in collaboration with:

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

Maastricht University

Glecaprevir-pibrentasvir for treating chronic hepatitis C

ERRATUM

This document contains errata in respect of the ERG report in response to the company's factual accuracy check.

The table below lists the page to be replaced in the original document and the nature of the change:

Page nr:	**Change:**
14	AiC marking has been added
15	We corrected the statement that four instead of three of the 24 subgroups included more than 100 patients.
16	"For some of the subgroups where G/P was not considered cost effective, the reason was that at least one of the comparators, which was considered cost effective, produced the same amount of QALYs at a lower cost. Thus, although G/P was dominated, it can be considered as equally effective as these comparators. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was as effective as at least one cost effective comparator." was replaced by "For seven of the 13 subgroups where G/P was not considered cost effective, as it was dominated, G/P could be considered as approximately equivalent (same QALYs at only slightly higher costs, i.e. max £200) to the most cost-effective comparator, SOL/VEL. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was nearly equivalent to SOL/VEL, the most cost effective comparator." AND: "The IFN eligibility was only considered for GT2, however it was not clear why there was no IFN containing regimen as a comparator for the GT2 TN CC (IFN-eligible) subgroup." was replaced by: "The IFN eligibility was only considered for GT2 TN NC, however, as there is no IFN containing regimen as a comparator for the GT2 TN CC subgroup."
17	"(e.g. whether a patient is intolerant or an inadequate responder to the previous therapy, or has already received a DAA treatment or maybe is DAA naïve, may all impact the effectiveness of G/P)." was replaced by "(e.g. whether a patient is intolerant or an inadequate responder to the previous therapy may impact the effectiveness of G/P)." We added: "However, scenario analysis by the company showed that the addition of these reinfection probabilities has only minimal impact on the results." and "since it was not clear that any alternative base-case assumptions would be properly justified, "

	was replaced by “since it was not clear that any alternative base-case assumptions regarding point estimates and structure would be properly justified, “
18	We corrected the statement that four instead of three of the 24 subgroups included more than 100 patients.
19	“since it was not clear that any alternative base-case assumptions would be properly justified,” was replaced by “since it was not clear that any alternative base-case assumptions regarding point estimates and structure would be properly justified,”
48	The number and % of ENDURANCE-3 patients has been corrected
53	AiC marking has been added We corrected the statement that four instead of three of the 24 subgroups included more than 100 patients.
54	The reference to (ENDURANCE-1 - GT1/NC/TN+TE) is corrected to (EXPEDITION-2).
55	Insomnia rates have been added and AiC marking has been added
58	AiC marking has been added
76	We corrected the statement that four instead of three of the 24 subgroups included more than 100 patients. AiC marking has been added
88	Text added: “In response to the clarification letter, the company performed a scenario analysis showing for one subgroup that the addition of these reinfection probabilities had only minimal impact on the results.”
92	We corrected the statement that four instead of three of the 24 subgroups included more than 100 patients.
108	Text added: “, showing only a small impact on the results.
112	“Given the high level of uncertainty associated with the input parameters of the model, the ERG chose to describe the cost effectiveness results in this section based on the £20,000 threshold.” was replaced by “Given the high level of uncertainty associated with some of the efficacy input parameters of the model (due the small sample sizes on which they are based), the ERG chose to describe the cost effectiveness results in this section based on the £20,000 threshold.” and “For some of the subgroups where G/P was not considered cost effective, the reason was that at least one of the comparators, which was considered cost effective, produced the same amount of QALYs at a lower cost. Thus, although G/P was dominated, it can be considered as equally effective as these comparators. This is indicated with shaded cells in Table 5.17. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups

