related quality of life (HRQoL).

2.3.5 Other relevant factors

Both atezolizumab and bevacizumab have existing Patient Access Schemes (PASs) which take the form of simple price discounts of [redacted]% for atezolizumab and [redacted]% for bevacizumab. Note that the discount for atezolizumab has increased since the CS. All results presented in this report use the new atezolizumab PAS unless explicitly stated.

Sorafenib, lenvatinib and regorafenib have simple PAS discounts. In accordance with NICE process, these price discounts were not considered in the CS, nor in this report. The results of the analyses including the cPAS discounts are provided in a confidential appendix to this ERG report.

3 CLINICAL EFFECTIVENESS

This section provides a structured critique of the clinical evidence submitted by the company to support the cost effectiveness of A+B for untreated locally advanced metastatic HCC patients.

3.1 Critique of the methods of review(s)

3.1.1 Searches

The company performed one clinical effectiveness search to identify all clinical effectiveness and safety studies of A+B and comparators sorafenib or lenvatinib for the first-line treatment of locally advanced metastatic HCC. The company reportedly searched several electronic bibliographic databases in March 2020: MEDLINE [via Ovid], MEDLINE in Process [via Ovid], EMBASE [via Ovid], Cochrane Database of Systematic Reviews [via EBM Reviews], Cochrane Central Register of Controlled Trials [via EBM Reviews], The Health Technology Assessment [via EBM Reviews]), and Database of Abstracts of Reviews of Effects [via EBM Reviews].

The applied search strategy terms, headings, thesauri terms, syntax, recognised RCT filters, limits applied (update search) and concept combinations were correctly applied. A minor suggestion relating to the HCC concept would be to include free-text terms such as "oncolog*" or "adenocarcinoma*" or "sarcoma*" or "adenoma*" in statement 3 of all database strategies. The reasons for the company's inclusion of terms for "chemoembolization" and "radiation therapy" were unclear to the ERG, as they are not relevant comparators designated for this submission (Appendix D, page 12, Table 7: Inclusion and exclusion criteria). The inclusion of these terms would only increase the number of records required to screen and are unlikely to impact the sensitivity of the search for eligible studies. Nevertheless, the database search strategies are fully reported and comprehensive to retrieve all published and eligible studies that are relevant to the review.

The company also searched the International Clinical Trials Registry Platform. Upon ERG request in the clarification letter (page 2, question A1), the company provided search terms used (HCC and hepatocellular carcinoma) and a list of included studies from the trial's registry search. It is unclear to the ERG how many trials were retrieved, screened and excluded by the company. Whilst the search terms applied were broad and should retrieve all eligible trials, the ERG was unable to undertake searches in the ICTRP registry at the time of the company's clarification response.

Supplementary searches by the company include searching several conference abstract websites in the last three years: American Society of Clinical Oncology, European Society for Medical Oncology, American Association for Cancer Research, International Society for Pharmacoeconomics and

Outcomes Research: European Meeting, Health Technology Assessment International, Society for Medical Decision Making (Appendix D, Page 11 of the CS).

The company also searched several HTA websites for previous technology submissions: Scottish Medicines Consortium, All Wales Medicines Strategy Group, Pharmaceutical Benefits Advisory Committee, Canadian Agency for Drugs and Technologies in Health including the pan-Canadian Oncology Drug Review.

3.1.2 Inclusion criteria

The company conducted one systematic review to identify evidence relevant to the scope, and also to populate a network meta-analysis (Appendix D of the CS). The comparators within the inclusion criteria were broader than those in the scope.

Comparators in the inclusion criteria (Appendix D Table 7 of the CS) included not only the comparators in the scope, sorafenib and lenvatinib, but also nivolumab, transarterial chemoembolisation (TACE), radiotherapy, camrelizumab and tislelizumab. Nivolumab would not be used in the UK. Camrelizumab and tislelizumab were included by the CS as they had been tested in Chinese populations, but would not be used in the UK. TACE would be used at an earlier stage in the clinical pathway than is relevant to the decision problem and radiotherapy is rarely used in the UK for HCC. BSC was not included as a comparator in the CS, but was listed in the scope; as discussed in Section 2.3.3, this was deemed appropriate by the clinical advisor.

The population, intervention and outcomes reflected in the inclusion criteria (Appendix D Table 7 of the CS) were consistent with the decision problem set out in the final NICE scope. The population was adults (men or women aged 18+ years) with locally advanced or metastatic and/or unresectable hepatocellular carcinoma (HCC), who had no previous systemic treatment for HCC. The intervention was A+B.

Study selection was conducted by one reviewer and checked by another, as is good practice in systematic reviews (Appendix D.1 of the CS).

3.1.3 Critique of data extraction

Data in the CS were extracted by one reviewer and checked by another, as is good practice in systematic reviews (Appendix D.1 of the CS).

Data in the CS were checked by the ERG against trial publications (Cheng 2019)[1] (Galle 2020)[2] and the IMbrave150 Clinical Study Report (CSR).[3]

3.1.4 Quality assessment

Quality assessment (QA) in the CS was conducted by one reviewer and checked by another (CS Clarification response A2), as is good practice in systematic reviews.

Quality items assessed by the company (CS Section B.2.5) were taken from the Centre for Reviews and Dissemination guidelines for undertaking reviews in health care.[4] Quality assessment of IMbrave150 using Cochrane risk of bias[5] by the company was provided in CS Appendix D.3. Both of these tools are standard and appropriate criteria for assessing the risk of bias in RCTs, and applicable to the IMbrave150 trial. Quality assessment was checked by the ERG against information provided by the company, the CSR,[3] trial protocol[6] and publications (Table 1). [1, 2]

Table 1: IMbrave150 QA by the CS and by the ERG

CRD item	QA by CS (CS Table 7)	Cochrane Risk of bias item	Risk of bias by CS (Appendix D.3)	ERG assessment
Was randomisation carried out appropriately?	Yes	Was the allocation sequence adequately generated?	Low - Randomization was performed via an interactive voice/web response system (IVRS/IWRS) using permuted blocks	Low risk of bias IVRS/IWRS permuted-block randomisation method (protocol)[6]
Was the concealment of treatment allocation adequate?	Yes	Was the concealment of treatment allocation adequate?	Low - Randomization was performed via an interactive voice/web response system using permuted blocks	Low risk of bias Central allocation by IVRS/IWRS [6]
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	NA	NA	Yes (CSR) [3]
Were the care providers, participants and outcome assessors blind to treatment allocation?	N/A (open label study)	Was knowledge of the allocated interventions adequately prevented from participants and personnel? Was knowledge of the allocated interventions adequately	Patients and participants High risk – open label. Outcome assessors Low – although open-label a blinded independent review of imaging for progression-free survival was selected	Patients and participants High risk – open label. Outcome assessors Mixed – high risk for PROs and investigator assessed outcomes Low risk for IRF assessed outcomes: Progression-Free

		prevented from outcome assessors?	for the co-primary endpoint	Survival; Objective Response; Time to Progression; Duration of response (protocol)[6]
Were there any unexpected imbalances in drop-outs between groups?	No	NA	NA	No (CSR)[3]
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	Are reports of the study free of suggestion of selective outcome reporting?	Low - The study reports outcomes of interest as specified	Not assessable until study is published. All outcomes relevant to the decision problem were provided in the CS
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Were incomplete outcome data adequately addressed?	Low - Primary and secondary endpoints reported	ITT analyses provided for primary endpoints AEs - safety evaluable population, as is appropriate
NA	NA	Was the study apparently free of other problems that could put it at a high risk of bias?	Low - The study appears to be free of other sources of bias	Risk of bias from funding source: F. Hoffmann-La Roche Ltd. (CSR) [3]

NA=not applicable

Randomised sequence generation and allocation concealment were conducted by interactive voice or web response technology (CS Appendix D.3 and study protocol[6]) giving a low risk of selection bias. IMbrave150 randomisation was stratified according to: geographic region (Asia excluding Japan vs. rest of world); macrovascular invasion and/or extrahepatic spread (presence vs. absence); Baseline AFP (<400 vs. ≥ 400 ng/mL); ECOG Performance Status (ECOG PS) (0 vs. 1) (study protocol).[6] Effectiveness analyses used the first three of these factors in stratified analyses (CS Section B.2.4), but not ECOG PS (CSR).[3] According to the CSR, *"ECOG PS was removed from the stratified analysis to avoid the potential risk of over-stratification"* (CSR).[3]

There was also a low risk of bias in respect of balance between groups as baseline characteristics appeared similar, and there were no unexpected imbalances in drop-outs between groups (CS Appendix D.3 and CSR[3]).

For effectiveness measures, an intention-to-treat analysis was presented for the primary outcomes of OS and PFS (CS Appendix D.3 and CSR[3]).

For secondary effectiveness outcomes, ITT or modified ITT analyses were employed. For overall response rates only patients with measurable disease were included. For PROs time-to-deterioration ITT analyses were provided. For the PRO measures of proportion of patients with clinically meaningful deterioration, patients required baseline and at least one follow-up measurement. However, patients were analysed within allocated groups in accordance with the principle of ITT (CS Section B.2.6.2) (CSR)[3]).

The IMbrave150 trial was open-label. Lack of blinding can lead to a high risk of performance and detection bias. Patient-reported outcome measures are more likely to be biased than objective measures such as overall survival.[4] Blinded outcome assessment by Independent Review Facility (IRF) was conducted for the measures of progression-free survival; objective response rate; time-to-progression; and duration of response (CS Section B.2.3.2 and study protocol[6]) which reduces the risk of detection bias. Given differences between the intervention and comparator in administration, blinding would require a double-dummy trial design. This would reduce bias for objective measures, but would disguise potential benefits to HRQoL resulting from mode of administration.

IMbrave150 is ongoing and therefore final results have not yet been published, so it cannot be assessed if the authors measured more outcomes than they published. However, data (from the clinical cut-off date 29 Aug 2019) for most outcomes of relevance to this review were provided by the company in the CS and accompanying documents. EuroQol 5 dimensions 5 level (EQ-5D-5L) questionnaire data were not provided; however, the company provided the ERG with mean utilities at each visit, in the model, using data derived from EQ-5D-5L data.

3.2 Critique of trial of the technology of interest

The search conducted in the CS, which sought more comparators than included in the scope (ERG report Section 4.1.2), identified 59 studies (CS Appendix D.1). Of these, one RCT of A+B, IMbrave150, met the inclusion criteria for the decision problem. The ERG does not believe that any relevant published RCTs of A+B that could have provided data have been omitted.

The clinical effectiveness evidence for A+B is based on one RCT, IMbrave150. IMbrave150 was ongoing at the time of writing, with the final OS analysis expected to occur June 2022 (CS Section

B.2.11). At time of writing, data were available from the final PFS analysis and the first interim OS analysis (clinical cut-off date 29 Aug 2019) (CS Section B.2.6).

Additional AE data were provided from the Phase Ib GO30140 study (CS Appendix F).

Other ongoing studies

The CS conducted a broad search for ongoing studies, and did not identify any studies of A+B in HCC. It did identify one RCT (NCT03755791) investigating combination therapy of atezolizumab and cabozantinib, compared with cabozantinib, and with sorafenib, with an estimated primary completion date of August 2020 (CS Clarification response A1). The CS search also identified four ongoing studies of sorafenib (NCT04000737, NCT03412773, NCT03794440, NCT03298451) (CS Clarification response A1). Of these, one RCT, with comparators durvalumab + tremelimumab and durvalumab monotherapy, had an estimated primary completion date of June 2020. Of the other three studies, the earliest estimated primary completion date was June 2021. Three ongoing studies of lenvatinib were identified (NCT04246177, NCT03713593, NCT03905967) with the earliest estimated primary completion date of May 2022 (CS Clarification response A1).

3.2.1 IMbrave150

IMbrave150 is a multicentre, international open-label RCT (CS Section B.2) with centres in Asia, Australia, Europe, and North America (Table 2). It includes 13 patients at n=4 centres in the UK.

Table 2: IMbrave150 study references

Study name	Study design	Published References	Other references provided by CS
IMbrave150 150 NCT03434379 YO40245	Phase III, open-label, multicentre, international, RCT	Protocol on clinical trials registry NCT03434379: A Study of Atezolizumab in Combination With Bevacizumab Compared With Sorafenib in Patients With Untreated Locally Advanced or Metastatic Hepatocellular Carcinoma (IMbrave150) https://clinicaltrials.gov/ct2/show/NCT03434379[7] Abstract of primary analysis results Cheng et al. (2019) European Society of Medical Oncology ASIA. Atezolizumab + bevacizumab vs sorafenib in patients with unresectable hepatocellular carcinoma: Phase 3 results from IMbrave150.[1] Abstract of patient-reported outcomes Galle et al. (2020) Patient-reported Outcomes from the Phase 3 IMbrave150 Trial of Atezolizumab + Bevacizumab Versus Sorafenib as First-line Treatment for Patients with Unresectable Hepatocellular Carcinoma. American Society of Clinical Oncology GastroIntestinal[2]	Study protocol Hoffmann La-Roche Ltd. (2019) IMbrave150 Study Protocol.[6] Clinical study report Hoffmann La-Roche Ltd. (2019) IMbrave150 Clinical Study Report (29 Aug 19 Clinical cut-off date (CCOD)). Report No.: 1092943.[3]

Patients were randomised to receive A+B or sorafenib until investigator-assessed unacceptable toxicity or loss of clinical benefit (CS Section B.2) (Table 3). Patients, in either trial arm, with disease progression were allowed to continue treatment if there was investigator-determined clinical benefit and absences of: symptoms and signs indicating unequivocal progression of disease; and decline in ECOG PS; and tumour progression at critical anatomical sites that cannot be managed by protocol-allowed medical interventions (CS Section B.2).

Dose modification or interruption was allowed for sorafenib (CS Section B.2) to allow management of toxicity. Dose modification was not allowed for A+B (CS Section B.2). However, dose interruption was allowed (CS Section B.2.10) to allow recovery from toxicity (CSR).[3]

The IMbrave150 trial allowed concomitant treatment with: oral contraceptives; hormone-replacement therapy; inactivated influenza vaccines; megestrol acetate administered as an appetite stimulant; mineralocorticoids; corticosteroids; low-dose aspirin; prophylactic use of low-dose anticoagulation, unfractionated heparin or low molecular weight heparin; palliative radiotherapy; radiotherapy to the brain; other local therapy (surgery, stereotactic radiosurgery, radiofrequency ablation) (CS Section B.2.3.2).

Table 3: IMbrave150 study characteristics (CS section B.2)

Study	Population	Intervention (n randomised)	Comparator (n randomised)	Primary outcomes
IMbrave150 150 NCT03434379 YO40245	Adults with locally advanced or metastatic and/or unresectable HCC who had not received prior systemic treatment	Combination atezolizumab 1200 mg IV infusions, Q3W plus bevacizumab 15 mg/kg IV Q3W (n=336)	sorafenib 400 mg oral, BID, continuously (n=165)	PFS: time from randomisation to the first documented disease progression as determined by an IRF according to response evaluation criteria in solid tumours (RECIST) Version 1.1 or death from any cause, whichever occurred first OS: time from randomisation to death due to any cause

IV=intravenous; Q3W= every three weeks; BID=twice a day

Eligibility criteria are provided in CS Section B.2.3.2. Included patients were adults with locally advanced or metastatic and/or unresectable HCC, with no previous systemic treatment (Table 4). Diagnosis was confirmed by histology/cytology or clinically by American Association for the Study of Liver Diseases criteria in cirrhotic patients.[8] Patients were also required to have at least one measurable (per RECIST v1.1) untreated lesion, and be scored as Child-Pugh class A and ECOG PS 0 or 1 within 7 days prior to randomisation.

Table 4: IMbrave150 eligibility criteria (reproduced from CS Section B.2.3.2 Summary of study methodology)

Inclusion criteria	**Exclusion criteria**
• Locally advanced or metastatic and/or unresectable HCC with diagnosis confirmed by histology/cytology or clinically by American Association for the Study of Liver Diseases criteria in cirrhotic patients • Disease that was not amenable to curative surgical and/or locoregional therapies, or progressive disease after surgical and/or locoregional therapies • No prior systemic therapy (including systemic investigational agents) for HCC • Patients who received prior local therapy (e.g., radiofrequency ablation, percutaneous ethanol or acetic acid injection, cryoablation, high-intensity focused ultrasound, transarterial	• History of malignancy other than HCC within 5 years prior to screening, with the exception of malignancies with a negligible risk of metastasis or death (e.g., 5-year OS rate >90%), such as adequately treated carcinoma in situ of the cervix, non-melanoma skin carcinoma, localised prostate cancer, ductal carcinoma in situ, or Stage I uterine cancer • Known fibrolamellar HCC, sarcomatoid HCC, or mixed cholangiocarcinoma and HCC • Moderate or severe ascites • History of hepatic encephalopathy • Co-infection of HBV and HCV Patients with a history of HCV infection who were negative for HCV RNA by PCR were considered non-infected with HCV.

chemoembolisation, transarterial embolisation, etc.) were eligible provided the target lesion(s) had not been previously treated with local therapy or the target lesion(s) within the field of local therapy had subsequently progressed in accordance with RECIST v1.1.
• Child-Pugh class A within 7 days prior to randomisation
• Adequate haematologic and end organ function within 7 days prior to randomisation
  o Serum bilirubin ≤3x ULN
  o Serum albumin ≥28 g/L (2.8 g/dL) without transfusion
  o For patients not receiving therapeutic anticoagulation: INR or aPTT ≤2x ULN
• Documented virology status of hepatitis, as confirmed by screening hepatitis B virus (HBV) and hepatitis C virus (HCV) serology test
• At least one measurable (per RECIST v1.1) untreated lesion
• ECOG PS of 0 or 1 within 7 days prior to randomisation
• For patients with active HBV:
  o HBV DNA <500 IU/mL obtained within 28 days prior to initiation of study treatment, and
  o Anti-HBV treatment (per local standard of care; e.g., entecavir) for a minimum of 14 days prior to study entry and willingness to continue treatment for the length of the study
	• Untreated or incompletely treated oesophageal and/or gastric varices with bleeding or high risk for bleeding
	
	Patients must undergo an esophagogastroduodenoscopy (EGD), and all size of varices (small to large) must be assessed and treated per local standard of care prior to enrolment. Patients who have undergone an EGD within 6 months prior to initiation of study treatment do not need to repeat the procedure
	• A prior bleeding event due to oesophageal and/or gastric varices within 6 months prior to initiation of study treatment
	Exclusion criteria related to medications
	Patients who met any of the following criteria were excluded from study entry:
	• Prior allogeneic stem cell or solid organ transplantation
	• History of severe allergic anaphylactic reactions to chimeric or humanized antibodies or fusion proteins
	• Known hypersensitivity to Chinese hamster ovary cell products or to any component of the atezolizumab or bevacizumab formulation
	• Treatment with strong CYP3A4 inducers within 14 days prior to initiation of study treatment, including rifampin (and its analogues) or St. John's wort
	• Treatment with any agent that may interfere with the immunostimulatory nature of atezolizumab
	All patients had to meet several bevacizumab-specific criteria based on the known safety profile of this drug. These criteria excluded patients with evidence of or a possibility for bleeding issues, uncontrolled hypertension, and/or gastrointestinal perforations

The primary outcomes of IMbrave150 were OS and PFS (Table 5). Other outcomes were objective response rate (ORR), Duration of Response (DOR), Time to Progression (TTP), safety and HRQoL. Pharmacokinetic outcomes in A+B group were measured, but are not considered in this ERG report (CSR Table 10). Patients underwent tumour assessments at baseline, then every 6 weeks (+/-1 week) for the first 54 weeks following treatment initiation, and every 9 weeks (+/-1 week) thereafter until radiographic disease progression per RECIST v1.1 or (for patients who continue treatment after radiographic disease progression) loss of clinical benefit as determined by the investigator (CS Clarification response A18).

Table 5: IMbrave150 outcome definitions (CS Section B.2.3 and CSR[3])

Outcome	Definition	Measured by
Overall survival (OS)	Time from randomisation to death due to any cause	
Progression-free survival (PFS)	Time from randomisation to the first documented disease progression, or death due to any cause, whichever occurred first	Independent Review Facility (IRF) Response Evaluation Criteria in Solid Tumors (RECIST) Version 1.1[9] IRF Modified Response evaluation criteria in solid tumours (mRECIST)[10] Investigator assessed RECIST v1.1
Objective response rate (ORR)	Complete or partial response	IRF RECIST v1.1 IRF mRECIST Investigator assessed RECIST v1.1
Duration of response (DOR)	Time from the first occurrence of a documented objective response to disease progression or death from any cause (whichever occurs first)	IRF RECIST v1.1 IRF mRECIST Investigator assessed RECIST v1.1
Time to progression (TTP)	Time from randomisation to the first occurrence of disease progression	IRF RECIST v1.1 IRF mRECIST Investigator assessed RECIST v1.1
Safety	Safety and tolerability of atezolizumab administered in combination with bevacizumab compared with sorafenib monotherapy	severity determined according to NCI CTCAE v4.0 (National Cancer Institute Common Terminology Criteria for Adverse Events, Version 4.0) [11]
HRQoL (Time to deterioration)	Time from randomisation to first deterioration (decrease from baseline of ≥10 points), maintained for two consecutive assessments or one assessment followed by death from any cause within 3 weeks: physical functioning; role functioning; and global health status/quality of life	Patient-reported outcomes (PROs) EORTC QLQ-C30 [12] and EORTC QLQ-HCC18 [13] questionnaires

IMbrave150 screened 725 patients, of whom 224 failed to meet eligibility criteria or withdrew consent prior to randomisation (CS Appendix D.2 and (CSR)[3]). Five hundred and one patients were randomised; 336 patients were randomised to A+B and 165 patients were randomised to sorafenib. These formed the ITT population for effectiveness analyses (CS Appendix D.2 and (CSR)[3]). Allocated treatment was not received by 7 patients in the A+B group, and 9 patients in the sorafenib group.

The safety population comprised any patients who received each treatment, regardless of allocated group: n=329 received A+B, and n=156 received sorafenib (CS Appendix D.2 and (CSR)[3]). The PRO-evaluable population comprised patients who had baseline PRO data and at least one other PRO assessment (A+B n=309, sorafenib n=145) (Galle 2020)[2] (CSR).[3]

Patient baseline characteristics were reported as being well balanced between treatment groups (CS Section B.2.3.3 and (CSR).[3]). CS B.2.3.3 Table 5 shows the key baseline demographics and disease characteristics. Median age was 64.0 years in the A+B group, and 66.0 in the sorafenib group (CS Section B.2.3.3). There was an ECOG PS score of 0 for 62.2% of the A+B group, and 62.4 % of the sorafenib group (CS Section B.2.3.3). Extrahepatic spread and/or macrovascular invasion was present in 76.8% of the A+B group, and 72.7% of the sorafenib group (CS Section B.2.3.3). The aetiology of HCC was HBV for 48.8% of the A+B group, and 46.1% of the sorafenib group (CS Section B.2.3.3). TACE had been received by 38.7% of the A+B group, and 42.4% of the sorafenib group (CS Section B.2.3.3).

[redacted]

Following discontinuation of study treatment, 20.5% of the A+B group, and 44.2% of the sorafenib group had subsequent HCC systemic therapy (CS Section B.2.3.3). Most of this was tyrosine kinase inhibitors, 18.8% of the A+B group, and 26.1% of the sorafenib group (CS Section B.2.3.3).

At the time of the clinical cut-off date (29 August 2019), 309 patients were still on study, of whom n=228 (67.9%) in the A+B group, and n=81 (49.1) in the sorafenib group (CS Appendix D.2). The most common reason for discontinuing the study was death (CS Appendix D.2). In the A+B group, n=108 discontinued the study, with death being the reason for n=95 (CS Appendix D.2). In the sorafenib group, n=84 discontinued the study, with death being the reason for n=65 (CS Appendix D.2).

At the time of the clinical cut-off date (29 August 2019), n=146 (43.5%) in the A+B group were still receiving study treatment, as were n=24 (15.4%) in the sorafenib group (Table 6). In the safety evaluable population, median study-treatment duration was 7.4 months for atezolizumab, 6.9 months for bevacizumab, and 2.8 months for sorafenib (CS Section B.2.10) (Cheng 2019[1]).

Table 6: IMbrave150 Reasons for discontinuation from study treatment in safety evaluable population (reproduced from CS Appendix D.2 Table 9 and CS Clarification response A4)

n, (%)	A+B n=329			Sorafenib n=156
	Atezolizumab	Bevacizumab	A+B	
Received at least one study treatment Yes	329 (100)	329 (100)	329 (100)	156 (100)
Treatment status				
Ongoing	146 (44.4)	137 (41.6)	146 (44.4)	24 (15.4)
Withdrawn from treatment	183 (55.6)	192 (58.4)	183 (55.6)	132 (84.6)
Withdrawn from treatment reason				
Death	15 (4.6)	16 (4.9)	15 (4.6)	7 (4.5)
Adverse event	29 (8.8)	49 (14.9)	26 (7.9)	16 (10.3)
Symptomatic deterioration	10 (3.0)	9 (2.7)	9 (2.7)	4 (2.6)
Progressive disease	111 (33.7)	100 (30.4)	98 (29.8)	93 (59.6)
Physician decision	3 (0.9)	4 (1.2)	3 (0.9)	4 (2.6)
Withdrawal by subject	15 (4.6)	14 (4.3)	14 (4.3)	7 (4.5)
Other	0	0	0	1 (0.6)

3.2.2 IMbrave150 effectiveness

At time of writing, data were available for the clinical cut-off date (CCOD) 29th Aug 2019. Median follow-up at CCOD was 8.6 months (CS Section B.2.6) (Cheng 2019).[1]

3.2.2.1 IMbrave150 OS

At time of writing, data were available from the first interim OS analysis (clinical cut-off date 29th Aug 2019) (CS Section B.2.6). Deaths from any cause occurred in n=96 (28.6%) in the A+B group, and n=65 (39.4%) in the sorafenib group.

Median OS was not estimable in the A+B group (Table 7). The Kaplan-Meier (KM) estimated median OS for the sorafenib group was 13.2 months (95% CI 10.4, NE) (CS Section B.2.6). There was a statistically significant advantage in OS for A+B over sorafenib, hazard ratio (HR) (stratified) 0.58 (95% CI 0.42, 0.79) log-rank *p*=0.0006 (CS Section B.2.6)(Cheng *et al.*2019[1])(Galle *et al.*[2]).

Table 7: IMbrave150 Overall Survival, first interim analysis, in the ITT population (adapted from CS Section B.2.6.1 Table 8 and B.2.10 and Cheng 2019[1] and Galle 2020[2])

	A+B n=336	**Sorafenib n=165**
Patients with event, n (%)	96 (28.6)	65 (39.4)
Median time to event, months (95% CI)	NE NE	13.2 (10.4, NE)
Stratified HR (95% CI) log-rank p value	0.58 (0.42, 0.79) *p*=0.0006	
Patients remaining event free at 6 months, % (95% CI)	84.8 (80.9, 88.7)	72.0 (65.1, 79.4)
Patients remaining event free at 12 months, % (95% CI)	67.2 (61.3, 73.1)	54.6 (45.2, 64.0)
Death due to progressive disease (safety evaluable population), n	71	51
Death due to AE (safety evaluable population), n	15	9
Other or unknown cause of death (safety evaluable population), n	7	4
Subgroups		
Geographic region (Asia excluding Japan), n/N	34/133	27/68
HR unstratified (95%CI)	0.53 (0.32,0.87)	
Geographic region (rest of world), n/N	62/230	38/97
HR unstratified (95%CI)	0.65 (0.44, 0.98)	
Macrovascular invasion and/or extrahepatic spread (presence), n/N	84/258	56/120
HR unstratified (95%CI)	0.55 (0.39, 0.77)	
Macrovascular invasion and/or extrahepatic spread (absence) n/N	12/78	9/45
HR unstratified (95%CI)	0.69 (0.29,1.65)	
Baseline AFP (<400) n/N	45/210	36/104
HR unstratified (95%CI)	0.52 (0.34, 0.81)	
Baseline AFP (≥ 400 ng/mL) n/N	51/126	29/61
HR unstratified (95%CI)	0.68 (0.43, 1.08)	
ECOG Performance Status 0, n/N	50/209	31/103
HR unstratified (95%CI)	0.67 (0.43, 1.06)	
ECOG Performance Status 1, n/N	46/127	34/62
HR unstratified (95%CI)	0.51 (0.33, 0.80)	

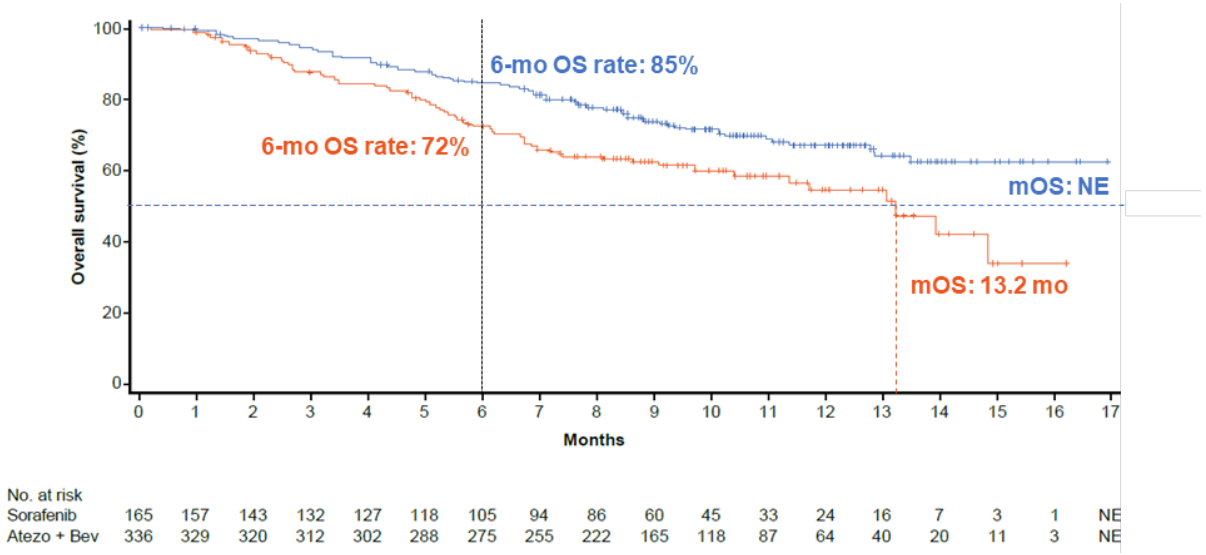
NE= not estimable

The OS event-free rate was higher for A+B than for sorafenib at 6 months (84.8% A+B, 72.0% sorafenib), and at 12 months (67.2% A+B, 54.6% sorafenib) (CS Section B.2.6).

Subgroups were investigated for OS; however, the study was not powered to detect differences in the individual subgroups, and so results should be interpreted with caution (CS Section B.2.7). Across subgroups (Table 7) there was a trend for a survival advantage for A+B over sorafenib (CS Appendix E).

The OS KM survival function within IMbrave150 is shown in Figure 2.

Figure 2: KM survival function for OS (reproduced from Figure 4 of the CS)



3.2.2.2 IMbrave150 PFS

At time of writing, data were available from the final PFS analysis based on IRF-assessment per RECIST v1.1 (the co-primary endpoint) (clinical cut-off date 29th Aug 2019).

Events counted were the first documented disease progression (IRF-assessment per RECIST v1.1), or death due to any cause, whichever occurred first. Events occurred in 197 (58.6%) patients in the A+B group, and n=109 (66.1%) patients in the sorafenib group (Table 8).

The KM estimated median PFS was 6.8 months (95% CI 5.6, 8.3) in the A+B group, and 4.3 months (95%CI 4.0, 5.6) for the sorafenib group (CS Section B.2.6.1).

There was a statistically significant advantage in PFS for A+B over sorafenib, HR (stratified) 0.59 (95% CI 0.47, 0.76) log-rank p<0.0001 (CS Section B.2.6.1).

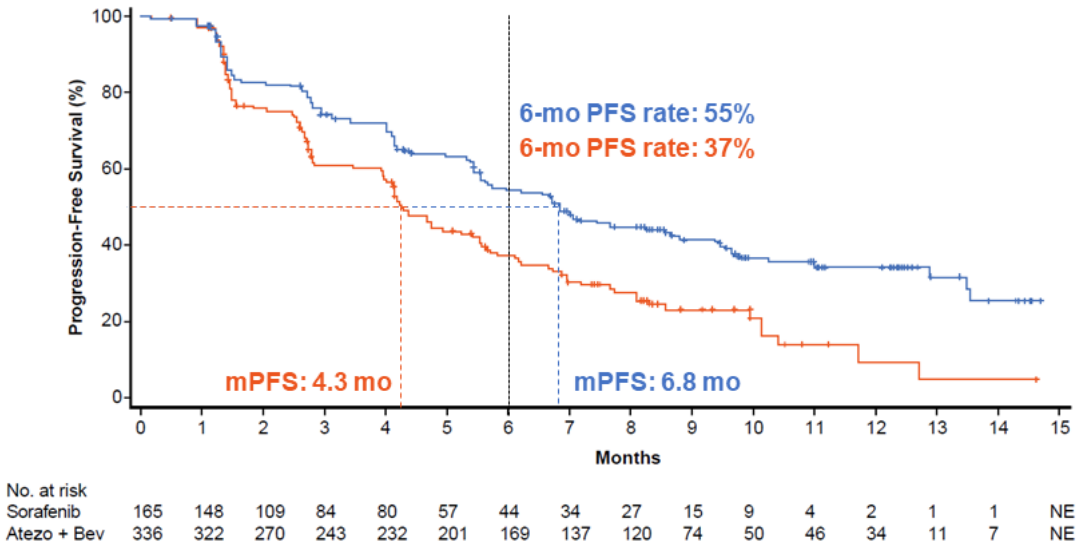
Table 8: IMbrave150 PFS IRF per RECIST v1.1 ITT population (from CS Section B.2.6.1 Table 9 and Cheng *et al.* 2019[1] and Galle *et al.* 2020[2])

	A+B n=336	**Sorafenib n=165**
Patients with event, n (%)	197 (58.6)	109 (66.1)
Earliest contributing event, (%) Death Disease progression	 34 163	 29 80
Median time to event, months (95% CI)	6.8 (5.7, 8.3)	4.3 (4.0, 5.6)
Stratified HR (95% CI) log-rank p value	0.59 (0.47, 0.76) p<0.0001	
Patients remaining event free at 6 months, % (95% CI)	54.5 (49.1, 60.0)	37.2 (29.0, 45.3)
Patients remaining event free at 12 months, % (95% CI)	34.0 (27.9, 40.1)	9.2 (0.0, 18.5)

The PFS event-free rate was higher for A+B than for sorafenib at 6 months (54.5% A+B, 37.2% sorafenib), and at 12 months (34.0% A+B, 9.2% sorafenib) (CS Section B.2.6).

The PFS KM) survival function within IMbrave150 is shown in Figure 3.

Figure 3: KM survival function for PFS (reproduced from Figure 5 of the CS)



For the secondary endpoint of PFS, measured by IRF-assessed HCC mRECIST, there were events in 199/336 (59.2%) patients in the A+B group, and 111/165 (67.3%) patients in the sorafenib group. Median PFS was 6.8 months (95% CI 5.7, 8.3) in the A+B group, and 4.3 months (95% CI 4.0, 5.6) in

the sorafenib group, stratified HR=0.59 (95%CI 0.47, 0.76) $p<0.0001$ (Cheng 2019)[1] (Galle 2020)[2] (CS Section B.2.6.2).

The secondary endpoint of TTP as measured by IRF-assessment RECIST v1.1, estimated median time to progression of 8.6 months (95% CI: 6.8, 9.9) in the A+B group, and 5.6 months (95% CI: 4.2, 7.7) in the sorafenib group. The stratified HR was 0.70 (95% CI 0.53, 0.92) (CS Section B.2.6.2) $p=0.0105$ (CSR).[3]

Subgroups were investigated for PFS measured by IRF-assessed RECIST v1.1; however, the study was not powered to detect differences in the individual subgroups, and so results should be interpreted with caution (CS Section B.2.7). Across subgroups there was a trend for a survival advantage for A+B over sorafenib (CS Appendix E).

3.2.2.3 IMbrave150 response rate

Objective response rate (IRF-assessment per RECIST v1.1) was measured from the ITT population with measurable disease at baseline, A+B n=326, sorafenib n=159 (CS Section B.2.6.2).

Non-measurable lesions were defined according to RECIST v1.1 as *"Non-measurable tumour lesions encompass small lesions (longest diameter <10 mm or pathological lymph nodes with short axis ≥10 mm but <15 mm) as well as truly non-measurable lesions. Lesions considered truly non-measurable include leptomeningeal disease, ascites, pleural or pericardial effusion, inflammatory breast disease, lymphangitic involvement of skin or lung, peritoneal spread, and abdominal mass/abdominal organomegaly identified by physical examination that is not measurable by reproducible imaging techniques"* (CSR).[3]

In the A+B group there were n=89 (27.3%) responders, of whom n=18 were assessed as having a complete response (Table 9). In the sorafenib group, there were n=19 (11.9%) responders. None of the sorafenib group were assessed as having a complete response (CS Section B.2.6.2). (Cheng 2019).[1] There was a statistically significant difference in confirmed ORR favouring A+B 15.4% (95% CI 7.9, 22.8) $p<0.0001$.

ORR measured by IRF-assessed mRECIST was measured from a population of n=325 A+B, and n=158 sorafenib. There was a statistically significant difference in this measure of ORR favouring A+B 19.9% (95% CI 12.1, 27.8) $p<0.0001$, based on a response rate of 33.2% in the A+B group, and 13.3% in the sorafenib group (CS Section B.2.6.2) (Cheng 2019[1]).

DoR was defined as time from the first occurrence of a documented objective response to disease progression or death from any cause (whichever occurs first). DoR based on IRF-assessment per RECIST v1.1 was measured in the responders from ORR (IRF-assessment per RECIST v1.1), n=89 in the A+B group, and n=19 in the sorafenib group. There was a significant difference between groups favouring A+B in DoR, stratified HR=0.23 (95% CI 0.08, 0.70) *p*=0.0051 (Table 9).

Table 9: IMbrave150 Confirmed ORR (and DOR) based on IRF-assessment per RECIST v1.1 (Population with measurable disease at baseline) (adapted from CS Section B.2.6.2 Table 10 and Table 12 and Cheng 2019)[1]

	A+B n=326	**Sorafenib n=159**
ORR		
Responders, n (%) 95% CI	89 (27.3) (22.5, 32.5)	19 (11.9) (7.4, 18.0)
Stratified analysis Difference in ORR, % (95% CI) Odds ratio (95% CI) Cochran-Mantel-Haenszel p value	 15.4 (7.9, 22.8) 2.90 (1.68, 5.01) *p*<0.0001	
Complete response, n (%) 95% CI	18 (5.5) (3.3, 8.6)	0 (0.0, 2.3)
Partial response, n (%) 95% CI	71 (21.8) (17.4, 26.7)	19 (11.9) (7.4, 18.0)
Stable disease, n (%) 95% CI	151 (46.3) (40.8, 51.9)	69 (43.4) (35.6, 51.5)
Progressive disease, n (%) 95% CI	64 (19.6) (15.5, 24.4)	39 (24.5) (18.1, 32.0)
Not evaluable, n (%) Missing, n (%)	8 (2.5) 14 (4.3)	14 (8.8) 18 (11.3)
DoR		
Patients with event, n/N (%)	12/89 (13.5)	6/19 (31.6)
Median time to event, months (95% CI)	NE (NE)	6.3 (4.7, NE)
Stratified HR (95% CI) log-rank p value	0.23 (0.08, 0.70) *p*=0.0051	

3.2.2.4 IMbrave150 HRQoL

Baseline HRQoL data were taken from the PRO-evaluable population, that is, patients who had baseline data and at least one other PRO assessment (A+B n=309; sorafenib n=145) (Galle 2020).[2]

While on study treatment, patients completed the EORTC QLQ-C30 and EORTC QLQ HCC18 questionnaires every 3 weeks, and following treatment discontinuation, every 3 months (Galle 2020).[2]

Questionnaire completion rates were high (≥92%) (Galle 2020)[2] until Cycle 17, which was beyond the time the majority of participants remained on allocated study treatment (CS Section B.2.6.3).

The EORTC QLQ-C30 measures global health/quality of life; patient functioning (measured on aspects of physical, emotional, role, cognitive, and social); symptom scales (fatigue, nausea/vomiting, pain); and single items of (dyspnoea, insomnia, appetite loss, constipation, diarrhoea, and financial difficulties).[12,14] The EORTC QLQ-HCC18 is measured on six symptom scales (fatigue, body image, jaundice, nutrition, fevers, and pain), and two single items (abdominal swelling and sexual interest).[13, 15]

Time to deterioration was measured using the European Organisation for Research and Treatment of Cancer quality of life questionnaire (EORTC-QLQ-C30) (CS Section B.2.3), measured on scales of 0-100. EORTC QLQ-C30 is a validated, reliable self-reported measure (Aaronson et al.1993; Fitzsimmons et al. 1999).[12, 14] Clinically meaningful deterioration was defined as decrease from baseline of at least ten points (Osoba 1998).[16]

There was a longer time to deterioration for A+B over sorafenib in three of the domains measured. Median time to deterioration for patient-reported physical functioning was 13.1 months (95% CI 9.7, not estimable) in the A+B group, and 4.9 months (95% CI 3.5, 6.2) in the sorafenib group, HR (stratified) 0.53 (95% CI 0.39, 0.73) (Galle 2020)[2] (CSR)[3] (CS Section B.2.3).