	where G/P was not cost effective, G/P was as effective as at least one cost effective comparator." was replaced by "For seven of the 13 subgroups where G/P was not considered cost effective as it was dominated, G/P could be considered as approximately equivalent (same QALYs at only slightly higher costs, i.e. max £200) to the most cost-effective comparator, SOL/VEL. This is indicated with shaded cells in Table 5.17. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was nearly equivalent to SOL/VEL, the most cost effective comparator. "
116	"in a PSA" added
126	'due to' replaced by 'despite' Text added: "The impact of including uncertainty appropriately for 100% SVR rates and 0% AE rates was already addressed in section 5.2.11."
127	Text added: "This scenario was performed by the company in response to the clarification letter for one subgroup, and was repeated by the ERG for all subgroups."
128	Text removed: 'relevant'
129	Text added: "However, a scenario analysis by the company showed that the addition of these reinfection probabilities has only minimal impact on the results." "For some of the subgroups where G/P was not considered cost effective, the reason was that at least one of the comparators, which was considered cost effective, produced the same amount of QALYs at a lower cost. Thus, although G/P was dominated, it can be considered as equally effective as these comparators. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was as effective as at least one cost effective comparator." replaced by "For seven of the 13 subgroups where G/P was not considered cost effective, as it was dominated, G/P could be considered as approximately equivalent (same QALYs at only slightly higher costs, i.e. max £200) to the most cost-effective comparator, SOL/VEL. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was nearly equivalent to SOL/VEL, the most cost effective comparator. "
132	AiC marking has been added AND: "For some of the subgroups where G/P was not considered cost effective, the reason was that at least one of the comparators, which was considered cost effective, produced the same amount of QALYs at a lower cost. Thus, although

	G/P was dominated, it can be considered as equally effective as these comparators. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was as effective as at least one cost effective comparator." replaced by "For seven of the 13 subgroups where G/P was not considered cost effective, as it was dominated, G/P could be considered as approximately equivalent (same QALYs at only slightly higher costs, i.e. max £200) to the most cost-effective comparator, SOL/VEL. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was nearly equivalent to SOL/VEL, the most cost effective comparator."
133	We corrected the statement that four instead of three of the 24 subgroups included more than 100 patients. AND: text added: regarding point estimates and structure

1. SUMMARY

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

The company's submission (CS) presents an evaluation of the clinical effectiveness and cost effectiveness of glecaprevir-pibrentasvir (G/P) for the treatment of chronic hepatitis C (CHC). The decision problem addressed by the CS was not completely in line with the final scope issued by the National Institute for Health and Care Excellence (NICE) with respect to the comparators. In particular, daclatasvir (DCV) in combination with sofosbuvir (SOF) (for GT1 and GT4); pegylated-interferon alfa (IFN) with RBV and SOF in combination with RBV (for GT1 and GT4) were not included in the decision problem. The rationale for these omissions, as supplied by the company, states that these treatment regimens are not used in current NHS practice.

The company's model does not include the development of resistance to G/P and other comparators based on the assumption that this outcome does not impact the cost effectiveness of G/P. Also, separate subgroup analyses for patients who are co-infected with HIV, previous treatment received (with or without DAA-containing regimens), people who have received treatment before liver transplantation, and those who have received it after liver transplantation, response to previous treatment (non-response, partial response, relapsed), and people with and without renal impairment were not presented, as it was deemed infeasible by the company.

1.2 Summary of clinical effectiveness evidence submitted by the company

Eighty-one publications (reporting on 79 studies) were identified as meeting the eligibility criteria. Fourteen of these publications, representing seven studies, involved G/P. In addition, information on four further clinical studies of G/P in patients with CHC are mentioned in the company submission. These studies were conducted in special populations of patients with HIV co-infection, renal impairment, failure on prior DAAs and a post-transplant population. Finally, the company mentioned two trials in Japanese patients with CHC: CERTAIN-1 and CERTAIN-2. These trials are only minimally discussed in the CS and not included in the economic model. According to the company, this exclusion was because "these two trials were conducted entirely in Japanese patients" which "precludes their generalisability to the UK patient population and subsequently their use in the economic model". Apart from these two trials in Japanese patients, none of the included studies presented comparative data for the licensed treatment duration of G/P with any of the comparators.

The G/P studies included patients with all genotypes; treatment-naïve and treatment-experienced patient populations; and patients with 'no cirrhosis or compensated cirrhosis'.

When split by cirrhosis status and previous treatment (naïve or experienced), SVR rates were consistently above 90% for all genotypes, except for GT2/TE/NC ([redacted] in SURVEYOR-II, Part 4; but [redacted] in SURVEYOR-II, Parts 1 and 2), GT3/TE/CC ([redacted] in SURVEYOR-II, Part 2; but [redacted] in SURVEYOR-II, Part 3) and GT6/TN/NC ([redacted] in SURVEYOR-II, Part 4).