The median time to deterioration for the role functioning domain was 9.1 months (95% CI 6.5, NE) in the A+B group, and 3.6 months (95% CI 2.2, 6.0) in the sorafenib group, HR (stratified) 0.62 (95% CI 0.46, 0.84) (Galle 2020)[2] (CS Section B.2.3).

Median time to deterioration in the Global Health Status/quality of life domain was 11.2 months (95% CI 6.0, NE) in the A+B group, and 3.6 months (95% CI 3.0, 7.0) in the sorafenib group, HR (stratified) 0.63 (95% CI 0.46, 0.85) (CS Section B.2.3). (Galle 2020)[2]

Exploratory endpoints were from EORTC-QLQ-C30 and EORTC QLQ-HCC18.[13, 15] Time to deterioration in symptoms was delayed for A+B over sorafenib, stratified HRs: Diarrhoea (QLQ-C30) HR=0.23 (95% CI 0.16, 0.34); Pain (QLQ-C30) HR=0.46 (95% CI 0.34, 0.62); Pain (QLQ-HCC18) HR=0.65 (95% CI 0.46, 0.92); Appetite Loss (QLQ-C30) HR=0.57 (95% CI 0.40, 0.81); Fatigue (QLQ-HCC18) HR=0.60 (95% CI 0.45, 0.80); Fatigue (QLQ-C30) HR=0.61 (95% CI 0.46, 0.81); Jaundice (QLQ-HCC18) HR=0.76 (95% CI 0.55, 1.07) (CS Section B.2.6.3) (Galle 2020).[2]

[redacted]

[redacted]

EQ-5D-5L were collected in IMbrave150 directly from the patients. These data were collected at "*each scheduled study visit prior to administration of study drug and prior to any other study assessment(s). During survival follow-up the EQ-5D-5L questionnaire was completed every 3 months (for 1 year) following disease progression or treatment discontinuation, unless the patient withdrew consent, whichever occurred first.*" Details on the use of EQ-5D-5L data are presented in Section 4.2.5.4 of this report.

3.2.3 Adverse events

The CS reference pack included the draft summary of product characteristics (SmPC) for atezolizumab. The following table showing frequency of adverse events with atezolizumab combination therapy is adapted from the draft SmPC (Table 10).

[redacted] **10:**

[redacted]

Adverse events of special interest for atezolizumab (CS Appendix F) are immune-mediated reactions, including immune-mediated hepatitis, hypo/hyperthyroidism, pneumonitis and rash; and autoimmune haemolytic anaemia; and infusion-related reaction. Adverse events of special interest for bevacizumab (CS Appendix F) include bleeding, thromboembolism, congestive heart failure, gastrointestinal (GI)

perforation, fistula/Abscess (non-GI), Posterior Reversible Encephalopathy Syndrome, proteinuria and wound healing complications.

3.2.3.1 IMbrave150 AEs

Table 11: IMbrave150 AE overview (copied from CS Section B.2.10 Table 22)

n, (%)	A+B n=329	Sorafenib n=156
Total number of patients with at least one AE	323 (98.2)	154 (98.7)
Total number of AEs, n	3058	1299
Total number of patients with at least one		
AE related to any study treatment	276 (83.9)	147 (94.2)
AE related to atezolizumab	252 (76.6)	n/a
AE related to bevacizumab	241 (73.3)	n/a
Grade 3–4 AE	186 (56.5)	86 (55.1)
Treatment-related Grade 3–4 AE	117 (35.6)	71 (45.5)
Grade 5 AE	15 (4.6)	9 (5.8)
Treatment-related Grade 5 AE	6 (1.8)	1 (0.6)
Serious AE	125 (38.0)	48 (30.8)
Related serious AE	56 (17.0)	24 (15.4)
AE leading to withdrawal from any study treatment	51 (15.5)	16 (10.3)
AE leading to withdrawal from atezolizumab	28 (8.5)	0
AE leading to withdrawal from bevacizumab	48 (14.6)	0
AE leading to withdrawal from A+B	23 (7.0)	0
AE leading to dose modification/interruption of any study treatment	163 (49.5)	95 (60.9)
AE leading to dose interruption of any study treatment	163 (49.5)	64 941.0)
AE leading to dose reduction of sorafenib	n/a	58 (37.2)

The majority of patients experienced at least one AE of any severity (Table 11), 98.2% of the A+B group, and 98.7% of the sorafenib group (CS Section B.2.10). The most common AEs in the A+B group were hypertension (29.8%), fatigue (20.4%) and proteinuria (20.1%) (CS Section B.2.10). The most common AEs in the sorafenib group were diarrhoea (49.4%), palmar-plantar erythrodysaesthesia (hand-foot) syndrome (48.1%), decreased appetite (24.4%) and hypertension (24.4%) (CS Section B.2.10).

Grade 5 AEs were experienced by n=15 (4.6%) of the A+B group, of which n=6 (1.8%) were considered by the investigator to be related to treatment (CS Section B.2.10). Grade 5 AEs were experienced by n=9 (5.8%) of the sorafenib group, of which n=1 (0.6%) were considered by the investigator to be related to treatment (CS Section B.2.10).

[redacted]

Grade 3/4 AEs were experienced by n=186 (56.5) of the A+B group, of which n=117 (35.6%) were considered by the investigator to be related to treatment (CS Section B.2.10).(Cheng 2019).[1] The most

common Grade 3 or 4 AEs in the A+B group were hypertension (10.3%), aspartate aminotransferase increased (4.3%) and proteinuria (2.7%) (CS Section B.2.10).

Serious AEs (SAEs) were defined as events meeting one of the following criteria: fatal; life-threatening; requiring or prolonging hospitalisation; resulting in persistent or significant disability/incapacity; congenital anomaly/birth defect caused by mother's exposure to treatment; or is significant in the investigator's judgment (CS Clarification response A3). Treatment-related serious AEs (SAEs) were experienced by n=56 (17.0%) of the A+B group (CS Section B.2.10). The most common SAEs were gastrointestinal haemorrhage (2.4%), oesophageal varices haemorrhage (2.4%), and pyrexia (2.1%) (CS Section B.2.10).

Grade 3 or 4 AEs were experienced by n=86 (55.1%) of the sorafenib group, of which n=71 (45.5%) were considered by the investigator to be related to treatment (CS Section B.2.10).(Cheng 2019)[1]

The most common Grade 3 or 4 AEs in the sorafenib group were hypertension (9.0%), palmar-plantar erythrodysaesthesia syndrome (8.3%), diarrhoea (3.8%), decreased appetite (3.8%), hypophosphataemia (3.2%), fatigue (3.2%), aspartate aminotransferase increased (2.6%), blood bilirubin increased (2.6%) and rash (2.6%) (CS Section B.2.10).

Treatment-related SAEs were experienced by n=24 (15.4%) of the sorafenib group (CS Section B.2.10). The most common SAEs were gastrointestinal haemorrhage (1.9%), oesophageal varices haemorrhage (0.6%), and pyrexia (1.3%) (CS Section B.2.10).

There was a higher rate of discontinuations for AEs for A+B than for sorafenib; however, there was a shorter duration of study treatment in the sorafenib group due to progression or death. In the safety evaluable population, median study-treatment duration was 7.4 months for atezolizumab, 6.9 months for bevacizumab, and 2.8 months for sorafenib (CS Section B.2.10). AEs led to withdrawal from study treatment for n=51 (15.5%) in the A+B group, and n=16 (10.3%) in the sorafenib group (CS Section B.2.10).

The most common AE leading to discontinuation in the A+B group was oesophageal varices haemorrhage (1.2%) (CS Section B.2.10). All other AEs leading to discontinuation occurred in <1% of patients, in either treatment arm (CS Section B.2.10).

AEs led to dose interruption in n=163 (49.5%) of the A+B group, and to dose interruption or modification in n=95 (60.9%) of the sorafenib group (CS Section B.2.10).

The most common AEs leading to dose interruption in the A+B arm were proteinuria (6.7%), hypertension (6.1%), aspartate aminotransferase increased (5.2%), alanine aminotransferase increased (3.3%), hyperthyroidism (2.7%), platelet count decreased (2.4%), and pyrexia (2.4%) (CS Section B.2.10). The most common AEs leading to dose reduction/interruption in the sorafenib group were palmar-plantar erythrodysaesthesia syndrome (17.3%), diarrhoea (10.9%), blood bilirubin increased (5.1%), fatigue (4.5%), decreased appetite (4.5%), hypertension (3.8%), platelet count decreased (3.2%), pyrexia (3.2%), vomiting (3.2%), rash (3.2%), aspartate aminotransferase increased (3.2%), ascites (2.6%), nausea (2.6%), abdominal pain (2.6%), alanine aminotransferase increased (2.6%), and asthenia (2.6%) (CS Section B.2.10).

3.2.3.2 Study GO30140 (NCT02715531) AEs

AE data were available from the Phase Ib GO30140 study (NCT02715531) (CS Appendix F) (clinical trials gov).[17]

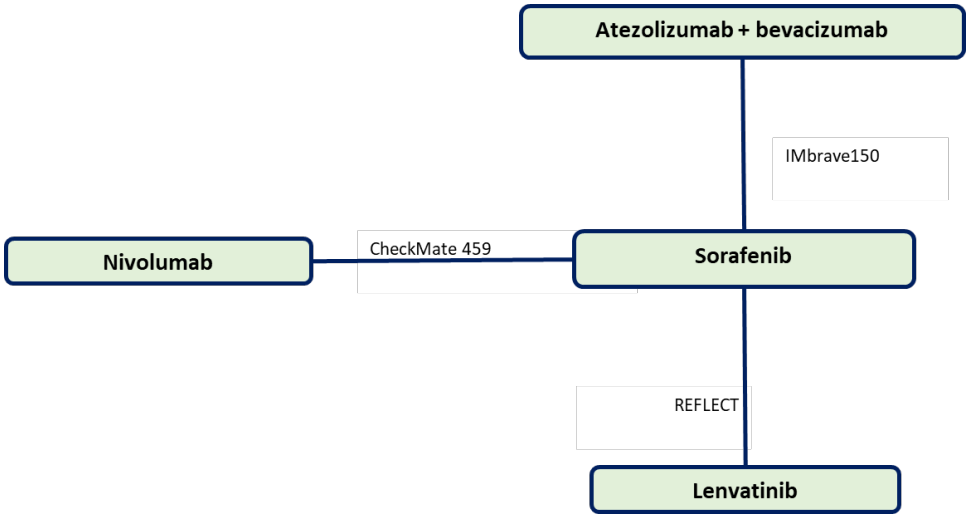
Participants in this study had advanced or metastatic and/or unresectable HCC who had received no prior systemic treatment (clinical trials gov).[17] The experimental group comparing A+B (n=60) to atezolizumab monotherapy (n=58) (CS Appendix F) administered atezolizumab at a dose of 1200 mg q3w, and for the combination group bevacizumab at 15 mg/kg q3w (clinical trials gov).[17]

In the A+B arm (median study treatment duration 5.21 months), 95.0% experienced at least one AE of any grade (CS Appendix F). In the atezolizumab monotherapy arm (median study treatment duration 1.61 months), 89.7% experienced at least one AE of any grade (CS Appendix F). No fatal AEs were reported in either group. Grade 3/4 AEs were experienced by 36.7% of the A+B arm, and in 13.8% of the atezolizumab monotherapy arm (CS Appendix F). In the A+B arm the more common grade 3+ events were hypertension (5.0%) and proteinuria (3.3%) (CS Appendix F).

3.3 Critique of trial identified and relevant in the indirect comparison

The company presented the evidence network in Figure 4.

Figure 4: Evidence network for level 1 comparators reporting OS and PFS (reproduced from Figure 7 of the CS)



REFLECT compared lenvatinib (n=478) with sorafenib (n=476) in first-line treatment of patients with unresectable HCC (Kudo 2018).[18] The primary endpoint was OS, which was assessed for non-inferiority of lenvatinib, and also for superiority over sorafenib. (Kudo 2018).[18] Sorafenib was administered orally at a dose of 400mg BID (as for IMbrave150). Lenvatinib was administered orally at a dose of 12 mg/daily (for bodyweight ≥60 kg) or 8 mg/daily (for bodyweight <60 kg). (Kudo 2018)[18]

Eligibility criteria for REFLECT included: unresectable HCC (confirmed histologically, cytologically, or clinically in accordance with American Association for the Study of Liver Diseases criteria); no previous systemic therapy for HCC; one or more measurable target lesions; Barcelona Clinic Liver Cancer stage B or C; Child-Pugh class A; and ECOG PS 0 or 1. (Kudo 2018)[18]

REFLECT was generally at low risk of bias, apart from being open-label (Table 12). Baseline characteristics were generally well balanced between treatment groups; however, there was a higher frequency of hepatitis C aetiology in the sorafenib group, and a higher frequency of patients in the lower AFP category in the sorafenib group, than in the lenvatinib group (Kudo 2018).[18]

Table 12: QA of REFLECT (adapted from CS Appendix D and Kudo 2018)[18]

	QA by CS (CS Appendix D.3)	QA by ERG
Was the allocation sequence adequately generated?	Low - The randomisation sequence was generated by an independent statistician by the system vendor, and the investigators obtained the randomisation assignments from the system directly	Low Independent statistician via IVRS
Was the concealment of treatment allocation adequate?	Low - Allocation of treatment group was done with an interactive voice–web response system, which also functioned as the allocation concealment method	Low IVRS
Was knowledge of the allocated interventions adequately prevented from participants and personnel	High – open label	High open label
Was knowledge of the allocated interventions adequately prevented from outcome assessors	Low - Although open label design masked independent assessments were conducted	Mixed High risk for investigator assessed endpoints and PROs. Low risk for "Post-hoc exploratory tumour assessments using mRECIST and RECIST version 1.1" which were conducted by "masked central independent imaging review" [18]
Were incomplete outcome data adequately addressed?	Low - Primary and secondary endpoints reported	Low ITT analyses for effectiveness outcomes
Are reports of the study free of suggestion of selective outcome reporting?	Low - The study reports outcomes of interest as specified	Low The effectiveness, safety and HRQoL outcomes in the protocol (clinicaltrials.gov/ct2/show/NCT01761266) are reported in Kudo 2018[18]

Baseline characteristics of REFLECT and IMbrave150 were similar in terms of age, sex, ECOG PS, BCLC stage, and participants were Child Pugh class A in both trials (Table 13). REFLECT had a higher proportion of patients from the Asia-Pacific region than IMbrave150, and the trials different differentiations of AFP category imply lower AFP in the REFLECT population than in IMbrave150. (Kudo 2018)[18]

Table 13: Baseline characteristics IMbrave150 and REFLECT (adapted from CS Table 5, and Kudo 2018[18])

	IMbrave150 A+B n=336	IMbrave150 Sorafenib n=165	REFLECT Sorafenib n=476	REFLECT Lenvatinib n=478
Median age, years	64.0	66.0	62.0 (range 22-88)	63.0 (range 20-88)
Male, n (%)	227 (82.4)	137 (83.0)	401 (84%)	405 (85%)
Geographic region	Asia (excl. Japan) 133 (39.6) Rest of World 203 (60.4)	Asia (excl. Japan) 68 (41.2) Rest of World 97 (58.8)	Asia-Pacific 319 (67%) Western 157 (33%)	Asia-Pacific 321 (67%) Western 157 (33%)
ECOG PS, n (%)				
0	209 (62.2)	103 (62.4)	301 (63%)	304 (64%)
1	127 (37.8)	62 (37.6)	175 (37%)	174 (36%)
BCLC stage at study entry, n (%)				
A1	5 (1.5)	3 (1.8)	0	0
A4	3 (0.9)	3 (1.8)	0	0
B	52 (15.5)	26 (15.8)	92 (19%)	104 (22%)
C	276 (82.1)	133 (80.6)	384 (81%)	374 (78%)
Aetiology of HCC				
HBV	164 (48.8)	76 (46.1)	228 (48%)	251 (53%)
HCV	72 (21.4)	36 (21.8)	126 (26%)	91 (19%)
Non-viral	100 (29.8)	53 (32.1)	122 (26%)*	136 (28%)*
Extrahepatic spread and/or macrovascular invasion present at study entry Yes	258 (76.8)	120 (72.7)	336 (71%)	329 (69%)
[redacted]	[redacted]	[redacted]	NR	NR
[redacted]	[redacted]	[redacted]	NR	NR
Weight n(%)				
<60kg	72 (21%)	41 (25%)	146 (31%)	153 (32%)
≥60kg	264 (79%)	124 (75%)	330 (69%)	325 (68%)
AFP category at screening, ng/mL	<400 210 (62.5) ≥400 126 (37.5)	<400 104 (63.0) ≥400 61 (37.0)	<200 286 (60%) ≥200 187 (39%)	<200 255 (53%) ≥200 222 (46%)

*alcohol/other/unknown

The median time on study treatment was 5.7 months for patients in the lenvatinib group and 3.7 months for the sorafenib group (Kudo 2018).[18] For information, this was relatively similar to the sorafenib group of IMbrave150, for whom median time on study treatment was 2.8 months (CS Section B.2.10).

Median overall survival in the lenvatinib group was 13·6 months (95% CI 12·1, 14·9), and in the sorafenib group median OS was 12·3 months (95% CI 10·4, 13·9). Lenvatinib was non-inferior to sorafenib (HR 0·92, 95% CI 0·79, 1·06) with respect to a non-inferiority margin of 1.08. (Kudo *et al.* 2018)[18]

For information, the sorafenib group in REFLECT had a similar median OS (12.3 months), to the sorafenib group in IMbrave150 which had an estimated median OS of 13.2 months.[1] The REFLECT authors speculated that the sorafenib group had a higher median OS than in previous trials partly due to the subsequent therapy after discontinuation of study treatment, with post-sorafenib treatment of systemic treatment (39% patients) or non-systemic treatment (27% patients).[18] In IMbrave150, 44.2% of the sorafenib group had subsequent HCC systemic therapy (CS Section B.2.3.3).

Lenvatinib showed statistically significant advantages over sorafenib for the secondary endpoints PFS, TTP and ORR. Median PFS assessed by masked review according to RECIST 1.1, was 7·3 months (95% CI 5·6, 7·5) for the lenvatinib group, and 3·6 months (95%CI 3·6, 3·9) for the sorafenib group. The HR for PFS was 0·65 (0·56, 0·77) $p<0·0001$. TTP assessed by masked review according to RECIST 1.1 produced a HR 0·61 (0·51, 0·72) $p<0·0001$, median TTP was 7·4 months (95%CI 7.3, 9.1) for the lenvatinib group, and 3·7 months (95% CI 3·6, 5.4) for the sorafenib group. Objective response rate assessed by masked review according to RECIST 1.1 produced a HR of 3.34 (2.17, 5.14) $p<0·0001$ was 18.8% (95% CI 15.3, 22.3) for the lenvatinib group, and 6.5% (95% CI 4.3, 8.7) for the sorafenib group.[18]

AEs were experienced by 99% of patients in each group, and grade 3+ adverse events were experienced by 67% of the sorafenib group, and 75% of the lenvatinib group.[18]

3.4 Critique of the indirect comparison and/or multiple treatment comparison

In the absence of head-to head evidence for all treatments of interest, a network meta-analysis (NMA) was performed to estimate the relative treatment effect of lenvatinib and sorafenib compared to A+B. An SLR identified twenty-three studies (including IMbrave150), which formed a connected network for inclusion into the SLR (Figure 6 of the CS). After exclusions, three studies were included in the base case evidence network: IMbrave150 (A+B vs. sorafenib),[19] REFLECT (lenvatinib vs. sorafenib)[20] and CheckMate 459 (nivolumab vs. sorafenib).[21] In response to clarification question A9, the company justified the inclusion of nivolumab by stating that "*when the NMA was being developed, it was unclear whether Nivolumab would be approved or not. In addition, the NMA covers all systemic therapies with information published in 1L in HCC since the sorafenib approval in 2007*." In addition, the company

provided results showing that removing CheckMate 459 (an RCT of nivolumab versus sorafenib) from the network did not affect the results.

The base case NMAs were conducted using a Bayesian random effects model of log HRs. The HR for OS was 0.63 (95% Credible Interval [CrI]: 0.32, 1.25) for A+B compared to lenvatinib, and 0.58 (95% CrI: 0.35, 0.99) for A+B compared to sorafenib. Hence, there was uncertainty whether A+B is superior to lenvatinib; the ERG assumes that the posterior probability that A+B is superior to lenvatinib reported in Table 18 of the CS is an error; the probability the A+B is superior to lenvatinib is given as 0.94 in Table 18 of the CS and yet the upper limit of the 95% CrI is 1.25. The HR for PFS was 0.91 (95% CrI: 0.23, 3.65) for A+B compared with lenvatinib, and 0.59 (95% CrI: 0.23, 1.58) for A+B compared with sorafenib.