[redacted]. In studies without a comparator, many treatment arms [redacted]

According to the company, G/P had a favourable safety profile that was similar to placebo and SOF/DCV, and that was similar across treatment durations of 8, 12, and 16 weeks. G/P was well tolerated across a broad and diverse population of patients, including patients with CC, HIV co-

infection, and CKD Stage 4/5. Common study adverse drug reactions (ADRs) occurring in ≥5% of patients were headache, fatigue, and nausea. Adverse drug reactions were mostly Grade 1 (mild) in severity. Serious ADRs and ADRs leading to premature study drug discontinuation were rare (≤0.1%).

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

The company submission (CS) and response to clarification provided sufficient detail for the ERG to appraise the searches for eligible trials. Searches were carried out in accordance with the NICE guide to the methods of technology appraisal sections 5.2.2 and 5.2.4 using a good range of databases. Additional searches of conference proceedings were conducted but no separate literature searches were undertaken to identify adverse events data, non-randomised and non-controlled evidence.

The conclusion from the G/P studies is that G/P has high SVR rates in all genotypes. In addition, G/P has a relative favourable safety and tolerability profile. However, patient numbers for most GT4, GT5 and GT6 populations are very low, often less than 10 patients in each group. Only four out of the 24 subgroups included more than 100 patients (GT1/TN/NC, GT1/TE/NC, GT2/TN/NC and GT3/TN/NC). Therefore, the uncertainty around SVR rates in most subgroups is considerable.

The main concern regarding clinical effectiveness is that data (for SVR12 and AEs) for G/P and comparators were taken from single arms. Therefore, the comparisons for SVR12 rates and AEs between G/P and comparators relies on a naïve indirect comparison. The company does not present any information about the comparability of populations between G/P studies and comparator studies; and about how response and adverse events for comparator studies were selected and whether all possible sources were used. In most cases the sources for SVR rates and AEs for comparators are the same as in TA430 (Sofosbuvir-velpatasvir for treating chronic hepatitis C). Therefore, the same critique as for TA430 applies: these methods increase the chance of bias and cherry-picking.

1.4 Summary of cost effectiveness submitted evidence by the company

The company developed a de novo cost effectiveness model to assess the cost effectiveness of G/P compared to nine different comparators: BSC-watchful waiting, DCV/SOF, DCV/SOF/RBV, EBR/GZR, LDV/SOF, OBV/PTV/DSV+DSV ± RBV, PR, SOF/PR, SOF/RBV and SOF/VEL. The cost effectiveness analyses performed by the company are in line with previous STAs for HCV treatments. The population considered in the cost effectiveness analyses was sub-divided into 26 different subgroups, where patients were stratified by genotypes (GT1, GT2, GT3, GT4, GT5 and GT6), treatment experience (treatment-naïve and treatment-experienced patients), cirrhosis status (cirrhotic and non-cirrhotic patients) and IFN-eligibility (only for GT2 TN patients). Full incremental cost effectiveness results were presented for all subgroups. A National Health Service (NHS) and Personal and Social Services (PSS) perspective was adopted with a lifetime time horizon. A 3.5% discount rate was used for both costs and quality-adjusted life years (QALYs).

The cost effectiveness model developed for this submission was a Markov model which consists of 13 health states. Non-cirrhotic patients start from states F0-F3, and cirrhotic patients start from F4. All treatment related outcomes (achieving SVR, treatment related adverse events and discontinuation) occur within the first year of the model. Patients who do not achieve SVR are at risk of progressing to more severe states. Patients who reach F4 can progress to DC and HCC states, which may lead to liver transplantation and liver-related death. The liver transplantation state was divided into two categories (first year and later years).

Treatment effectiveness was modelled as the probability of achieving SVR. Other treatment-specific parameters included adverse event rates, treatment duration, and treatment-related utility adjustments.

All these parameter estimates were based on naïve indirect comparison of clinical trials assessing the efficacy of G/P and its comparators in the relevant subgroups.

The model uses health state based utilities from the literature (utilities that were used in Wright et al. 2006 and Ratcliffe et al. 2002) in line with previous STAs for HCV treatments. A utility increment due to SVR is applied based on Shepherd et al. 2007 and Hartwell et al. 2011. Treatment-related health utility changes were applied to adjust for the impact on HRQoL of treatment, e.g. due to adverse events.