There are no feedback loops in the evidence base so it is not possible to formally assess inconsistency in the evidence base (i.e. there is only direct evidence specific to each trial). Biased estimates of treatment effect would arise if there was an imbalance in the distribution of treatment effect modifiers across studies comparing difference pairs of treatments.

With limited studies, it was necessary to incorporate external information about the between-study standard deviation in the random effects models. However, the company did not report posterior estimates of the between-study standard deviations. Furthermore, the company only reported summaries of the mean of the random effects distributions, although it is recommended to also present summaries of predictive distributions of effects in new studies. In addition, posterior predictive distributions of effects in new studies are what is recommended to use to characterise uncertainty in economic models to account for heterogeneity.

It is not unreasonable to conduct an NMA of HRs for the purpose of answering the question whether there is evidence of an average treatment effect ignoring any treatment-by-time interaction over the duration of the observed studies. However, using HRs from NMAs in the context of an economic evaluation is inappropriate in order to estimate population mean benefit, and it is inconsistent to generate the lenvatinib survival function using hazard ratios; this is discussed further in Section 4.3.4.2.

3.5 Additional work on clinical effectiveness undertaken by the ERG

No additional work related to clinical effectiveness was undertaken by the ERG.

3.6 Conclusions of the clinical effectiveness section

The ERG believes that no RCTs of A+B meeting the inclusion criteria of the final scope have been missed by the CS. The search for clinical evidence reflected the decision problem set out in the final scope, but did not include best supportive care (BSC), and had a broader selection of active comparators. The ERG believes this allowed identification of relevant sorafenib and lenvatinib studies. Clinical effectiveness evidence was available from one RCT, IMbrave150, that compared A+B to sorafenib. Safety data were available from IMbrave150 and the phase 1b study GO30140.

IMbrave150 randomised adults with adults with locally advanced or metastatic and/or unresectable hepatocellular carcinoma (HCC), who had no previous systemic treatment for HCC, to A+B (atezolizumab 1200 mg IV infusions every three weeks, and bevacizumab 15 mg/kg every three weeks, n=336) or sorafenib (400 mg orally twice per day n=165). The quality of the IMbrave150 RCT was assessed using well-established and recognised criteria. IMbrave150 was an open label trial, but was of otherwise good methodological quality.

According to clinical advice, the demographics of the IMbrave150 trial population were broadly representative of the UK population eligible for A+B treatment, although there are a smaller proportion of Asian patients, and patients with an aetiology of hepatitis B in the UK than in the trial, and more patients in the UK would have aetiology of alcohol or non-alcohol related fatty liver disease, than in the trial. Prior treatments in IMbrave150 were broadly similar to a UK population, although prior radiotherapy is rare in the UK, whereas 10% of trial patients received prior radiotherapy.

OS was statistically significantly higher for A+B, than for sorafenib HR (stratified) 0.58 (95% CI 0.42, 0.79) p=0.0006. Median OS for A+B was not estimable (NE), median OS for sorafenib was 13.2 months (95% confidence interval [CI] 10.4, NE). There was a statistically significant treatment group difference for PFS HR (stratified) 0.59 (95% CI 0.47, 0.76) p<0.0001. Median PFS was 6.8 months (95% CI 5.6, 8.3) in the A+B group, and 4.3 months (95% CI 4.0, 5.6) for the sorafenib group.

The majority of A+B treated patients experienced at least one AE of any severity (98.2%). The most common AEs of any grade in the A+B group were hypertension (29.8), fatigue (20.4%) and proteinuria (20.1%). The most common NCI-CTCAE Grade 3 or 4 AEs experienced in the A+B group were hypertension (10.3%), aspartate aminotransferase increased (4.3%) and proteinuria (2.7%). The most common Grade 3 or 4 AEs in the sorafenib group were hypertension (9.0%), palmar-plantar erythrodysaesthesia syndrome (8.3%), diarrhoea (3.8%), decreased appetite (3.8%), hypophosphataemia (3.2%), fatigue (3.2%), aspartate aminotransferase increased (2.6%), blood bilirubin increased (2.6%) and rash (2.6%).

There was a longer time to deterioration for A+B over sorafenib in three of the HRQoL domains measured: Global Health Status/quality of life; physical functioning; and role functioning.

4 COST EFFECTIVENESS

This section provides a structured critique of the economic evidence submitted by the company to support the cost effectiveness of A+B for untreated locally advanced metastatic HCC patients.

The company present a systematic literature review (SLR) of relevant economic evidence and then present a *de novo* economic evaluation. The company also provided an electronic version of their economic model developed in Microsoft Excel.

4.1 ERG's comment on company's review of cost-effectiveness evidence

4.1.1 Objective of cost effectiveness review

The company performed systematic literature searches for: i) published cost-effectiveness studies in the first-line treatment of patients with unresectable HCC (CS Appendix G); ii) HRQoL and health state utilities in patients with HCC in the first-line setting (CS Appendix H), and iii) cost and healthcare resource use of first-line HCC (CS Appendix I)**.**

An extensive range of databases, HTA conference websites and grey literature sources were searched by the company. In the economic SLR search (CS Appendix G), the following sources were searched: MEDLINE [via Ovid], MEDLINE In-Process, Epub Ahead of Print and Daily [via Ovid], Embase [via Ovid], Cochrane Database of Systematic Reviews [via EBM Reviews], Cochrane Central Register of Controlled Trials [via EBM Reviews], The Health Technology Assessment [via EBM Reviews]), and Database of Abstracts of Reviews of Effects [via EBM Reviews], American College of Physicians [via EBM Reviews], Cochrane Clinical Answers [via EBM Reviews], Cochrane Methodology Register [via EBM Reviews], and EconLit [via Ovid] in October 2019. The ERG notes that in contrast to the clinical effectiveness SLR (Section 3.1), the company did not attempt to update the search so eligible studies post 2019 would be excluded in the review.

A comprehensive list of intervention and comparator search terms combined with an economic search filter was applied in the company's MEDLINE and Embase search. As seen in the EconLit database search, a more sensitive strategy and approach would be obtained by combining HCC terms with an economic search filter (CS Appendix G Table 18, page 54). However, the ERG acknowledges that the number of records retrieved may have been unmanageable for the company to review.

Supplementary searches by the company include searching several conference abstract websites in the last three years (2017-2019): American Society of Clinical Oncology, European Society for Medical Oncology, American Association for Cancer Research, International Society for Pharmacoeconomics

and Outcomes Research: European Meeting, Health Technology Assessment International, Society for Medical Decision Making (Appendix G, page 89).

The company also searched several HTA websites for previous technology submissions: NICE, the Scottish Medicines Consortium, the All Wales Medicines Strategy Group, the Pharmaceutical Benefits Advisory Committee, the Canadian Agency for Drugs and Technologies in Health including the pan-Canadian Oncology Drug Review.

Supplementary searches in several HTA databases and the Research Papers in Economics (via EconPapers) were undertaken by the company for published literature. The ERG notes that there is significant overlap of coverage of content indexed in the International Network of Agencies for Health Technology Assessment, the University of York Centre for Reviews and Dissemination and the National Institute for Health Research HTA database. However, the ERG is unable to confirm whether the searches were applied consistently across these sources as the search strategies were absent in the CS. In addition, the company's purpose for searching Google Scholar was not explicitly stated.

For the HRQoL and health state utilities studies searches, the company searched the same sources as in the cost-effectiveness review but with the inclusion of two web sources: the EuroQol website and the University of Sheffield ScHARRHUD database (CS Appendix H). The company's searches were comprehensive (HCC population combined with a recognized and published HSU search filter) with no observable and consequential errors in the strategies.

In the economic, cost and healthcare resource and humanistic burden searches of first-line HCC (CS Appendix I), the company searched the same sources as those reported in the cost-effectiveness SLR. The company's searches were comprehensive (HCC population combined with cost and resource terms) with no observable and consequential errors in the strategies.

4.1.2 The inclusion and exclusion criteria used in the study selection

The inclusion criteria used by the company to facilitate study selection are presented in Table 14. The ERG considers the inclusion criteria to be appropriate to capture recent and relevant published evidence.

Table 14: Inclusion/exclusion criteria for the company's economic review

Category	Inclusion criteria
Population (P)	Aligned with patients enrolled in the IMbrave150 study: Age: adults aged ≥18 years Gender: any Race: any Disease: patients with locally advanced or metastatic HCC who have received no prior systemic therapy for HCC
Intervention (I)	A+B
Comparators (C)	Any pharmacological intervention whether single agent or in a combination including sorafenib, nivolumab, TACE, radiotherapy, other investigational agents, and others being examined in ongoing studies.
Outcome (O)	Cost-effectiveness estimates (costs, health outcomes, and ICERs)
Study design	- Cost-effectiveness analysis - Cost-utility analysis - Cost-minimisation analysis - Cost-benefit analysis
Language	English language publications or non-English language publications with an English abstract.

HCC, hepatocellular carcinoma; ICER, incremental cost-effectiveness ratio; TACE, transarterial chemoembolisation

4.1.3 Findings of the cost effectiveness review

Fifty-seven studies were identified that were relevant to the decision problem (27 full publications, 22 conference abstracts, and eight previous HTA submissions); however, none of these included A+B as an option. The majority of studies were cost-utility analyses reporting incremental cost per QALY gained with most of the models using a Markov approach. Table 30 in the CS summarises the evidence found in the 27 full publications identified.

4.1.4 Conclusions of the cost effectiveness review

As the company's searches did not identify any relevant studies of A+B, they developed a *de novo* health economic model.

4.2 **Summary of the company's submitted economic evaluation**

4.2.1 Population

The population included in the company's health economic analysis reflects adult patients with locally advanced or metastatic HCC who have not received prior systemic treatment. The modelled patient characteristics reflect those of the full patient population within the IMbrave150 study[1] with a mean age of 63.4 years, and 18% of the population are assumed to be female. The mean baseline body weight was 71.7 kg and the mean baseline body surface area was 1.82 m^2.

In response to clarification question B2, the company provided separate subgroup analyses for the IMbrave150 population excluding Asia (except for Japanese patients).

4.2.2 Interventions and comparators

Atezolizumab is provided at a fixed dose of 1200mg whereas bevacizumab is given at a weight-based dose of 15mg/kg. Both drugs are administered intravenously every 3 weeks until loss of clinical benefit, disease progression, or unacceptable toxicity. Both atezolizumab and bevacizumab can be given separately if patients became intolerant to the other intervention.

In line with the final NICE scope, the included comparators were sorafenib and lenvatinib. Both are administered orally at a fixed dose of 400 mg, twice daily, for sorafenib, and a weight-dependant dose, once daily, for lenvatinib (12 mg for patients ≥60 kg and 8 mg for patients <60 kg).

4.2.3 Perspective, time horizon and discounting

The base case model adopts an NHS and Personal Social Services (PSS) perspective. The base case model uses a 20-year time horizon; shorter time horizons were included in the company's scenario analyses. Both costs and quality-adjusted life years (QALYs) were discounted at 3.5% per annum as recommended by NICE.[22]

4.2.4 Model structure

As part of its submission to NICE, the company developed a fully executable partitioned survival model (PSM) in Microsoft® Excel that included three mutually exclusive and exhaustive health states: (i) progression-free survival (PFS); (ii) progressed disease (PD); and (iii) death. The model is similar to that of other treatments for advanced/metastatic cancer previously submitted to NICE as part of the STA process. The model structure is shown in Figure 5. A weekly cycle length was used, and half-cycle correction was implemented.

Figure 5: The company's model structure (reproduced from Figure 9 of the CS)

All patients are assumed to enter the model in the progression-free health state and remain there until progression or death. As with a standard PSM, the health state membership for A+B and sorafenib is inferred via survival functions fitted to the IMbrave150 study PFS and OS data. As lenvatinib was not included in IMbrave150, the company undertook an NMA to estimate HRs for PFS and OS for patients treated with lenvatinib compared to those treated with A+B. The lenvatinib HRs were then applied to the A+B PFS and OS functions.

Parametric survival models were fitted to time to treatment discontinuation (TTD) data from the IMbrave150 study for atezolizumab, bevacizumab and sorafenib separately. The company assumed that time to treatment discontinuation on lenvatinib is equivalent to time to progression as TTD was not explicitly reported in REFLECT.[18]

4.2.5 Evidence used to inform the company's model parameters

The main groups of the company's base case model parameters and the evidence sources used to populate these are summarised in Table 15. These are discussed in further detail in the subsequent sections.

Table 15: Evidence sources used to inform model parameters in the company's base case

Parameter type	Parameter	Source(s)
Time-to-event parameters	PFS – A+B	The IMbrave150 study[1]
	PFS – sorafenib	The IMbrave150 study[1]
	PFS – lenvatinib	HRs from the company's ITC[23] applied to A+B data
	OS – A+B	The IMbrave150 study[1]
	OS – sorafenib	The IMbrave150 study[1]
	OS – lenvatinib	HRs from the company's ITC[23] applied to A+B data
	TTD – A+B	The IMbrave150 study[1]
	TTD – sorafenib	The IMbrave150 study[1]
	TTD – lenvatinib	Assumed the same as the lenvatinib PFS
Adverse event rate (grade 3+ experienced by at least 5% of the patients)	A+B	The IMbrave150 study[1]
	Sorafenib	The IMbrave150 study[1]
	Lenvatinib	The REFLECT study[18]
HRQoL	Two sets of health utilities (on/off treatment), each has four utility values defined in terms of proximity to death	EQ-5D-5L questionnaires collected in The IMbrave150 study[1]
Resource use and costs	A+B acquisition cost (including PAS)	The CS by Roche[23]
	Sorafenib and lenvatinib acquisition cost (list prices)	The British National Formulary (BNF)[24]
	Drug dosing as planned per individual characteristics in the IMbrave150 study	The IMbrave150 study[1]
	A+B subsequent therapy %	Assumed to be 0%
	Sorafenib subsequent therapy %	Assumed that 44.2% of patients on sorafenib who went on to receive subsequent therapy, would only receive regorafenib
	Lenvatinib subsequent therapy %	Assumed to be 0%

	Subsequent therapy acquisition costs (list prices)	The British National Formulary (BNF)[24]
	Drug administration costs	NHS Reference costs 2018-19[25]
	Medical resource use for progression-free and progressed health states	Expert elicitation
	Adverse event costs	NHS Reference costs 2018-19[25]
	End of life care costs	Costs were sourced from Georghiou and Bardsley[26] and inflated using PSSRU HCHS indices[27]

A+B, atezolizumab plus bevacizumab; CS, company's submission; HCHS, Hospital and Community Health Services; HR, hazard ratio; HRQoL, health-related quality of life; ITC, indirect treatment comparison; NHS, National Health Services; OS, overall survival; PAS, patient access scheme; PFS, progression free survival; PSSRU, Personal Social Services Research Unit; TTD, time to treatment discontinuation

4.2.5.1 Treatment effectiveness and extrapolation in the base case

The data for the full ITT population from IMbrave150 were used to model PFS and OS independently in the A+B and sorafenib arms in the company's base case. At the time of data cut-off (29th August, 2019), 71% of the patients on A+B were still alive compared to 61% on sorafenib. Approximately 41% of patients were still progression-free on A+B compared to 34% on sorafenib.

The company followed guidance for fitting and selecting survival models based on NICE Decision Support Unit Technical Support Document 14.

The company investigated the use of a range of parametric survival models: exponential, Weibull, Gompertz, log-normal, log-logistic and generalised gamma distributions fitted independently to each treatment arm. The company also incorporated in the model the option of using the KM survival functions directly from IMbrave150 and extrapolating beyond the duration of follow-up using one of the six aforementioned models.

4.2.5.1.1 Estimating OS

The company considered independently fitted parametric distributions (i.e. for each treatment arm separately) in the economic model. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values were compared to assess relative goodness-of-fit to the observed data. For the fit to the A+B OS data, the generalised gamma was the distribution with the lowest BIC, although there was little to distinguish between this and log-normal and log-logistic distributions as the difference in BIC values were below 2. For the fit to the sorafenib OS data, the log-normal was the distribution with

the lowest BIC, although there was little to distinguish between this and the log-logistic and generalised gamma distributions as the difference in BIC values were below 2. There was positive, but not strong evidence that the generalised gamma distribution was a better fit to the observed data than the Weibull distribution.

AIC and BIC values are presented in Table 16; the company chose the exponential distribution for its base case. The clinical advisor to the ERG did not rule out that the hazard of a death could be constant throughout the patient's life and proportional for each treatment.

Table 16: Goodness-of-fit of parametric models to OS data observed in IMbrave150

	A+B		Sorafenib		Summed Totals*	
Parametric distribution	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Exponential	547.24 (6)	551.06 (5)	325.09 (5)	328.20 (5)	872.33 (6)	879.26 (5)
Weibull	538.43 (4)	546.06 (4)	322.36 (4)	328.57 (4)	860.79 (4)	874.63 (4)
Log-normal	534.56 (2)	542.19 (2)	**320.63 (1)**	**326.84 (1)**	**855.19 (1)**	**869.03 (1)**
Generalised Gamma	**534.55 (1)**	**542.19 (1)**	322.26 (3)	331.57 (3)	856.81 (2)	873.76 (3)
Log-logistic	536.30 (3)	543.93 (3)	320.88 (2)	327.09 (2)	857.18 (3)	871.02 (2)
Gompertz	544.90 (5)	552.53 (6)	325.28 (6)	331.49 6)	870.18 (5)	884.02 (6)

*Calculated by the ERG. Rounding errors may be present
Numbers in brackets provide the rank ordering of each distribution. The best-fitting distribution to the observed data is highlighted in bold.

An HR of [redacted] (Table 18 of the CS) was applied to the selected A+B parametric model to estimate OS survival function for lenvatinib. The company stated that they deemed it appropriate to apply HRs from the NMA to an accelerated failure time model. The ERG does not agree that a HR should be applied to models that are not proportional hazard models. Furthermore, the ERG does not accept that it is appropriate to use hazard ratios to estimate a survival function or to estimate the effect of lenvatinib from a different model to one estimating the effect of sorafenib. These issues are discussed in Section 4.3.4.2.

The company then assessed the tails of the parametric distributions for their clinical plausibility using judgements from six UK clinicians. The aim was for clinicians to ensure that the survival functions estimated in the populations defined by the IMbrave50 and REFLECT studies were consistent with what is seen in UK clinical practice. The clinicians concluded "*that only the exponential model and the Generalised Gamma model represented clinically plausible estimates, as the remaining four models*

projected a higher OS for sorafenib than lenvatinib, which is not aligned with the REFLECT trial clinical data which showed lenvatinib OS to be non-inferior to sorafenib." (page 99 of the CS)

The clinicians considered that the exponential model represented the most realistic survival rates at 12, 24, 36, 48 and 60 months for both sorafenib and lenvatinib but noted that it might under-estimate the long-term proportion of patients surviving when treated with A+B. The company selected the exponential distribution in its base case and explored the impact of a generalised gamma distribution in scenario analyses. Figure 6 presents the selected exponential models for the three treatment arms and their respective KM survival functions which have been marked academic-in-confidence by the company. Comparisons between IMbrave150 and REFLECT represent a naïve indirect comparison. The full range of parametric models are presented in Figure 16 of the CS for A+B and in Figure 17 of the CS for sorafenib.

In response to clarification question A7, as described in Section 3.4, the company also fitted a first-degree Bayesian fixed effect fractional polynomial NMA to allow for time-varying HRs for OS. The impact of this model was explored as a scenario analysis.

Additionally, the company's economic model offered the option to maintain the treatment effect until a user-selected time point after which the probabilities of death on A+B are assumed to be the same as that of sorafenib.

Figure 6: The company's base case OS extrapolation (adapted from the company's model)

4.2.5.1.2 Estimating PFS

The company fitted six parametric models to the A+B and sorafenib data independently. AIC and BIC values were compared to assess relative goodness-of-fit to the observed data and are presented in Table 17. The log-normal distribution produced the smallest BIC (highlighted) for both treatment arms and was the model used in the company base case.

Table 17: Goodness-of-fit of parametric models to PFS data observed in IMbrave150

	A+B		Sorafenib		Summed Totals*	
Parametric distribution	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
Exponential	841.34 (5)	845.16 (5)	381.53 (6)	384.64 (6)	1222.87 (6)	1229.80 (5)
Weibull	836.71 (4)	844.34 (4)	370.46 (4)	376.67 (4)	1207.17 (4)	1221.01 (4)
Log-normal	815.65 (2)	**823.28 (1)**	**360.21 (1)**	**366.42 (1)**	1175.86 (2)	**1189.70 (1)**
Generalised Gamma	**813.19 (1)**	824.64 (2)	362.19 (2)	371.50 (2)	**1175.38 (1)**	1196.14 (2)
Log-logistic	825.06 (3)	832.70 (3)	364.08 (3)	370.29 (3)	1189.14 (3)	1202.99 (3)
Gompertz	843.11 (6)	850.75 (6)	378.92 (5)	385.14 (5)	1222.03 (5)	1235.89 (6)

*Calculated by the ERG. Rounding errors may be present
Numbers in brackets provide the rank ordering of each distribution. The best-fitting distribution to the observed data is highlighted in bold.