List prices were used as treatment costs for G/P and the comparator treatments in the cost effectiveness analysis. Health state costs (disease management costs based on disease stage) and other costs for adverse events were based on literature, expert opinion, UK reference costs and previous appraisals for HCV (especially TA430).

The base-case cost effectiveness results showed that for non-cirrhotic patients, G/P was always cost effective except for two subgroups: GT2 treatment-naïve IFN-eligible patients and GT3 treatment-experienced patients. For the subgroups of non-cirrhotic patients for which G/P was cost effective, the relevant comparator was always no treatment, which resulted in very low ICERs (always below £5,000 per QALY). For patients with compensated cirrhosis, G/P was only cost effective for GT1 treatment-naïve and GT3 treatment-naïve patients. For seven of the 13 subgroups where G/P was not considered cost effective, as it was dominated, G/P could be considered as approximately equivalent (same QALYs at only slightly higher costs, i.e. max £200) to the most cost-effective comparator, SOF/VEL. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was nearly equivalent to SOF/VEL, the most cost-effective comparator. In the remaining six subgroups, G/P was clearly not cost effective (ICER above cost effectiveness threshold).

Additionally, the company conducted probabilistic, deterministic and scenario analyses. Probabilistic results were reported as the probability that G/P is cost effective against one single comparator for each subgroup at £20,000 and £30,000 thresholds. The comparator was selected as the one against which G/P had the lowest incremental net monetary benefit when valuing a QALY at £20,000. The result of the deterministic sensitivity analyses showed that in general the ICER was most sensitive to changes in SVR rates. Two scenario analyses conducted by the company first demonstrated how the cost effectiveness of G/P might change after the CMU price agreement (when comparators from other companies were based on list prices). Second, it was shown that using trial-based utilities increased total QALY estimates compared to the base-case when literature-based utilities were used as input.

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

The CS and response to clarification provided sufficient detail for the ERG to appraise the cost effectiveness searches. Searches were well documented but not all searches were reproducible in line with the NICE guide to the methods of technology appraisal. However, a good range of databases were searched and additional searches of conference proceedings were also undertaken.

The following treatments were not included in the cost effectiveness analyses because, according to the company, these are not used in current NHS practice: 1) DCV in combination with SOF, with or without RBV (for specific people with GT1 or GT4; as recommended by NICE); 2) IFN with RBV (for GT1–6; except in GT2 non-cirrhotic treatment-naïve patients); 3) SOF in combination with RBV, with or without IFN (for specific people with GT1 and GT4; as recommended by NICE). The IFN eligibility was only considered for GT2 TN NC, however, as there is no IFN containing regimen as a comparator for the GT2 TN CC subgroup.

Despite being included in the final scope, the company did not perform subgroup analyses for patients who are co-infected with HIV and post-liver transplantation. The subgroup of patients who are intolerant to or ineligible for interferon treatment were only considered for GT2 TN patients. Since these excluded groups (e.g. HIV co-infected patients) were also not taken into consideration while deriving some of the model input estimates (e.g. utility), transferability of the current results for these groups is disputable. Furthermore, heterogeneity of the treatment-experienced population was not taken into account. (e.g. whether a patient is intolerant or an inadequate responder to the previous therapy may impact the effectiveness of G/P).

Onward transmission is not included in the economic model. Incorporating onward transmission would require a dynamic transmission model to capture an ongoing risk of infection for individuals in a population, and therefore could not be incorporated into the current modelling framework. Similarly, the company assumed a zero-reinfection probability after reaching SVR and assumed that no natural recovery takes place, despite contrary evidence reported in the clinical literature. However, scenario analysis by the company showed that the addition of these reinfection probabilities has only minimal impact on the results.

SVR rates, adverse event rates, treatment duration, and treatment-related utility adjustments were based on naïve indirect comparisons of clinical trials assessing the efficacy of G/P and its comparators in the relevant subgroups. The ERG has concerns on the plausibility of this approach, which is not in line with evidence synthesis best practices and is susceptible to bias. Furthermore, some of the SVR rates were either derived from very small sample sizes or the effectiveness in a subgroup was assumed to hold in another subgroup. Since SVR rates are the main driver of costs and effectiveness, all these assumptions create a substantial uncertainty on the cost effectiveness of G/P.

Furthermore, it was not clear to the ERG why age-dependent transition probabilities were not updated every year.

The health state utilities from RCTs could have been used by the company in their cost effectiveness analysis instead of the utilities from the literature. The ERG questions to what extent utility values published in 2006 (originating from EQ-5D questionnaires completed in 2002), i.e. before the DAA-era, can be seen as representative of UK patients currently suffering with CHC. Similarly, the RCT-based utility values show a difference in utility with or without SVR ranging from 0.025 to 0.029, substantially lower than the increment of 0.05 applied by the company based on Wright et al. 2016 thus raising doubt about the validity of the latter value.

The ERG was unsure about the completeness of the health state cost estimates used in the model, as items such as GP visits and home care costs are not included.

The ERG is concerned over the validation status of the cost effectiveness analysis by the company. The tests conducted for the technical verification of the model were not presented and the only validation effort was the external validation of the model estimates of the cirrhosis risk in 20 years from the clinical literature.

Despite the several uncertainties present in the CS base-case, the ERG did not produce an alternative base-case, since it was not clear that any alternative base-case assumptions regarding point estimates and structure would be properly justified, as in most situations the assumptions made by the company were reasonable and in line with previous appraisals. Instead of setting a preferred base-case, the ERG conducted a number of exploratory scenario analyses.

failure to include a large number of SVR and AE rates (i.e. all that have a value of 100% and 0%) in the probabilistic analyses. As a consequence, the ERG considers the PSA results in the CS unreliable. Given the time constraints and the model complexity, the ERG could not produce detailed (corrected) PSA results for all subgroups, only for a few example subgroups. If it is judged that the analysis of uncertainty is a major concern for this submission, the PSA analyses should be repeated after tackling the issues discussed in this report.

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

1.6.1 Strengths

The majority of searches for eligible studies in the CS were well documented. Searches were carried out in a good range of databases and strategies utilised study design filters. In response to clarification questions, a number of searches were repeated to ensure all relevant evidence had been included. Supplementary searches of conference proceedings were also undertaken.

The company's submitted evidence on clinical effectiveness broadly covered the final scope set out by NICE. The review of G/P studies included all relevant studies in which G/P had been used. Reviews for other treatments were likely to have identified the majority of trials of other relevant treatments. The submission covers the key clinical outcomes, including SVR rates, adverse events and mortality.

The structure of the economic model developed by the company is in line with previous models presented in appraisals for HCV submitted to NICE. Thus, the model structure (not necessarily inputs) reflects the main aspects of the chronic HCV disease. The model also includes relevant adverse events, utilities and costs.

1.6.2 Weaknesses and areas of uncertainty

Clinical effectiveness searches were re-run in response to clarification questions but did not include a number of comparators from the original search. Conference searches also did not look for the intervention of interest in addition to some comparator interventions. Cost effectiveness searches that were re-run in response to clarification questions added a restrictive UK country filter, which may have resulted in relevant evidence being missed. There is also concern about the effectiveness of the Embase search for health-related quality of life as the company did not present the full set of records that they claimed to have screened. Some searches were also not reproducible in line with NICE guide to the methods of technology appraisal. There were no searches for adverse events data, non-randomised and non-controlled evidence.

The main concern regarding clinical effectiveness is that data (for SVR12 and AEs) for G/P and comparators were taken from single arms. Therefore, the comparisons for SVR12 rates and AEs between G/P and comparators rely on a naïve indirect comparison. The company does not present any information about the comparability of populations between G/P studies and comparator studies; and about how responses and adverse events for comparator studies were selected and whether all possible sources were used. In most cases, the sources for SVR rates and AEs for comparators are the same as in TA430 (Sofosbuvir-velpatasvir for treating chronic hepatitis C). Therefore, the same critique as for TA430 applies: these methods increase the chance of bias and cherry-picking.

In addition, patient numbers for most GT4, GT5 and GT6 populations in G/P studies are very low, often less than 10 patients in each subgroup. Only four out of the 24 subgroups included more than 100 patients (GT1/TN/NC, GT1/TE/NC, GT2/TN/NC and GT3/TN/NC). Therefore, the uncertainty around SVR rates in most subgroups is considerable.