As for OS, the company assessed the tails of the parametric distributions for their clinical plausibility based on judgments from six UK clinicians "*to ensure the curves represented UK clinical practice*." (page 106 CS) In addition, the company assessed whether the fitted lenvatinib PFS survival function is in line with the lenvatinib KM survival function from REFLECT. The company ruled out the generalised gamma distribution as it exceeded the OS exponential models selected in the company's base case. The company excluded the Weibull and Gompertz distributions because they did not produce "*a lenvatinib PFS curve in line with lenvatinib KM data from the REFLECT trial*."

In addition to being the model that represented the observed PFS survival functions best, the log-normal distribution was selected by the company for its base case as it was used in the lenvatinib submission to extrapolate PFS. The impact of using exponential and log-logistic distributions was explored in scenario analyses.

Figure 7 presents the fitted lognormal models for the three treatment arms and their respective KM survival functions.

An HR of [redacted] (Table 21 of the CS) was applied to all of the A+B parametric models to estimate the PFS survival function for lenvatinib. The company stated that they deemed it appropriate to apply HRs

from the NMA to an accelerated failure time model. The ERG does not agree that a hazard ratio should be applied to models that are not proportional hazard models. Furthermore, the ERG does not accept that it is appropriate to use hazard ratios to estimate a survival function or to estimate the effect of lenvatinib from a different model to one estimating the effect of sorafenib. These issues are discussed in Section 4.3.4.2.

Additionally, the company's economic model includes an option to maintain treatment effect for A+B until a user-selected time point after which the probabilities of progression associated with sorafenib were applied.

Figure 7: The company's base case PFS extrapolation (adapted from the company's model)

4.2.5.2 Duration of treatment

In IMbrave150 patients were allowed to receive atezolizumab until loss of clinical benefit or unmanageable toxicity meaning that patients could receive the regimen following disease progression. Treatment duration data were collected in the IMbrave150 study and TTD KM survival functions were estimated separately for each of atezolizumab, bevacizumab and sorafenib. TTD data were not available for lenvatinib; hence, the company assumed that the PFS survival function for lenvatinib was a proxy for the TTD survival function of lenvatinib in the economic model.

Based on the KM survival function at 16 months 30% of patients on A+B were still receiving at least one of the two treatments compared to 12% of patients on sorafenib. The company decided that the TTD data were relatively complete and used parametric survival modelling only to extrapolate beyond 14 months. In a similar approach to that used for OS and PFS, the company fitted the six parametric distributions to TTD data. Goodness-of-fit statistics are provided in Table 38 of the CS.

Based on AIC and BIC, the generalised gamma and Weibull distributions represented the data best for both atezolizumab and bevacizumab. For sorafenib, the lognormal distribution gave the best fit based on BIC. However, the company stated that the '*Weibull, Log-Normal and Log-Logistic provided poor fit to the observed data and very long unrealistic tails. The Gompertz and Exponential reported the same values. Therefore, the Exponential parametric distribution is used in the base case for the extrapolation of TTD, because whilst it does not provide the best statistical fit, it does demonstrate the best visual fit out of all potential distributions, as well as clinical validity.*'

The company decided that because "*the observed TTD data for Atezo+Bev and sorafenib in IMbrave150 are relatively complete, it was deemed appropriate to use the TTD KM curve followed by the Exponential distribution, as this was the parametric model showing the best visual fit to the observed data, for atezolizumab, bevacizumab and sorafenib*." The company used the KM survival function, without allowing for uncertainty, for the first 14 months, and extrapolated beyond that timepoint using an exponential distribution. The time point of 14 months was said to '*to ensure robustness in terms of patient numbers at risk*' although no comment was made on why 14 months was deemed preferable to 13 or 15 months but this value could be changed in the model. Figure 8 shows the base case TTD survival functions used in the company's economic model.

Figure 8: The company's base case TTD estimation

4.2.5.3 Treatment safety

In the model, AEs were associated with additional costs. All grade 3, 4 or 5 AEs were included in the model where at least 5% of patients experienced them in at least one of A+B and sorafenib within

IMbrave150. The clinical advice provided to the ERG indicated that there were no known, rare, AEs that would have a high clinical burden or large cost.

For lenvatinib the rates within the lenvatinib NICE submission[28] were used which represents a naïve indirect comparison. In its response to clarification question A5, the company acknowledged that the '*rate of lenvatinib adverse events would be lower as the length of follow-up in the REFLECT trial is much longer than in the IMbrave150 trial. This would result in a more accurate rate of lenvatinib adverse events, as the current naive indirect comparison relies on the assumption that time doesn't determine the relative frequency of adverse events, as well as comparability of populations.*' This approach is therefore unfavourable to lenvatinib.

The incidence rates used to inform the economic model are presented in Table 39 of the CS. The company applied the impact of adverse events on costs for each cycle to patients still on treatment after converting the incidence rates into weekly probabilities. The costs per AE are discussed in Section 4.2.5.5.5.

In the base case the company did not include the impact of AEs on health-related quality of life (HRQoL) assuming that any disutility due to an AE was already captured in the EQ-5D data collected in the study and incorporating extra disutility could be considered double counting. In the clarification response (question A6), the company stated that '*A regression analysis was carried out to determine if the EQ-5D questionnaire was measured whilst an adverse event was active. The results show that adverse events did have a statistically significant impact on the EQ-5D measurement (-0.04 - this is for any AE), but this impact is not considered clinically significant.*' Thus, it appears that the EQ-5D did capture AEs that were being experienced when the questionnaire was completed, which was on treatment administration and three-monthly thereafter for one year. The average duration of adverse events was provided in Table 3 of the clarification response to indicate the likelihood that AEs would be experienced when the EQ-5D was completed. The company also presented a scenario analysis where an additional disutility was applied; this had little impact on the model results.

4.2.5.4 Health related quality of life

The SLR carried out by the company identified 23 unique HRQoL studies relevant to the technology appraisal; however, only 15 of these were presented as full publications. The company identified only two publications that fully met NICE reference case, where utilities were derived directly from patients using the preferred EQ-5D tool and the UK tariff was used to value the resulting health states; however, both were hepatitis C virus (HCV)-related. The first study (Chong *et al.*) reported an EQ-5D-3L utility value of 0.65 [95% CI 0.44-0.86] and was collected from 15 HCV-related HCC patients.[29] The second

publication was an HTA report examining health benefits of antiviral therapy for mild chronic HCV and sourced the utility value (0.45) for HCC from a prospective multi-centre UK trial.[30, 31]

HRQoL data were collected using the EQ-5D-5L within IMbrave150 at each visit prior to treatment administration (i.e. every 3 weeks) or at each follow-up visit every 3 months for one year following disease progression or treatment discontinuation. The Van Hout *et al*[32] crosswalk mapping algorithm was applied to these data to estimate the corresponding EQ-5D-3L values as recommended in the NICE position statement on the EQ-5D-5L.[32, 33]

The company explored four different approaches to include the mapped mean utility values in the model: i) an on/off treatment approach where three sets of utility scores were derived for patients on A+B, patients on sorafenib, and patients off treatment; ii) a pre- and post-progression approach where also three sets of utility scores were derived for progressed-free patients on A+B, progressed-free patients on sorafenib, and progressed patients; iii) a pre- and post-progression including AE disutility approach where the analysis performed in ii included a covariate for grade 3 or higher AEs; and iv) a proximity to death approach where a mixed linear model was constructed using treatment status and proximity to death as covariates.

The utility values produced by the four approaches are presented in Table 18. The time to death approach was used in the company's base case assuming that it was the most relevant to the population under consideration reflecting the decline in HRQoL of cancer patients as they approach death.

It is noted that a utility value of 0.78 would be associated with that of a population aged 60 years. In response to clarification question B34 the company stated that "*Assuming a higher utility for patients on treatment more than 15 weeks from death than for the IMbrave 150 age-matched general population is a plausible assumption. This is because the age-matched general population are also composed of observations that are closer to the death of the patient, which have a negative impact on the general utility level (the brazier regression does not include any time coefficients that may control for this effect). This means that a patient who is more than 15 weeks from death can have a higher utility than the general population average.*" The ERG was not convinced that, on average, this would be correct.

Additionally, the ERG noted that in the time to death approach, the difference in midpoint utility estimate increases as the proximity to death increases, with a large difference when a patient is within 5 weeks of death. The ERG does not know whether this is a true finding, although comments that the duration of time associated with this utility is relatively small.

Table 18: Utility values produced by each of the four methods

Category	Utility (95% CI)	Utility (95% CI)
On/Off treatment		
On treatment: A+B	0.79 (0.777, 0.803)	
On treatment: Sorafenib	0.75 (0.734, 0.771)	
Off treatment: Pooled	0.68 (0.666, 0.702)	
Pre- and post-progression		
Pre-progression: A+B	0.78 (0.765, 0.792)	
Pre-progression: Sorafenib	0.77 (0.749, 0.786)	
Post-progression: Pooled	0.74 (0.723, 0.753)	
Pre- and post-progression including AE disutility		
Pre-progression: A+B	0.74 (0.728, 0.764)	
Pre-progression: Sorafenib	0.72 (0.695, 0.744)	
Post-progression: Pooled	0.72 (0.700, 0.735)	
Time to death approach	On Treatment	Off Treatment
≤ 5 weeks before death	0.64 (0.573, 0.713)	0.37 (0.303, 0.430)
> 5 & ≤ 15 weeks before death	0.73 (0.702, 0.759)	0.62 (0.572, 0.658)
> 15 & ≤ 30 weeks before death	0.78 (0.750, 0.805)	0.66 (0.585, 0.722)
> 30 weeks before death	0.80 (0.763, 0.834)	0.71 (0.607, 0.816)

The company's model additionally allowed for a scenario analysis adjusting utilities according to age as per Ara and Brazier,[34] although this did not have a marked impact on the ICER.

4.2.5.5 Resources and costs

The costs and resource use included in the base case model comprised: drug acquisition costs; post-discontinuation subsequent therapy costs; drug administration costs; medical resource use (MRU) associated with progression status; AE costs; and end of life care costs. These are discussed in the following sections.

4.2.5.5.1 Drug acquisition costs

Atezolizumab is available as a 1,200 mg vial at a price of £3807.69 ([redacted] when incorporating the Patient Access Scheme (PAS) discount). Bevacizumab is available in two vial sizes; 400 mg and 100 mg vial at a cost of £924 and £242.66 respectively ([redacted] and [redacted] when incorporating the PAS discount). The costs for A+B are planned to be incurred every 3 weeks. Sorafenib is available in packs of 112 x 200 mg tablets (this represents a supply of 28-days) at a cost of £3,576.56, whereas lenvatinib is supplied as a package of 30 x 4 mg capsules (this represents 15 days' supply for patients weighing <60kg and 10 days' supply otherwise) at a cost of £1,437. As requested by NICE, the company's economic model did not include the PAS discounts for either sorafenib or lenvatinib. The results when these PASs are included are contained in a confidential appendix. All costs were sourced from the British National Formulary (BNF).[24]

Atezolizumab has a fixed IV dose of 1,200 mg, whereas bevacizumab is weight-based, being administered as 15 mg/kg every 3-week cycle. Sorafenib is administered at a dose of 400 mg twice daily, whilst lenvatinib is given at a dose of 8 mg if a patient weighs less than 60kg or 12 mg once daily otherwise.

Following the clarification process, three dosing approaches were considered in the company's economic model.

(i) The base case approach used individual patient characteristics to calculate the planned dose per patient and compute the mean dose per the whole IMbrave150 population

(ii) The second approach used the mean actual dose, considering patient characteristics (mainly weight) and the dose intensity observed in the in IMbrave150. ([redacted] (atezolizumab), [redacted] (bevacizumab), and [redacted] (sorafenib) as reported in the CSR [3]

(iii) The third approach (planned mean dose) replicated the second approach but assumed 100% relative dose intensity (RDI).

For atezolizumab, the base case approach yielded a mean of one vial per patient every three weeks - the third approach provided the same result. Applying the RDI of atezolizumab in the actual dosing approach gave a mean of 0.951 vials every three weeks. The ERG believes that using the reduced RDI (approach (ii) is reasonable given that dose modification was not allowed in the study.

For bevacizumab, the base case approach generated a mean planned dose of 1076.16 mg (2.43 400 mg vials plus 1.49 100 mg vials). Applying the RDI of bevacizumab generated an actual dose of 1047.34 mg (2.27 400 mg vials plus 1.39 100 mg vials) in the second approach. The third approach used the mean weight of IMbrave150 cohort (71.74 kg) and resulted in the same dose as in the first approach albeit using a slightly different number of vials (2 400 mg vials plus 3 100 mg vials) as calculations were done on the mean dose requirements and not individual dose requirements.

For sorafenib, the base case approach yielded a mean daily dose of 400 mg which was the same as in the third approach. In the second approach the actual mean dose was 335.2 mg when the sorafenib-specific RDI was applied. For sorafenib the RDI approach may be less reasonable as it assumes that all reductions in RDI were due to planned reductions rather than patients not being able, or forgetting, to take a tablet intermittently.

For lenvatinib, the base case approach resulted in a planned mean daily dose of 11.10 mg (equating to 2.77 tablets), which was the same as for as the second approach given that the company assumed an RDI of 1 for lenvatinib. However, the third approach considered only the mean weight of the population,

which was above 60 kg, leading to a mean daily dose of 12 mg equivalent to 3 tablets per day. When asked about the assumption of an RDI of 1 for lenvatinib, clarification question B26, the company responded that that the "*RDI for lenvatinib was assumed to be 1.00 in the absence of trial information. The RDI for lenvatinib can easily be updated to equal the sorafenib RDI (83.8%), or alternatively a value of 88% can be applied, sourced from Kudo et al*.[18]" As with sorafenib, using an RDI-based approach for lenvatinib would be less reasonable than for vial-based treatments such as atezolizumab.

The company assumed that 5% of patients shared vials although did not provide justification for this value. In addition, the company assumed that a vial would not be opened if the patient requires less than 5% of its content, in line with NHSE bevacizumab dose banding table. Clinical advice to the ERG suggested that this was appropriate. Scenario analyses undertaken by the company suggested that the ICER would increase slightly if these assumptions were removed. Given the small increase in the ICER these issues have not been explored further by the ERG.

Table 19 summarises the total drug acquisition costs for all the comparators every 3-week cycle using the three dosing approaches and assuming that a vial would not be opened if a patient required less than 5% of it. Drug acquisition costs were applied to the TTD distributions every three weeks for A+B and were adjusted for sorafenib and lenvatinib to be applied every week.

Table 19: Drug acquisition costs every 3-weeks using the three dosing approaches considered in the company's model and assuming a vial would not be opened if a patient required less than 5% of it

Comparator	Approach 1: Planned individual patient dosing (base case)	Approach 2: Actual mean dosing	Approach 3: Planned mean dosing
Atezolizumab	[redacted]	[redacted]	[redacted]
Bevacizumab (5% vial sharing)	[redacted]	[redacted]	[redacted]
Bevacizumab (no vial sharing)	[redacted]	[redacted]*	[redacted]
Sorafenib	£2,682.42	£2,247.87	£2,682.42
Lenvatinib	£2,790.81	£2,790.81	£3,017.70

* The resulting discrepancy (£609.13 being lower than £609.28 with vial sharing) is because the company did not amend all model sheet cells following their response to clarification questions

4.2.5.5.2 Post-discontinuation subsequent therapy costs

Upon discontinuation of A+B or sorafenib, patients in IMbrave50 were allowed to receive a range of subsequent therapies. These are presented in Appendix 1. A summary of these subsequent-line therapies is provided in Table 20.

Table 20: Subsequent therapies observed in IMbrave150 (reproduced from Table 44 of the CS)

	Sorafenib n=165	A+B n=336
Number of patients with at least 1 systemic treatment	[redacted]	[redacted]
Therapy type		
Tyrosine Kinase Inhibitors	[redacted]	[redacted]
Angiogenesis Inhibitors	[redacted]	[redacted]
Immunotherapy	[redacted]	[redacted]
Chemotherapy	[redacted]	[redacted]
Others	[redacted]	[redacted]

Currently, the only recommended second-line therapy in the UK is regorafenib, which is only recommended for use after sorafenib. This means that patients on A+B would not receive expensive subsequent therapies, and those who received sorafenib would only receive regorafenib. Therefore, IMbrave150 was not in line with UK clinical practice. The company performed a Cox regression analysis to examine how subsequent therapies administered in IMbrave150 affected the OS on both treatment arms. The detailed results produced by the Cox regression analyses were not provided by the company. It is unclear whether there was sufficient follow-up post-progression or events within the study for differences in underlying survival rates to be observed. Consequently, the survival functions estimated from the data in IMbrave50 might not reflect the survival functions that might have been estimated if patients had been treated according to UK clinical practice.

The company initially explored three approaches to account for second-line treatment in the economic model. The OS models were not changed for any change in assumptions related to subsequent treatment. Regorafenib is supplied as 84 x 40 mg tablets at a cost of £3,744 at list price. It is administered at a dose of 160 mg once daily for 21 days every 28-day treatment cycle as was assumed to be taken for 13.3 weeks based on IMbrave150 data. The PAS for regorafenib is excluded from the company analysis, as recommended by NICE, but included in the confidential appendix by the ERG. Within all approaches, costs of post-discontinuation subsequent therapies were applied as a one-off cost for patients once they discontinue their first-line treatment.

The first approach, which was used in the company's base case, assumed that the 44.2% of patients on sorafenib who received subsequent systemic treatments would receive regorafenib (and thus incur its associated costs), whereas patients on either A+B or lenvatinib were assumed to receive no further treatments and incur no further treatment costs. This led to a second-line treatment cost of £6,745 for patients on sorafenib.

The second approach altered the proportion of patients who received sorafenib arm who would get regorafenib to be 20%, based on advice provided to the company by clinical experts at an advisory board. This led to a second-line treatment cost for sorafenib patients of £3,052.

The third approach used the IMbrave150 study data to account for immunotherapy (nivolumab) and tyrosine kinase inhibitors (TKIs) administered as a second-line option. For A+B, 1.2% of patients received nivolumab and 18.8% received TKIs; for sorafenib 18.8% of patients received nivolumab and 26.1% received TKIs. TKIs were costed as the weighted mean of sorafenib, lenvatinib, and regorafenib. Further details are provided in Tables 46 and 47 in the CS, although there is a typographical error in Table 46 of the CS as the duration of TKI therapy after A+B was 14.4 weeks, not 13.4 weeks (Clarification response C5). This approach resulted in subsequent treatment acquisition costs of £1,641 for A+B and £6,152 for sorafenib. Patients on lenvatinib were assumed to have the same costs of subsequent therapies as for patients on A+B; based on the use of subsequent treatments in REFLECT, which provides information relating to the relative efficacy of sorafenib and lenvatinib, this assumption appears reasonable.

Following the clarification process the company undertook, at the request of the ERG, three further analyses relating to the use of subsequent treatments. These were: assuming that the full costs for subsequent treatments were costed for both A+B and sorafenib; using statistical analyses to adjust OS removing treatments not recommended in England, and not including costs for subsequent treatments. The ERG acknowledges that the third scenario is not plausible but might provide useful information to the committee in understanding the sensitivity of the ICER to assumptions related to subsequent treatments. The ERG comments that no details were provided on the statistical methods used to adjust OS for removing treatments and as such, the results of this analysis should be treated with caution.

4.2.5.5.3 Drug administration costs

The costs for IV administration were sourced from NHS Reference costs 2018-19 (codes SB14Z (for A+B) and SB15Z (for nivolumab, when costed). Oral chemotherapy delivery costs were sourced for sorafenib, lenvatinib (code SB11Z).[25] A+B administration costs (£371) were applied every three weeks

while nivolumab was provided every two weeks with administration costs of £332. For sorafenib and lenvatinib, administration costs (£195) were applied only once in the first cycle of the model.

4.2.5.5.4 Medical resource use associated with progression status

MRU costs included visits to different health care practitioners, various laboratory tests and scans, and hospitalisation. In the company's base case MRU data were estimated based on consultation with 6 UK clinical experts treating patients with HCC. Experts gave their views on the proportion of patients in need of a resource with the associated frequency per month required. MRU varies according to progression status, and MRU was estimated separately for progression-free and progressed health states. Table 50 of the CS shows the elicited values for MRU in the company's base case.