Since the key parameters in the cost effectiveness analyses (SVR rates) were based on the treatment effectiveness data, all health economic analyses suffer from the uncertainty of clinical effectiveness (i.e. comparative SVR rates). Furthermore, all analyses were conducted on list prices, which may not reflect the actual costs of the treatments to the NHS. Both probabilistic and sensitivity analyses presented by the company were performed incorrectly. As a consequence, the ERG considers the sensitivity analysis results in the CS unreliable. If it is judged that the analysis of uncertainty is a major concern for this submission, these analyses should be repeated after tackling the issues discussed in this report. The company submission would also benefit from a more transparent electronic model.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG has not presented an alternative base-case, since it was not clear that any alternative base-case assumptions regarding point estimates and structure would be properly justified, as in most situations the assumptions made by the company were reasonable and in line with previous appraisals. Instead of setting a preferred base-case, the ERG conducted a number of exploratory scenario analyses. In the scenario analyses assumptions surrounding the utility gain due to SVR, impact of the treatment on utility, impact of age on utility were challenged. In addition, alternative inputs for transition probabilities between fibrosis stages and re-infection rates were explored. Even though these scenarios changed the total costs and/or total QALYs estimates, the impact on incremental results was minimal. The cost effectiveness of G/P in each subgroup did not change, hence the cost effectiveness results of the base-case seem to be robust to changes in utility and treatment-unrelated clinical model inputs.

Additionally, the exploratory PSA analyses conducted by the ERG showed that that the inclusion of parameter uncertainty around all SVR and AE rates (which was not included in the company's base-case when rates were 100% or 0%) can have a major impact on the G/P cost effectiveness probability for certain subgroups; but also that this impact can go in either direction (although more frequently it decreases the G/P cost effectiveness probability). This was especially striking for GT5 TN NC patients, for whom the company might have overestimated the probability of G/P being cost effectiveness by 66 percent. Furthermore, the ERG showed that the PSA outcomes were enormously scattered over the CE plane quadrants for a number of subgroups, which illustrates the main limitation of presenting cost effectiveness probabilities alone (as in the CS).

Table 4.1: Baseline characteristics for relevant G/P studies (ENDURANCE and EXPEDITION) - continued

	ENDURANCE-1[18, 39]	ENDURANCE-3[20, 43]	EXPEDITION-1[47]
Baseline characteristics, n (%)	G/P 8 weeks (N=351)	G/P 8 weeks (N=157)	G/P 12 weeks (N=146)
HCV genotype			
1 (total)	[redacted]	-	87 (59.6)
1a	152 (43.3)	-	[redacted]
1b	[redacted]	-	[redacted]
2 (total)	-	-	34 (23.3)
3 (total)	-	157 (100)	-
4 (total)	-	-	16 (11.0)
5 (total)	-	-	2 (1.4)
6 (total)	-	-	7 (4.8)

Source: CS, Tables 16, 17, 20 and 21, pages 75-89.

CC = compensated cirrhosis; DCV = daclatasvir; G/P = glecaprevir/pibrentasvir (300 mg/120 mg); GT = genotype; SOF = sofosbuvir; HCV = hepatitis C virus; NC = non-cirrhotic; RBV = ribavirin

Genotype	Subgroup	Study	Regimen	SVR12
		SURVEYOR-II, Part 3[24, 52]	G/P 16 weeks:	[redacted]
GT4–6	TN NC	GT4: SURVEYOR-II, Part 4[52]	G/P 8 weeks	[redacted]
		GT5: SURVEYOR-II, Part 4[52]	G/P 8 weeks	[redacted]
		GT6: SURVEYOR-II, Part 4[52]	G/P 8 weeks	[redacted]
	TN CC	GT4: EXPEDITION-1[47]	G/P 12 weeks	[redacted]
		GT5: EXPEDITION-1[47]	G/P 12 weeks	[redacted]
		GT6: EXPEDITION-1[47]	G/P 12 weeks	[redacted]
	TE NC	GT4–6: SURVEYOR-II, Part 4[52]	G/P 8 weeks	[redacted]
	TE CC	GT4–6: EXPEDITION-1[47]	G/P 12 weeks	[redacted]
Source: CS, section B2.7.1, page 108 *ITT population excluding prior SOF+ RBV ± peg-IFN failures				

ERG comment: As can be seen from Table 4.8, numbers for most GT4, GT5 and GT6 populations are very low, often less than 10 patients in each group. Only four out of the 24 subgroups included more than 100 patients (GT1/TN/NC, GT1/TE/NC, GT2/TN/NC and GT3/TN/NC). Therefore, the uncertainty around SVR rates in most subgroups is considerable.