The company explored a scenario analysis where MRU estimates were sourced from lenvatinib NICE submission[28] as was shown in Table 51 of the CS. Unit costs were estimated using NHS Reference costs 2018-19[35] and Personal Social Services Research Unit 2018-19[27] as presented in Table 49 in the CS. Table 21 summarises the MRU weekly costs applied in the economic model.

Table 21: MRU costs per week used in the economic model

Approach	Base case (company's expert elicitation)	Scenario (lenvatinib submission data)
MRU costs/week for progression-free patients	£129.91	£137.52
MRU costs/week for progressed patients	£131.07	£299.14

4.2.5.5.5 AE costs

The rationale and frequency for the AEs included in the model is provided in Section 4.2.5.2.3. The costs associated with each AE were primarily sourced from NHS Reference Costs 2018-19.[25] Table 53 in the CS presents the costs associated with the management of a single occurrence for each AE. This resulted in mean weekly costs of £4.68, £11.62 and £19.10 to resolve AEs associated with A+B, sorafenib, and lenvatinib, respectively for patients whilst on treatment. No costs were considered for AEs after cessation of the first treatment, which is likely to underestimate AEs.

4.2.5.5.6 End of life care costs

End of life care costs (health and social care costs) for HCC patients reported within Georghiou and Bardsley were considered in line with lenvatinib NICE submission.[26, 28] Table 52 in the CS provides the itemised costs and the inflated costs that were derived for the model using Hospital and Community Health Services (HCHS) indices.[36] This resulted in a one-off cost of £8,186 which was applied for all patients upon entry to the 'Dead' health state.

4.2.6 Model validation and face validity check

The company validated its economic model using two approaches. The first was via *"a number of UK clinical experts"* who validated the key aspects and assumptions of the model. The second approach was an internal quality control of the company's model by a third party.

4.2.7 Cost effectiveness results

Following the clarification process, the company submitted a revised version of the model that included updated estimates of the cost-effectiveness of A+B. All the results presented in this section and in Section 4.2.8 use the revised model and include the increased PAS for atezolizumab and the list prices for sorafenib, lenvatinib, and regorafenib. A confidential appendix presents the same results with the PAS considered for all of the five treatments.

Table 22 shows the results of the company's base case analysis based on the deterministic and probabilistic versions of the company's revised model. The probabilistic sensitivity analyses (PSA) results are based on 2,000 iterations run by the ERG. Based on the probabilistic version of the model, A+B is expected to generate [redacted] additional QALYs at an additional cost of [redacted], compared with sorafenib. The corresponding ICER is £22,419 per QALY gained. The deterministic version of the company's model produces a similar ICER of £22,267 per QALY gained. A+B dominated lenvatinib, generating [redacted] more QALYs at a reduced cost of [redacted] based on the probabilistic version of the model. Figure 9 shows the cost-effectiveness acceptability curves (CEAC) for all three options based on a re-run of the PSA by the ERG. Figure 10 plots the PSA results on the cost-effectiveness plane. Figure 11 presents the resultant survival functions for the first 12 years of the company's model.

Table 22: Company's base case results

Treatment	Total QALYs	Total Costs	Incremental QALYs	Incremental costs	ICER (£ per QALY gained)
Deterministic					
Sorafenib	1.05	£44,983	-	-	
Lenvatinib	1.13	£62,580	[redacted]	[redacted]	Dominated
A+B	[redacted]	[redacted]	[redacted]	[redacted]	£22,267
PSA (run by the Evidence Review Group)					
Sorafenib	1.05	£45,002	-	-	
Lenvatinib	1.19	£63,557	[redacted]	[redacted]	Dominated
A+B	[redacted]	[redacted]	[redacted]	[redacted]	£22,419

A+B, atezolizumab plus bevacizumab; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Figure 9: Company's base case cost-effectiveness acceptability curves

Figure 10: Company's base case cost-effectiveness plane

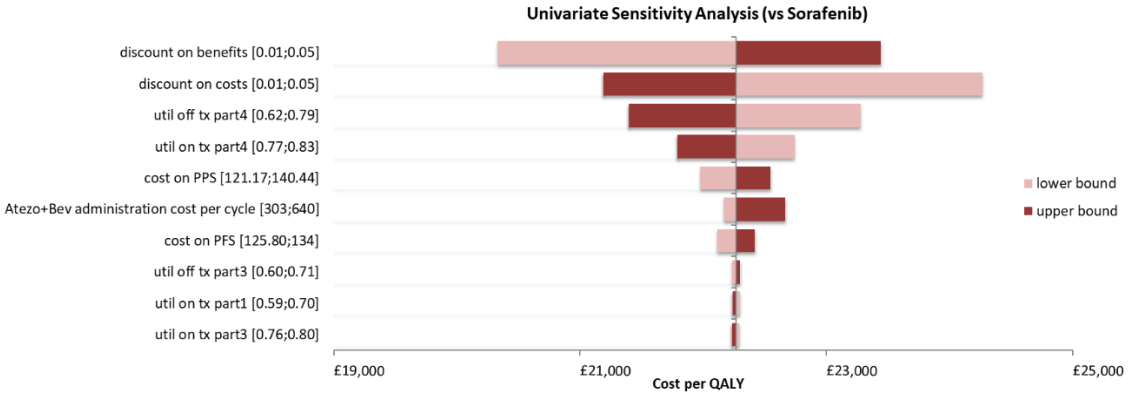
Figure 11: Company's base case survival functions (model traces) – A+B (top left), sorafenib (top right) and lenvatinib (bottom)

4.2.7.1 Tornado diagrams

A tornado plot showing the ten most influential parameters in terms of impact on the ICER of A+B and sorafenib is presented in Figure 12. Within this analysis, all parameters were varied between the upper and lower bounds of the 90% percentile values obtained from the distributions used in the PSA. If such distributions were not available, the parameter was varied by ±20%. The tornado plots were reproduced by the ERG from the revised model. The ERG notes that the relative efficacy for sorafenib and lenvatinib were not included in the tornado diagram which is a limitation given that the CrIs suggest that lenvatinib could be more efficacious than A+B (see Section 3.4).

The most influential parameters, of those explored, on the ICER of A+B versus sorafenib were related to the discount rates applied and utility values used for patients with more than 30 weeks to die. None of the ICERs on the tornado plot exceeded £25,000 per QALY gained. A+B dominance of lenvatinib remained for all parameter changes and thus a tornado diagram has not been presented.

Figure 12: Tornado diagrams of A+B versus sorafenib



4.2.8 Sensitivity analyses

The company conducted a range of scenario analyses, which included the effects of using alternative survival models and parameter inputs on the results.

4.2.8.1 Scenario and subgroup analyses

The ERG updated the results of the scenarios outlined in Table 57 of the CS and added those conducted by the company in response to clarification questions. These are provided in Table 23.

Table 23: The company's scenario analyses results

No.	Scenario	ICER versus sorafenib	ICER versus lenvatinib
Base case		22,267	A+B dominant
1	5-year time horizon	24,470	A+B dominant
2	10-year time horizon	22,531	A+B dominant
3	15 - year time horizon	22,296	A+B dominant
4	Atezo+Bev OS - Generalised Gamma	18,657	A+B dominant
5	Atezo+Bev OS - Log-logistic distribution	21,443	A+B dominant
6	Sorafenib OS - Generalised Gamma distribution	24,054	A+B dominant
7	Sorafenib OS - Log-logistic distribution	28,073	A+B dominant
8	Atezo+Bev PFS – Exponential distribution	22,283	5,950
9	Atezo+Bev PFS - Log-logistic distribution	22,261	A+B dominant
10	Sorafenib PFS – Exponential distribution	22,266	A+B dominant
11	Sorafenib PFS - Log-logistic distribution	22,271	A+B dominant
12	Atezo TTD – Exponential distribution	21,029	A+B dominant
13	Atezo TTD – Weibull distribution	29,111	4,625
14	Discount rate – costs - 0%	25,153	A+B dominant
15	Discount rate – costs - 5%	21,192	A+B dominant

16	Discount rate – effects - 0%	19,565	A+B dominant
17	Discount rate – effects - 5%	23,441	A+B dominant
18†*	Stopping rule – Yes	15,827	A+B dominant
19	Treatment duration: Until progression	16,465	A+B dominant
20	Dose: Planned ind. dose without vial sharing	22,299	A+B dominant
21	Utilities: IMbrave150 (On/Off treatment)	22,303	A+B dominant
22	Utilities: IMbrave150 (Off/On progression)	22,532	A+B dominant
23	Utilities: IMbrave150 (Off/On progression)+ AE3+	23,078	A+B dominant
24	Modelling sorafenib using HRs from ITC rather than IMbrave150 study data	21,376	A+B dominant
25	Resource use estimates: TA551[28]	27,516	6,351
26	Subsequent therapy: IMbrave150 study data	23,064	A+B dominant
27	Subsequent therapy: Sorafenib arm only receive regorafenib (clinical expert opinion)	27,626	A+B dominant
28†	Modelling sorafenib and lenvatinib: Fractional polynomial NMA	21,813	A+B dominant
29†	Lenvatinib dose prescribed assuming all patients weigh ≥ 60kg	22,267	A+B dominant
30†	Lenvatinib dose prescribed assuming all patients weigh < 60kg	22,267	16,391
31†*	All patients receive full recommended dose of bevacizumab (i.e. no vial use threshold)	22,308	A+B dominant
31†	Subsequent therapy: IMbrave150 study data (all treatments are costed in)	A+B dominant	A+B dominant
32†	OS adjusted excluding subsequent treatments not recommended in England	20,307	A+B dominant
33†	Subsequent therapy: no costs are applied	32,054	A+B dominant

A+B, atezolizumab plus bevacizumab; AE, adverse event; HR, hazard ratio; ICER, incremental cost-effectiveness ratio; ITC, indirect treatment comparison, NMA, network meta-analysis; OS, overall survival; PFS, progression free survival; TA, technology appraisal, TTD, time to treatment discontinuation

†Reported in the clarification response

*The ERG could not reproduce the results reported by the company

Most scenarios produced ICERs that were similar to the company's base case ICER. All ICERs for A+B versus sorafenib were less than £30,000 per QALY gained, except when subsequent therapies were assumed to have zero cost (Scenario 33).

Similarly, A+B dominated lenvatinib in all but four scenarios; Scenario 8, where the PFS of A+B (and also lenvatinib because of the use of a HR) is decreased as PFS was used as a proxy for TTD for lenvatinib; Scenario 13, where the TTD of A+B was increased; Scenario 25 where resource use associated with TA 551 was used; and Scenario 30, where all patients weigh < 60kg and the cost per patient of lenvatinib is reduced. However, none of the resulting ICERs were above £20,000.

Additionally, the ERG asked for a subgroup analysis for IMbrave150 population excluding patients from Asia (except Japan); the results of this analysis are presented in Table 24. The rationale for this request was that in the NICE appraisal of sorafenib,[37] data from the SHARP study[38], in which 70% of patients were European, was preferred to that from the Asia Pacific study[39] which recruited patients from China, Korea and Taiwan and where there was endemic HBV.

Table 24: The company's subgroup results for IMbrave150 population excluding Asia (except Japan)

Option	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
A+B	[redacted]	[redacted]	-	-	-
Sorafenib	44,802	1.05	[redacted]	[redacted]	22,368
Lenvatinib	59,103	1.04	[redacted]	[redacted]	Dominant

A+B, atezolizumab plus bevacizumab; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

4.3 Critique of company's submitted economic evaluation by the ERG

4.3.1 Methods for reviewing the company's economic evaluation and health economic model

The ERG adopted a number of approaches to explore, interrogate and critically appraise the company's submitted economic evaluation and the underlying health economic model upon which this was based. These included:

- Scrutiny of the company's model and discussion of issues identified amongst the members of the ERG.
- Verification of the implementation of the company's model.
- Examination of the correspondence between the description of the model reported within the CS and the company's executable model.
- Re-running the scenario analyses and PSA presented within the CS.
- Where possible, checking the parameter values used in the company's model against their original data sources.
- The use of expert clinical input to judge the credibility of the company's economic evaluation and the assumptions underpinning the model.

4.3.2 Adherence of the company's model to the NICE reference case

Table 25 compares the company's economic evaluation with the NICE reference case.[22]

Table 25: Adherence of the company's model to the NICE reference case

Element	Reference case	ERG comments
Type of economic evaluation	Cost-utility analysis with fully incremental analysis	The model is in line with the NICE reference case.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	The company used a time horizon of 20 years which was sufficiently long to meet the NICE reference case. The number of patients alive in the model at 20 years was effectively zero.
Synthesis of evidence on health effects	Based on study outcome data and systematic review	The company used data directly drawn from IMbrave150 to model the relative effectiveness of A+B and sorafenib. An ITC, albeit with some limitations, was undertaken to assess the relative effectiveness of lenvatinib with A+B and sorafenib.
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL in adults	The model uses the EQ-5D measure as preferred in the NICE reference case.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	EQ-5D-5L data reported by patients were collected in IMbrave150. This meets the NICE reference case.
Source of preference data for valuation of changes in HRQoL	Representative sample of the UK population	The company followed advice by NICE regarding mapping of the EQ-5D-5L to 3L and meets the NICE reference case. The company adopted a '*proximity to death*' approach which involved the use of a mixed linear model which had theoretical limitations (See Section 4.3.4.4).
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	The company's model is in line with the NICE reference case, although the company makes a case for the end of life criteria being met.
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	The company's model is in line with the NICE reference case.
Discount rate	The same annual rate for both costs and health effects (currently 3.5%)	The company's model is in line with the NICE reference case.

4.3.3 ERG Critique of the modelling performed by the company

4.3.3.1 Model verification

The ERG checked and verified the implementation of the model and the methods for generating results in the model submitted after the clarification process. Only small errors were identified which are detailed in Section 4.3.4.1.

The ERG could not replicate the scenario analyses on two occasions (see Table 23) for unknown reasons.

4.3.3.2 Correspondence of the model inputs and the original sources of parameter values

The ERG is satisfied that, where checked, model parameters corresponded with their original source values. These were in line also with the parameter values reported in the CS.

4.3.4 Issues identified from the ERG's critical appraisal

Box 1 summarises the issues identified within the company's health economic model. These points are discussed in the following subsections. Where possible, the ERG has performed exploratory analyses as described in Section 4.4 with the impact on the ICER being provided in Section 5.

Box 1: Summary of the issues identified within the company's health economic model

- Perceived modelling errors
- Limitations in the estimation of time-to-event data and choice of distribution used to estimate OS
- Actual dosage not considered in the company's base case
- Insufficient wastage of oral chemotherapy considered
- Inappropriate use of utility values for patients with unresectable HCC that were greater than average age and gender-matched patients
- Using a naïve indirect comparison to estimate AEs associated with lenvatinib
- Underestimating the relative efficacy of lenvatinib
- Lack of details associated with the analyses removing treatments not recommended by NICE from IMbrave150

4.3.4.1 Perceived modelling errors

The ERG identified three modelling errors in the calculations for bevacizumab dose after receiving the company's response to clarification questions. The first error was related to question B18. The company amended the formulae in cells Q30:S30 in the 'Dosing' worksheet which related to calculations for bevacizumab without vial sharing. However, the same formulae in cells U30:V30 were not amended. This resulted in the discrepancy shown in Table 19, whereby the cost of the 'vial sharing' approach was more than that of the 'no vial sharing' approach.

The second error was related to calculating the number of bevacizumab number of vials needed per patient. In its response to clarification question B16, the company did not account for the 5% vial use threshold stated in response to question B5. Accordingly, the model calculations assumed that a patient gets a vial even if they require less than 5% of the vial's amount which is not in line with NHSE bevacizumab dose banding table.

4.3.4.2 Limitations in the estimation of time-to-event data and choice of distribution used to estimate OS.

The ERG considers the following issues to be limitations with the survival modelling:

- It is inappropriate to estimate the relative treatment effects of lenvatinib versus A+B using an HR from a random effects NMA, whilst estimating the effect of sorafenib versus A+B using arm-based parametric survival models

- There is no reason to assume proportional hazards for the comparison of lenvatinib versus A+B and to do so when hazards are not proportional will generate a biased estimate of the lenvatinib absolute survival function and population mean survival.
- Furthermore, the Cox HR from REFLECT would not have the same numerical value as a hazard ratio that would be estimated by fitting a parametric model to both treatment arms. If it is believed that a particular parametric model correctly represents the underlying data-generation process and proportional hazards is accepted, then there is no reason to not use the parametric model to estimate the relative treatment effect.
- Overlaying the HR from one analysis onto a baseline arm from a different analysis will overstate the uncertainty in the analysis because the covariation between baseline and treatment effect that would be expressed in a single coherent analysis is lost, resulting eventually in an incorrect characterisation of the uncertainty in incremental net benefit.
- When the proportional hazards assumption does not hold, then the HR obtained from a Cox model is meaningless, and applying it to a parametric form for the baseline survival function will give incorrect inference.
- Not all parametric distributions, such as the log-logistic distributions, can be parameterized as a proportional hazards model and yet the company generated survival functions for lenvatinib in such cases.
- The company used a random effects NMA to estimate the relative effect of lenvatinib versus A+B but used a different model to estimate the effect of sorafenib versus A+B, effectively treating the effect of sorafenib versus A+B as fixed. Consequently, the uncertainty associated with the two estimates of treatment effect is modelled differently and allows for greater uncertainty for the relative effect of lenvatinib than for the relative effect of sorafenib.
- The underlying hazard functions may vary between treatments for various reasons including differences in the mechanism of action of treatments. Nevertheless, the company has sought to find a single standard parametric model that can be used to model the time-to-event data for each treatment. In response to clarification question A11 the company stated, "*More flexible survival models such as spline models were not deemed necessary due to the good fit of the data to standard distributions. More flexible survival models are more commonly used when the survival curves do not follow a specific distribution and the data is slightly more complex to extrapolate. That was not the case when fitting the IMbrave150 data or the ITC data to parametric distributions*." The ERG notes that no standard parametric distribution is likely to be the true model and that interest is not only in identifying a model that is a reasonable representation of the sample data but also one that provides plausible predictions. While the ERG is aware that spline models

generally require many events to estimate parameters and make assumptions about the extrapolation phase, they do provide a way of relaxing the assumption of a single underlying standard parametric model for each treatment. In response to clarification question A15, the company justified their choice of model for the data based on the shape of the survival functions rather than the underlying hazard functions and stated that "*the tail of the curve OS would naturally differ due to the different mechanism of action between an immunotherapy and a TKI, as one may expect a prolonged tail of patients who have not relapsed on Atezo+Bev therapy. Nevertheless, it was agreed that as extrapolated survival curves are based on assumptions, it is justified to assume the same hazard function for all treatments, and it was deemed a conservative approach by the clinical community*."

- At the clarification stage, the ERG asked the company to perform network meta-analyses allowing for time varying treatment effects (i.e. not necessarily HRs) for different survival models including all treatments of interest and referred the company to methods proposed by Ouwens *et al*. for standard parametric models[40] and Jansen for fractional polynomials.[41] The methods require the same reconstruction of KM product limit estimates as sample data from KM survival functions. In response, clarification question A7, the company used a first-degree Bayesian fixed effect FP NMA to analyse OS data. However, the company did not present fitted survival functions or make an assessment of the relative goodness-of-fit of this model to the original models, and did not comment on the clinical plausibility of the extrapolations. In addition, it is not clear why the company used a fixed effect model rather than a random effects model as it did in the original NMA. Consequently, the ERG is not able to assess the credibility of this analysis.
- In response to clarification question A7, the company stated that "*the Ouwens approach was unfeasible, as that also required individual patient level data.*" It is not clear to the ERG why the company was not able to fit time-varying models as described by Ouwens *et al* given that it used the same likelihood function for FPs.
- If the company had performed random effects meta-analyses allowing for time-varying treatment effects and FPs, then the appropriate input to the economic model would be samples from the posterior predictive joint distribution of the effects of treatments in a new study; this would generate greater uncertainty than has currently been allowed for.
- At the clarification stage, and in response to clarification question A7, the company stated that the "*FP NMA was unattainable for PFS due to the different methodologies in data collection in the REFLECT and Imbrave150 trials, mRECIST and RECIST 1.1, respectively."* The ERG notes that this issue also applies to its base case using HRs.
- The company claims that *"the exponential model and the Generalised Gamma model represented clinically plausible estimates, as the remaining four models projected a higher OS for sorafenib*

than lenvatinib, which is not aligned with the REFLECT trial clinical data which showed lenvatinib OS to be non-inferior to sorafenib." This ignores several important issues that the ERG believe affects the company's choice of survival model:

  - Non-inferiority of lenvatinib versus sorafenib in REFLECT was judged according to an HR estimated over an approximately 40-month study period assuming no treatment by time interaction. While an HR is a convenient summary in clinical trials, survival functions may not be proportional in practice and they are not particularly relevant in HTA allowing for the observed and extrapolated (i.e. approximately another 40 months in the case of lenvatinib based on Figure 9 of the CS) periods.
  - Non-inferiority of proportional hazards model means that the true lenvatinib survival function could be worse than sorafenib. Indeed, the conclusion regarding non-inferiority was based on a 95% CI for the HR (0.79, 1.06), an average treatment effect over the observed study period, which suggests that lenvatinib could be worse than sorafenib, and even worse than indicated by the upper limit of the 95% CI.

- Patients in IMbrave150 and REFLECT were not treated according to UK clinical practice post--progression. Consequently, the survival functions estimated from the data in IMbrave50 and REFLECT might not reflect the survival functions that might have been estimated if patients had been treated according to UK clinical practice. This may lead to biases in survival in favour of A+B and lenvatinib because subsequent therapies not approved in UK clinical practice were allowed in IMbrave150 and REFLECT, respectively, whereas patients receiving sorafenib can receive regorafenib, if sufficiently fit.
- The ERG has concern with the process use to elicit judgements of experts:
  - The process is not transparent, including a lack of clarity regarding the evidence that each expert was familiar with and the questions that were asked.
  - The ERG is concerned that experts' judgements are sought "*to ensure the curves represented outcomes seen in UK clinical practice*". It is the opinion of the ERG that this could be misleading. The survival functions estimated in the studies reflect the mix of patients defined by the inclusion/exclusion criteria of the studies and not necessarily the target population when treatments are used in clinical practice.
  - The ERG notes that clinical trials provide a sample estimate of a survival function not the true survival function. We expect sampling variation and differences in the central estimates of survival functions in different studies even if a study is repeated under identical circumstances and the clinicians are not required to account for sampling variation or parameter uncertainty when expressing their judgements.

An additional potential problem with the use of an HR for lenvatinib is that it has been assumed to be maintained throughout the time horizon of the model and does not take into account the actual time on A+B and lenvatinib treatment.

In accordance with usual practice, the baseline survival function for the target population is taken from the A+B arm of the IMbrave150 study. The ERG notes that the mix of patients in clinical trials may be different to that in clinical practice and the two baseline survival functions may differ. This would affect the expected absolute survival functions for all treatments and the estimate of incremental survival, although it is not possible to say whether this would be smaller or larger in this case.

4.3.4.2.1. Estimation of OS survival functions

Refer to Section 4.3.4.2 for detailed comments.

The company choose to model OS using an exponential distribution despite it not being one of the better fitting models, and it being associated with a constant hazard of death for the lifetime of patients and a constant treatment effect. The ERG was not convinced by the reasons given by the company for discarding other survival model and undertook further modelling using log-normal, generalised gamma and log-logistic models.

4.3.4.2.2. Estimation of PFS survival functions

No exploratory analyses were undertaken by the ERG relating to PFS as this was shown to only have a small impact on the ICER in the company's scenario analyses and the company's choice of base case model appeared appropriate.

4.3.4.2.3. Estimation of TTD survival functions for A+B and sorafenib

The company fitted the same set of models to TTD data that it used to analyse OS and PFS data. The ERG does not know whether the use of the 14-month cut-off point for switching between the KM survival function and the parametric distribution is optimal, but changes to this time point does not change the ICER markedly.

4.3.4.3 Actual dosage not considered in the company's base case

The company provided three analyses relating to the dosage (and costs) of interventions as shown in Table 19. The ERG believes that in addition to Approach 2, which considers the actual RDI used in IMbrave150, another informative scenario would be to use the RDI for atezolizumab, which is vial-based but planned dosage for sorafenib, as it is plausible that savings on unused tablets are not recouped (termed Approach 2b). The correct dosages for decision making are likely to lie between Approach 2 and Approach 2b.

Further, the company assumed an RDI of 1 for lenvatinib "*in the absence of trial information.*" (Clarification question B26). The company stated that the model could be updated with the RDI observed for sorafenib (0.84) or with the value from the REFLECT study (0.88)[18] but did not perform these analyses.

4.3.4.4 Insufficient wastage of oral chemotherapy considered

The company model assumes acquisition costs for both sorafenib or lenvatinib are incurred in each weekly cycle and uses the half-cycle corrected proportion for patients on treatment to calculate acquisition costs. Wastage that occurs from patients discontinuing or dying with pills dispensed has therefore not been considered. In contrast, drug wastage was considered in the STAs of sorafenib and lenvatinib. where the Appraisal Committee for sorafenib suggested that the most plausible ICER should account for drug wastage for up to 7 days.[28, 37]

4.3.4.5 Inappropriate use of utility values for patients with unresectable HCC that were greater than average age and gender-matched patients

The utility used in the company base case (Table 15) indicate that for patients on treatment the utility could be 0.78 or higher for those who were further than 15 weeks from death. For reference, the utility associated with the age- and sex-matched population is [redacted].[34] The ERG believes it extremely unlikely that, on average, patients with unresectable HCC have a higher utility than an age- and sex-matched population.

Additionally, the ERG noted that the models used by the company to map EQ-5D-5L data to EQ-5D-3L data make the assumption of normality, effectively appealing to the Central Limit Theorem. In response to clarification question B32, the company stated that this is standard practice and estimates will be unbiased. However, the ERG notes that these data are multimodal, right bounded at 1 with a substantial gap to the next set of observations and also left bounded. These features present significant statistical challenges and it is known that standard approaches do not perform well for this reason; using standard approaches therefore introduces the possibility of bias.[42] With incremental QALYs forming the denominator of the ICER calculation, small adjustments to small values can result in significant and sometimes decision altering changes.[43] More nuanced models, such as adjusted limited dependent variable mixture models,[42] are available which address these issues. However, if the health states being considered are all well-populated with data, then a non-parametric calculation of the mean for each state (with appropriate weighting for repeated values from individuals) is also appropriate and simple to implement without introducing bias.[44]

4.3.4.6 Using a naïve indirect comparison to estimate AEs associated with lenvatinib

In response to clarification question A5, the company acknowledged that the method used for estimating the rate of AEs associated with lenvatinib treatment was a naïve indirect comparison and that if a relative frequency had been applied the rate would be lower.

4.3.4.7 Estimating the relative efficacy of lenvatinib

The main issue is that the currently the effect of lenvatinib compared to A+B is estimated inconsistently. In particular, the lenvatinib survival function is generated with respect to an HR from a random effects NMA, whereas the effect of sorafenib relative to A+B is effectively from a fixed effect arm-based comparison of survival functions from IMbrave150.

Using the HR from the NMA to generate the lenvatinib survival function (and mean survival) assumes that the A+B survival function (and mean survival) will be above that for lenvatinib over the lifetime of patients, which may not be true and is an unnecessary modelling assumption. Some indication of whether the effect of lenvatinib relative to A+B is constant over time is provided by the results of the analysis using fractional polynomials (clarification question A8) subject to the limitations discussed in Section 4.3.4.2. This shows a small but consistent increase in the time-varying hazard ratio from 0.647 at month 0.1 to 0.705 at month 75. Nevertheless, these results suggest some shrinkage in the hazard ratio towards one over the lifetime of patients. That said, the ERG is concerned with the apparent inconsistency between the average treatment effect estimated from the random effects NMA (HR of lenvatinib versus A+B: 0.63) and the time-varying HRs estimated from the model using fractional polynomials. There is no assumption of a constant treatment effects when comparing sorafenib and A+B.

Uncertainty is treated differently when comparing lenvatinib with A+B and sorafenib with A+B in that the former is from a random effects model and the latter is effectively a fixed effect comparison. The ERG would prefer to see a single random effects NMA allowing for time-varying treatment effects. The company did present results of a fractional polynomial NMA but this was from a fixed effect model rather than a random effects model and it did not sufficiently critique the results.

For further details refer to Section 4.3.4.2.

4.3.4.8 The lack of details associated with the analyses removing treatments not recommended by NICE from IMbrave150

In clarification question B30, the company were asked to provide a scenario analysis which attempted to use statistical methods to exclude treatments not recommended in England. The company provided

a set of analyses but with no explanation of the method undertaken. As such, the ERG cannot critique these values and believes that these results should be treated with caution. The ERG cannot provide a robust opinion on the likely direction and magnitude of the bias of this scenario.

4.3.4.8 The assumption that oral chemotherapy administration costs are incurred once only

The company assumed that the costs of oral chemotherapy administration (£195) was only incurred once for each patient. Previous comments by NHS England staff suggests that this could be incurred with each prescription. Communication with NICE staff has indicated that the is not a standard NICE position on this.

4.4 Exploratory analyses undertaken by the ERG

This section details the exploratory analyses undertaken by the ERG.

Where possible, the ERG undertook exploratory analyses to address the limitations listed in Box 1. The following two limitations could not be addressed by the ERG within the timescales of producing the report: (1) additional analyses relating to the inconsistent modelling of the relative efficacy of lenvatinib compared with A+B, and (2) assessing the appropriateness of the statistical methods used to attempt to remove the impact of subsequent treatments not recommended in England. Additionally, the ERG undertook exploratory analyses that was also performed by the company to show the impact of these given other changes made by the ERG.

4.4.1 Correction of perceived modelling errors.

The ERG amended the formulae used to calculate the number of bevacizumab vials in the actual dose approach in the 'with vial sharing' cells to be consistent with the 'no vial sharing' calculations. Accordingly, the ERG implemented the same equations of cells R30:S30 in cells U30:V30 in the 'Dosing' sheet.

In addition, the ERG amended the formulae used to calculate the number of vials per patient in the 'planned individual dosing' to take into account the 5% vial threshold. Hence, if a patient required less than 5mg of a 100mg vial, they would not receive it. For example, after implementing the correction, the patient in row 49, who is in need of 903mg of bevacizumab, is correctly receiving 2 vials of 400mg bevacizumab and 1 vial of 100mg bevacizumab (instead of 2 vials of 400mg bevacizumab and 2 vials of 100mg bevacizumab)

4.4.2 Exploratory analyses relating to the estimation of time to event data

Two exploratory analyses were undertaken which are described in the following sub-sections.

4.4.2.1 Exploratory analyses using different assumptions relating to OS for A+B, sorafenib and lenvatinib

The following analyses were undertaken.

- Using the log-normal for both A+B and sorafenib as this was the best fitting model to the observed data when adding BICs across treatment arms
- Using the generalised gamma for both A+B and sorafenib as this was the second-best fitting model to the observed data when adding BICs across treatment arms
- Using the log-logistic for both A+B and sorafenib as this was the third-best fitting model to the observed data when adding BICs across treatment arms

The KM survivor function and the survival estimates produced by the log-normal, the generalised gamma and the log-logistic are shown in Figure 13, Figure 14 and Figure 15 respectively.

Figure 13: Estimates of OS associated with log-normal distributions

Figure 14: Estimates of OS associated with generalised gamma distributions

Figure 15: Estimates of OS associated with log-logistic distributions

The OS function for lenvatinib changed when the A+B model was changed as the company applied an HR to the A+B OS survival function to obtain the OS survival function for lenvatinib. The ERG acknowledges that applying a HR to a baseline survival function imposes an unjustifiable constant treatment effect and that it should not be applied to models that are not proportional hazards models. However, given that the company did not conduct a coherent random effects NMA allowing for time-varying treatment effects, this is this best the ERG could do and enables some assessment of the robustness of results that may be of interest to the Appraisal Committee.

4.4.2.2 Exploratory analyses using different assumptions relating to TTD for A+B, sorafenib and lenvatinib

The ERG undertook analyses varying the time at which the exponential distribution was used rather than the KM survival function from 14 months to 13 months and 15 months. The model did not allow for uncertainty in the KM survival function to be considered.

4.4.3 Exploratory analyses relating to the dosage and acquisition costs of the interventions

4.4.3.1 Exploratory analyses using the RDI for A+B and sorafenib from IMbrave150 and the RDI for lenvatinib from Kudo *et al*.

Analyses were undertaken when the RDI for lenvatinib observed in REFLECT, 0.88, [18] was used.

4.4.3.2 Exploratory analyses using the RDI for A+B from IMbrave150 and the planned dosage for sorafenib

Analysis were performed whereby the RDI observed in IMbrave150 was used for A+B but the planned dosage was used for sorafenib. These were explored as it was not known whether the reduced RDI for sorafenib was planned or due to patients intermittently not taking a tablet.

4.4.4 Exploratory analyses incorporating seven day's wastage of sorafenib and lenvatinib when a patient discontinues treatment.

In order to account for oral chemotherapy wastage, the ERG amended the calculations for the acquisition costs of sorafenib and lenvatinib by taking the patient proportion still on treatment at the start of a given cycle (i.e. columns BV and AC instead of columns BW and AD in 'Sorafenib' and 'lenvatinib' sheets respectively) and multiplying it by the weekly acquisition costs. This accounted for an average of 3.5 days of drug wastage for patients who discontinued through this cycle. A further 3.5 days' worth of drug acquisition costs (£447.07 and £465.14 for sorafenib and lenvatinib respectively based on list prices) were added to discontinuing patients to account for a total to 7 days of drug wastage for the discontinuing proportion.

4.4.5 Exploratory analyses capping the utility of patients with unresectable HCC to that of the age- and sex-matched population.

An analysis was undertaken where it was assumed that the utility associated with patients with unresectable HCC was capped at the age- and sex-matched population value. This was implemented by limiting all utility values not to exceed the age- and sex-adjusted general population utility value calculated from Ara and Brazier.[34]

4.4.6 Exploratory analyses removing AEs for lenvatinib.

To assess the influence that the rate of AEs associated with lenvatinib had on the ICER an extreme analysis was undertaken that assumed there were no lenvatinib-related AEs.

4.4.7 Exploratory analyses assuming that oral chemotherapy administration costs are incurred at every prescription rather than once only.

The ERG has run an exploratory analysis assessing the impact of administration costs being incurred every 28 days for sorafenib and lenvatinib.

4.4.8 Exploratory analyses relating to the costs of subsequent treatments after A+B, sorafenib and lenvatinib

The ERG provided results produced under three alternative assumptions relating to the costs of subsequent treatments which were: costing only TKIs and nivolumab; costing all subsequent treatments; and costing none of the further treatments. The latter two scenarios provide the potential range in the ICER based on extreme scenarios. The first alternative is the ERG's preference, as it explicitly incorporates the most widely treatments that could impact on OS.

4.4.9 Exploratory analyses removing AEs for lenvatinib.

The actual over-estimation of AEs associated with lenvatinib is unknown. To assess the influence that the rate of AEs associated with lenvatinib had on the ICER, an extreme analysis was undertaken that assumed there were no lenvatinib-related AEs. A small change in the ICER would indicate that the over-estimation of AEs associated with lenvatinib was not a key driver of the decision problem.

5 IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

Sixteen modifications to the company's base case were explored. Two ERG base cases are presented both of which include the first four modifications. ERG base case A, adds scenario 5, whilst ERG base case B, adds scenario 6. Together these form the ERG's preferred ICER range.

The ERG performed subgroup analyses based on patient weight for those weighing under 60kg and for those weighing 60kg and over. This was motivated by the fact that the acquisition cost of lenvatinib differs between these subgroups, hence the relative cost-effectiveness of A+B against lenvatinib is influenced by the weight category of the patient. As the ERG did not have data by patient weight it was assumed that only costs were changed and the estimated OS for all patients was generalisable to both weight groups.

95% CIs around the mean probabilistic ICER have been calculated using the method described in Hatswell *et al*.[45] It was seen that 1,000 PSA iterations appeared sufficient to reduce Monte Carlo sampling error.

The ERG ran all results deterministically, whereas probabilistic results using 1,000 iterations were obtained for the ERG's base case and subgroup analyses. A condensed summary detailed exploratory analyses undertaken by the ERG is provided in Table 26. More detailed results are presented in Table 27 and Table 28. The confidential PASs for sorafenib, lenvatinib and regorafenib are not included; these data are considered in a confidential appendix.

In the analyses using the list prices all ICERs for A+B compared with sorafenib and lenvatinib were under £50,000 per QALY gained.

Table 26: Summary of ERG-preferred ICER (cost per QALY gained) ranges for the four scenarios

	All patients in IMbrave150		**Non-Asian plus Japanese patients in IMbrave150**	
	Sorafenib	**Lenvatinib**	**Sorafenib**	**Lenvatinib**
Costs associated with patients weighing under 60kg	£16,567 to £21,843	£83 to £3,962	£15,387 to £21,488	Dominant to £3,381
Costs associated with patients weighing 60kg or more	£21,427 to £26,653	Dominant to Dominant	£20,837 to £27,017	Dominant to Dominant

The ERG additionally explored the impact of applying adjustments 8-16, detailed in Table 29, to the ERG base case with the least favourable ICER for A+B. This base case is 'ERG base case B assuming costs for patients weighing more than or equal to 60kg excluding Asia (except Japan)', in which A+B has a deterministic ICER of £26,525 compared with sorafenib (Table 29**).**

An additional analysis was performed applying adjustments 8-16 to a favourable scenario for A+B, which was 'ERG base case A assuming costs for patients weighing less than 60kg'. The deterministic ICER for A+B compared with sorafenib was £16,296 per QALY gained (Table 28). Results of these analyses are shown in Table 30.

These analyses provide the committee with an indication of the ICER value when assumptions other than that in the ERG's base cases are chosen.

Table 27: The ERG's exploratory model results

Analysis	Discounted costs			Discounted QALYS			ICER (A + B versus sorafenib)	ICER (A + B versus lenvatinib)
	A+B	Sorafenib	Lenvatinib	A+B	Sorafenib	Lenvatinib		
Company deterministic base case	[redacted]	£44,983	£62,580	[redacted]	1.05	1.13	**£22,267**	**A+B dominant**
1) Adjusting for perceived modelling errors	[redacted]	£44,983	£62,580	[redacted]	1.05	1.13	£22,250	A+B dominant
2) Use of log-normal functions to model OS	[redacted]	£47,739	£63,920	[redacted]	1.34	1.23	£22,066	A+B dominant
3) Including 7 days oral chemotherapy wastage on discontinuation	[redacted]	£45,865	£63,491	[redacted]	1.05	1.13	£20,969	A+B dominant
4) Capping utilities for people with unresected HCC at that of the age- and sex-matched population	[redacted]	£44,983	£62,580	[redacted]	1.03	1.10	£23,083	A+B dominant
5) Costing subsequent TKIs and nivolumab treatments	[redacted]	£47,508	£71,600	[redacted]	1.05	1.13	£23,064	A+B dominant
6) Implementing the 'actual dose' approach for A+B and using an RDI of 1 for sorafenib and lenvatinib	[redacted]	£44,983	£62,580	[redacted]	1.05	1.13	£19,849	A+B dominant
7) Implementing the 'actual dose' approach and RDI of 0.88 for lenvatinib	[redacted]	£41,761	£58,176	[redacted]	1.05	1.13	£24,593	£485
8) Use of generalised gamma functions to model OS	[redacted]	£45,765	£63,761	[redacted]	1.13	1.22	£19,537	A+B dominant
9) Use of log-logistic functions to model OS	[redacted]	£47,039	£62,582	[redacted]	1.26	1.09	£26,296	A+B dominant
10) Using MRU costs associated with the STA of lenvatinib (TA551)[28]	[redacted]	£53,654	£68,560	[redacted]	1.05	1.13	£27,516	£6,351
11) Including costs of all subsequent treatments	[redacted]	£78,863	£102,433	[redacted]	1.05	1.13	A+B dominant	A+B dominant
12) Excluding costs of all subsequent treatments	[redacted]	£38,335	£62,580	[redacted]	1.05	1.13	£32,054	A+B dominant
13) Exponential tail for modelling TTD starts at 13 months	[redacted]	£44,573	£62,580	[redacted]	1.05	1.13	£22,424	A+B dominant
14) Exponential tail for modelling TTD starts at 15 months	[redacted]	£45,494	£62,580	[redacted]	1.05	1.13	£20,598	A+B dominant
15) Excluding costs of AE for lenvatinib	[redacted]	£44,983	£62,285	[redacted]	1.05	1.13	£22,267	A+B dominant
16) Including oral chemotherapy administration costs at each prescription	[redacted]	£45,952	£64,762	[redacted]	1.05	1.13	£20,841	A+B dominant

A+B, atezolizumab plus bevacizumab; HCC, hepatocellular carcinoma; ICER, incremental cost-effectiveness ratio; OS, overall survival; QALY, quality adjusted life year; RDI, relative dose intensity; TKIs, tyrosine kinase inhibitors; TTD, time to treatment discontinuation

Table 28: The ERG's base case ICERs varying dose intensity of oral chemotherapy, costs associated with body weight, and the population (all patients or non-Asian patients plus Japanese patient)

Analysis	Discounted costs			Discounted QALYS			ICER (A + B versus sorafenib)	ICER (A + B versus lenvatinib)
	A+B	Sorafenib	Lenvatinib	A+B	Sorafenib	Lenvatinib		
ERG base case A (scenarios 1 – 6) assuming costs for patients weighing less than 60kg	[redacted]			[redacted]				
• Deterministic		£50,783	£60,897		1.32	1.21	£16,296	£1,031
• Probabilistic		£50,946	£61,698		1.33	1.26	£16,567 [95% CI: £16,275 to £16,875]	£83 [95% CI: A+B dominant to £1,526]
ERG base case B (scenarios 1 – 5 plus 7) assuming costs for patients weighing less than 60kg	[redacted]			[redacted]				
• Deterministic		£47,305	£56,883		1.32	1.21	£21,372	£6,043
• Probabilistic		£47,620	£58,966		1.34	1.29	£21,843 [95% CI: £21,387 to £22,331]	£3,962 [95% CI: £2,650 to £5,225]
ERG base case A assuming costs for patients weighing less than 60kg excluding Asia (except Japan)	[redacted]			[redacted]				
• Deterministic		£49,936	£57,865		1.25	1.09	£15,036	£629
• Probabilistic		£50,197	£59,301		1.27	1.18	£15,387 [95% CI: £15,049 to £15,748]	A+B dominant [95% CI: A+B dominant to £333]
ERG base case B assuming costs for patients weighing less than 60kg excluding Asia (except Japan)	[redacted]			[redacted]				
• Deterministic		£46,484	£54,092		1.25	1.09	£21,096	£5,797
• Probabilistic		£46,716	£56,418		1.26	1.18	£21,488 [95% CI: £20,902 to £22,128]	£3,381 [95% CI: £1,956 to £4,752]
ERG base case A assuming costs for patients weighing more than or equal to 60kg	[redacted]			[redacted]				
• Deterministic		£51,252	£77,623		1.32	1.21	£20,967	A+B dominant
• Probabilistic		£51,421	£81,062		1.33	1.30	£21,427	A+B dominant

							[95% CI: £21,009 to £21,871]	[95% CI: A+B dominant to A+B dominant]
ERG base case B assuming costs for patients weighing more than or equal to 60kg • Deterministic • Probabilistic	[redacted]	£47,717 £47,919	£71,602 £73,641	[redacted]	1.32 1.33	1.21 1.27	£26,071 £26,653 [95% CI: £26,056 to £27,289]	A+B dominant A+B dominant [95% CI: A+B dominant to A+B dominant]
ERG base case A assuming costs for patients weighing more than or equal to 60kg excluding Asia (except Japan) • Deterministic • Probabilistic	[redacted]	£50,405 £50,609	£73,586 £76,195	[redacted]	1.25 1.26	1.09 1.18	£20,432 £20,837 [95% CI: £20,286 to £21,438]	A+B dominant A+B dominant [95% CI: A+B dominant to A+B dominant]
ERG base case B assuming costs for patients weighing more than or equal to 60kg excluding Asia (except Japan) • Deterministic • Probabilistic	[redacted]	£46,897 £47,330	£67,927 £70,954	[redacted]	1.25 1.28	1.09 1.17	£26,525 £27,017 [95% CI: £26,177 to £27,940]	A+B dominant A+B dominant [95% CI: A+B dominant to A+B dominant]

A+B, atezolizumab plus bevacizumab; CI, confidence interval; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Table 29: Assessing the impact of ERG's exploratory model results to ERG base case B for patients weighing more than 60kg and excluding Asian patients (except Japanese: deterministic results)

Analysis	Discounted costs			Discounted QALYS			ICER (A + B versus sorafenib)	ICER (A + B versus lenvatinib)
	A+B	Sorafenib	Lenvatinib	A+B	Sorafenib	Lenvatinib		
ERG base case B assuming costs for patients more than or equal to 60kg excluding Asia (except Japan)	[redacted]	**£46,897**	**£67,927**	[redacted]	**1.25**	**1.09**	**£26,525**	**A+B dominant**
8) Use of generalised gamma functions to model OS	The ERG does not believe these results are clinically plausible. See accompanying text							
9) Use of log-logistic functions to model OS	[redacted]	£46,777	£66,769	[redacted]	1.23	0.97	£36,218	A+B dominant
10) Using MRU costs associated with the STA of lenvatinib (TA551)[28]	[redacted]	£57,877	£74,219	[redacted]	1.25	1.09	£32,954	£2,825
11) Including costs of all subsequent treatments	[redacted]	£78,109	£98,665	[redacted]	1.25	1.09	A+B dominant	A+B dominant
12) Excluding costs of all subsequent treatments	[redacted]	£37,786	£58,954	[redacted]	1.25	1.09	£37,449	A+B dominant
13) Exponential tail for modelling TTD starts at 13 months	[redacted]	£46,622	£67,927	[redacted]	1.25	1.09	£25,537	A+B dominant
14) Exponential tail for modelling TTD starts at 15 months	Not estimable as the KM survivor function stops at14 months for this subgroup of patients							
15) Excluding costs of AE for lenvatinib	[redacted]	£46,897	£67,631	[redacted]	1.25	1.09	£26,525	A+B dominant
16) Including oral chemotherapy costs at each prescription.	[redacted]	£47,857	£69,980	[redacted]	1.25	1.09	£24,802	A+B dominant

Table 30: Assessing the impact of ERG's exploratory model results to ERG base case A for patients weighing less than 60kg

Analysis	Discounted costs			Discounted QALYS			ICER (A + B versus sorafenib)	ICER (A + B versus lenvatinib)
	A+B	Sorafenib	Lenvatinib	A+B	Sorafenib	Lenvatinib		
ERG base case A assuming costs for patients below 60kg	[redacted]	£50,783	£60,897	[redacted]	1.32	1.21	£16,296	£1,031
8) Use of generalised gamma functions to model OS	[redacted]	£48,809	£60,738	[redacted]	1.12	1.19	£14,925	£811
9) Use of log-logistic functions to model OS	[redacted]	£50,083	£59,559	[redacted]	1.24	1.06	£18,558	A+B dominant
10) Using MRU costs associated with the STA of lenvatinib (TA551)[28]	[redacted]	£63,177	£68,111	[redacted]	1.32	1.21	£22,210	£12,667
11) Including costs of all subsequent treatments	[redacted]	£81,289	£90,881	[redacted]	1.32	1.21	A+B dominant	A+B dominant
12) Excluding costs of all subsequent treatments	[redacted]	£41,973	£52,239	[redacted]	1.32	1.21	£25,226	£8,455
13) Exponential tail for modelling TTD starts at 13 months	[redacted]	£50,374	£60,897	[redacted]	1.32	1.21	£16,500	£685
14) Exponential tail for modelling TTD starts at 15 months	[redacted]	£51,293	£60,897	[redacted]	1.32	1.21	£14,602	£175
15) Excluding costs of AE for lenvatinib	[redacted]	£50,783	£60,603	[redacted]	1.32	1.21	£16,296	£1,406
16) Including oral chemotherapy costs at each prescription.	[redacted]	£51,752	£63,098	[redacted]	1.32	1.21	£14,851	A+B dominant

As seen in Table 29 the ERG did not believe that results produced when using generalised gamma distributions for estimating OS were clinically plausible. This was because the estimated OS for Non-Asian patients plus Japanese patients crossed for A+B and sorafenib as shown in Figure 16. Figure 17 and Figure 18 presents the OS estimates using the lognormal (ERG's base case) and the log-logistic (explored in ERG's scenario analyses) distributions for the aforementioned population respectively.

Figure 16: Estimates of OS associated with generalised gamma distributions for Non-Asian patients plus Japanese patients

Figure 17: Estimates of OS associated with lognormal distributions for Non-Asian patients plus Japanese patients

Figure 18: Estimates of OS associated with log-logistic distributions for Non-Asian patients plus Japanese patients

6 END OF LIFE

In Table 29 of the CS the company puts forward the case that A+B meets the NICE End of Life criteria. These criteria are:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months and;
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared to current NHS treatment.

The company's base case model estimates mean life years to be 1.50 years for patients receiving sorafenib and 1.54 years for patients receiving lenvatinib. Both values appear to meet the short life expectancy criteria.

A+B is estimated to provide [redacted] life years resulting in estimated extensions of life of [redacted] years compared with sorafenib and [redacted] compared with lenvatinib. Both values are in excess of the three-month period specified in the end of life criterion.

The ERG's base cases did not materially affect these conclusions.

7 OVERALL CONCLUSIONS

The clinical evidence for A+B was based on one sorafenib controlled RCT, IMbrave150, which was open-label but of otherwise good methodological quality, and whose population was considered broadly generalisable to a UK population. There was a statistically significant advantage for A+B over sorafenib for OS, PFS and OR. OS was statistically significantly higher for A+B, than for sorafenib HR (stratified) 0.58 (95% CI 0.42, 0.79) p=0.0006. Median OS for A+B was not estimable (NE), median OS for sorafenib was 13.2 months (95% confidence interval [CI] 10.4, NE). There was a statistically significant treatment group difference for PFS HR (stratified) 0.59 (95% CI 0.47, 0.76) p<0.0001. Median PFS was 6.8 months (95% CI 5.6, 8.3) in the A+B group, and 4.3 months (95% CI 4.0, 5.6) for the sorafenib group.

The most common NCI-CTCAE Grade 3 or 4 AEs experienced in the A+B group were hypertension (10.3%), aspartate aminotransferase increased (4.3%) and proteinuria (2.7%).

The company's economic model indicated that the probabilistic ICER for A+B compared with sorafenib was £22,419 per QALY gained, whilst A+B was assumed to dominate lenvatinib (provided more QALYs at a lower cost). The ERG included the following five exploratory analyses in its base case: 1) correcting perceived errors; 2) using log-normal distributions for estimating OS; incorporating actual dosages rather than planned dosages; 3) including seven days of wastage when discontinuing oral chemotherapy; 4) capping utility at age- and sex-matched values; 5) costing the use of subsequent TKIs and nivolumab, and combined these with two different assumptions related to actual dosage used. An ERG-preferred range was provided as the costs associated with reduced RDI for patients receiving lenvatinib and sorafenib are uncertain, and have different implications on whether the reduced RDI was planned or unintentionally. Four subgroups encompassing combinations of patient weight, less than 60kg or not, and whether the full IMbrave150 population was considered or only non-Asian and Japanese patients were considered.

The ERG-preferred ranges are summarised in Table 26. It is seen that the cost per QALY gained for A+B never exceeded £30,000 per QALY when compared with either sorafenib or lenvatinib. Alternative assumptions, as detailed in Table 29 and Table 30 could push the ICER higher.

These values, however, do not include PAS discounts related to sorafenib, lenvatinib or regorafenib; results including these PAS discounts contained in a confidential appendix to this report.

8 REFERENCES

1. Cheng AL, Qin S, Ikeda M, Galle PR, Ducreux M, Zhu AX, *et al.* Atezolizumab + bevacizumab vs sorafenib in patients with unresectable hepatocellular carcinoma: Phase 3 results from IMbrave150. ESMO ASIA, abstract no. 39.
2. Galle PR, Finn RS, Qin S, Ikeda M, Zhu AX, Kim TY, *et al.* Patient-reported Outcomes From the Phase 3 IMbrave150 Trial of Atezolizumab + Bevacizumab Versus Sorafenib as First-line Treatment for Patients With Unresectable Hepatocellular Carcinoma. ASCO GI, abstract no. 40.
3. Hoffmann La-Roche Ltd. IMbrave150 Clinical Study Report (29 Aug 19 CCOD) no. 1092943; 2019.
4. Centre for Reviews and Dissemination (CRD). CRD's Guidance for Undertaking Reviews in Health Care. 2009.
5. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
6. Hoffmann La-Roche Ltd. IMbrave150 Study Protocol. 2019.
7. Clinical Trials Gov. NCT03434379: A Study of Atezolizumab in Combination With Bevacizumab Compared With Sorafenib in Patients With Untreated Locally Advanced or Metastatic Hepatocellular Carcinoma [IMbrave150]. 2020. https://clinicaltrials.gov/ct2/show/NCT03434379?id=NCT02849080+OR+NCT02607865+OR+NCT03015220+OR+NCT02863328+OR+NCT02863419+OR+NCT03434379&draw=2&rank=1&load=cart (Accessed February 2020).
8. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, *et al.* Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. *Hepatology* 2018;68:723-50.
9. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228-47.
10. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin Liver Dis* 2010;30:52-60.
11. National Cancer Institute. National Cancer Institute Common Terminology Criteria for Adverse Events. (CTCAE) v4.0. 201.
12. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, *et al.* The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;85:365-76.
13. Blazeby JM, Currie E, Zee BCY, Chie W-C, Poon RT, Garden OJ. Development of a questionnaire module to supplement the EORTC QLQ-C30 to assess quality of life in patients with hepatocellular carcinoma, the EORTC QLQ-HCC18. *European Journal of Cancer* 2004;40:2439-44.
14. Fitzsimmons D, Johnson CD, George S, Payne S, Sandberg AA, Bassi C, *et al.* Development of a disease specific quality of life (QoL) questionnaire module to supplement the EORTC core cancer QoL questionnaire, the QLQ-C30 in patients with pancreatic cancer. *European Journal of Cancer* 1999;35:939-41.
15. Chie WC, Blazeby JM, Hsiao CF, Chiu HC, Poon RT, Mikoshiba N, *et al.* International cross-cultural field validation of an European Organization for Research and Treatment of Cancer questionnaire module for patients with primary liver cancer, the European Organization for Research and Treatment of Cancer quality-of-life questionnaire HCC18. *Hepatology* 2012;55:1122-9.
16. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology* 1998;16:139-44.
17. Clinical Trials Gov. NCT02715531: A Study of the Safety and Efficacy of Atezolizumab Administered in Combination With Bevacizumab and/or Other Treatments in Participants With Solid Tumors. 2020. https://www.clinicaltrials.gov/ct2/show/NCT02715531 (Accessed February 2020).

18. Kudo M, Finn RS, Qin S, Han KH, Ikeda K, Piscaglia F, *et al.* Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. *Lancet* 2018;391:1163-73.
19. Cheng AL, Qin S, Ikeda M, Galle P, Ducreux M, Zhu A, *et al.* IMbrave150: Efficacy and safety results from a ph III study evaluating atezolizumab (atezo) 1 bevacizumab (bev) vs sorafenib (Sor) as first treatment (tx) for patients (pts) with unresectable hepatocellular carcinoma (HCC). *Annals of Oncology* 2019;30 (Supplement 9):ix186-ix7.
20. Yamashita T, Kudo M, Ikeda K, Izumi N, Tateishi R, Ikeda M, *et al.* REFLECT-a phase 3 trial comparing efficacy and safety of lenvatinib to sorafenib for the treatment of unresectable hepatocellular carcinoma: an analysis of Japanese subset. *Journal of Gastroenterology* 2020;55:113-22.
21. Yau T, et al. CheckMate 459: A Randomized, Multi-Center Phase 3 Study of Nivolumab vs Sorafenib as First Line Treatment in Patients With Advanced Hepatocellular Carcinoma. *Annals of Oncology* 2019;30 (suppl_5):v851-v934.
22. National Institute for Health and Care Excellence. Guide to the processes of technology appraisal 2018; 2018.
23. Hoffmann La-Roche Ltd. Atezolizumab with Bevacizumab for Untreated, Unresectable or Advanced Hepatocellular Carcinoma [ID 1655]. Company evidence submission.; 2020.
24. British National Formulary. BNF Update. 2017. https://www.medicinescomplete.com/mc/bnf/current/ (Accessed August 2018).
25. Department of Health. NHS Reference Costs 2018-19- Available from: https://improvement.nhs.uk/resources/national-cost-collection. 2019.
26. Georghiou T, Bardsley M. Exploring the cost of care at the end of life. *Nuffield Trust* 2014.
27. Personal Social Services Research Unit. Unit Costs of Health and Social Care 2018. 2018. https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2018/ (Accessed 09 May 2019).
28. National Institute for Health and Care Excellence. Lenvatinib for untreated advanced hepatocellular carcinoma [TA551]. 2018.
29. Chong C, Gulamhussein A, Heathcote EJ, Lilly L, Sherman M, Naglie G, *et al.* Health-state utilities and quality of life in hepatitis C patients. *American Journal of Gastroenterology; 98(3):630-8* 2003.
30. Wright M, Grieve R, Roberts J, Main J, Thomas HC. Health benefits of antiviral therapy for mild chronic hepatitis C: randomised controlled trial and economic evaluation. *Health Technol Assess* 2006;10:1-113, iii.
31. Ratcliffe J, Longworth L, Young T, Bryan S, Burroughs A, Buxton M, *et al.* Assessing healthrelated quality of life pre and post liver transplantation: a prospective multi-centre study. *Liver Transpl* 2002;8:263-70.
32. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, *et al.* Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;15.:708-15.
33. National Institute for Health and Care Excellence. Position statement on use of the EQ-5D-5L valuation set for England (updated October 2019). 2019. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l. (Accessed 25th March 2020).
34. Ara R, Brazier J. Populating an Economic Model with Health State Utility Values: Moving toward Better Practice. *Value in Health* 2010;13.
35. NHS Improvement. Reference costs. 2018. https://improvement.nhs.uk/resources/reference-costs/ (Accessed 29 July 2019).
36. Curtis L, Burns A. Unit Costs of Health and Social Care 2019. 2019.
37. National Institute for Health and Care Excellence. Sorafenib for treating advanced hepatocellular carcinoma [TA474]; 2017.
38. Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, *et al.* Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med* 2008;359:378-90.

39. Cheng AL, Kang YK, Chen Z, Tsao CJ, Qin S, Kim JS*, et al.* Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *Lancet Oncol* 2009;10:25-34.
40. Ouwens MJ, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. *Res Synth Methods* 2010;1:258-71.
41. Jansen JP. Network meta-analysis of survival data with fractional polynomials. *BMC Medical Research Methodology* 2011;11:61.
42. Hernández Alava M, Wailoo AJ, Ara R. Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. *Value Health* 2012;15:550-61.
43. Pennington B, Davis S. Mapping from the Health Assessment Questionnaire to the EQ-5D: The Impact of Different Algorithms on Cost-Effectiveness Results. *Value in Health* 2014;17:762-71.
44. Wailoo AJ, Hernandez-Alava M, Manca A, Mejia A, Ray J, Crawford B*, et al.* Mapping to Estimate Health-State Utility from Non–Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value in Health* 2017;20:18-27.
45. Hatswell AJ, Bullement A, Briggs A, Paulden M, Stevenson MD. Probabilistic Sensitivity Analysis in Cost-Effectiveness Models: Determining Model Convergence in Cohort Models. *Pharmacoeconomics* 2018;36:1421-6.

9 APPENDICES

Appendix 1: Subsequent treatments used in the IMbrave150 study

Table 31: Post-discontinuation subsequent therapies used in the IMbrave150 trial (source: the company's economic model)

Category	Therapy	A+B arm (n=69 equivalent to [redacted] of patients)		Sorafenib arm (n=73 equivalent to [redacted] of patients)	
		No. of patients	Mean duration (months)	No. of patients	Mean duration (months)
Tyrosine Kinase inhibitors	Sorafenib	29	3.4	2	1.26
	Lenvatinib	20	4.08	17	3.85
	Regorafenib	7	2.32	20	3.54
	Sorafenib tosilate	9	2.38	0	0
	Lenvatinib mesilate	3	2.07	5	5.27
	Cabozantinib	2	2.48	6	4.33
	Cabozantinib s-malate	1	6.41	0	0
	Apatinib mesylate	0	0	1	0.13
Immunotherapy	Nivolumab	3	5.62	16	3.92
	Pembrolizumab	1	Not reported	5	4.98
	Atezolizumab	0	0	2	1.76
	Durvalumab	0	0	4	5.68
	IRX-2 (cytokines)	0	0	2	2.17
	Tremelimumab	0	0	4	7.37
	Sintilimab	0	0	2	2.2
	Tislelizumab	0	0	2	6.26
	Triprizumab	0	0	1	0
	Investigational drug	1	7.46	0	0
Angiogenesis inhibitors	Bevacizumab	0	0	2	2.56
	Ramucirumab	2	3.74	3	1.64
Chemotherapy	Fluorouracil	2	7.35	2	5.15
	Oxaliplatin	3	5.05	3	4.21
	Calcium folinate	1	7.46	2	5.15
	Capecitabine	1	0.46	2	1.18
	Pegylated arginine deiminase	1	7.46	1	0.59
	Bufalin/ Cinobufagin/ Resibufogenin	0	0	1	0.13
	Carboplatin	1	5.36	0	0
	Cyclophosphamide	0	0	3	1.72
	Etoposide	0	0	1	2.47
	Folinic acid	1	7.23	0	0
	Gemcitabine	0	0	1	2.33
	Gemcitabine hydrochloride	1	5.36	0	0

	Tegafur/ Uracil	0	0	1	1.87
Others	Thalidomide	0	0	2	1.63
	Generic component(s) not known	0	0	2	1.98
	Antineoplastic agent	0	0	1	0.72
	BLU-554 (FGFR4 inhibitor)	1	3.75	0	0
	Chinese traditional medicine	1	6.84	0	0
	PI3K inhibitor	0	0	1	1.71