4.2.2 Health-related quality of life

[redacted] In ENDURANCE-1, '[redacted]' (G/P vs Placebo) and in ENDURANCE-3 '[redacted]' (G/P vs SOF/DCV). In studies without a comparator, many treatment arms [redacted], according to the company.

4.2.3 Subgroup analyses

Subgroup analyses are described in section B2.8 (pages 128-129) of the CS and Appendix E (CS Appendix, pages 385-392). Basic results presented above are already reported by genotype, for people with and without cirrhosis and based on previous treatment (naïve or experienced). Additional subgroups mentioned in the scope are:

- co-infection with HIV
- previous treatment received (with or without DAA-containing regimens)
- people who have received treatment before liver transplantation, and those who have received it after liver transplantation

- response to previous treatment (non-response, partial response, relapsed)
- people who are intolerant to or ineligible for interferon treatment
- people with and without renal impairment

From these subgroups, the company provided results for people co-infected with HIV (EXPEDITION-2). No results are provided for any of the other subgroups that were used in the economic model.

4.2.4 Adverse events

The summary of the safety profile for G/P in the SmPC[11] shows that in patients treated for eight, 12 or 16 weeks with compensated liver disease (with or without cirrhosis), based on Phase 2 and 3 studies which evaluated approximately 2,300 patients, the most commonly reported adverse reactions (incidence ≥ 10%) were headache and fatigue. Less than 0.1% of patients treated with G/P had serious adverse reactions (transient ischaemic attack). The proportion of patients treated with G/P who permanently discontinued treatment due to adverse reactions was 0.1%. The type and severity of adverse reactions in patients with cirrhosis were overall comparable to those seen in patients without cirrhosis.[11]

The most commonly reported adverse reactions identified in patients treated with G/P are reported in Table 4.9. The adverse reactions are listed below by body system organ class and frequency.

Table 4.2: Adverse reactions identified with G/P

Frequency	Adverse reactions
Nervous system disorders	
Very common	headache
Gastrointestinal disorders	
Common	diarrhoea, nausea
General disorders and administration site conditions	
Very common	fatigue
Common	asthenia
Source: Glecaprevir & Pibrentasvir (Maviret) Draft SPC_26-06-2017[11] Very common: ≥ 1/10), common: ≥ 1/100 to < 1/10)	

Adverse events (AEs) in the CS are reported in four groups. First, AEs from a placebo-controlled study (ENDURANCE-2); second, AEs from an active-controlled study (ENDURANCE-3); third, AEs from all randomised patients from 21 arms of the Phase II/III studies who received at least one dose of G/P 300 mg/120 mg OD without RBV; and fourth, AEs from a study including patients with chronic kidney disease (CKD Stage 4/5; EXPEDITION-4).

Placebo-controlled study: ENDURANCE-2

In the placebo-controlled analysis set, 302 (202 G/P, 100 placebo) patients received at least one dose of study drug in ENDURANCE-2. Patients were genotype GT2, NC, TN or TE with IFN, peg-IFN ± RBV, or SOF + RBV ± peg-IFN. Treatment was 12 weeks of G/P at a dose of 300 mg/120 mg. Adverse events from ENDURANCE-2 are reported in Table 4.10.

Table 4.3: ENDURANCE-2 and ENDURANCE-3 adverse events summaries

Adverse events, n (%)	ENDURANCE-2		ENDURANCE-3	
	G/P (300 mg/ 120 mg), 12 weeks (N=202)	Placebo 12 weeks (N=100)	G/P (300 mg/ 120 mg) 12 weeks (N=233)	SOF + DCV 12 weeks (N=115)
≥1 AE	131 (64.9)	58 (58.0)	177 (76.0)	80 (69.6)
≥1 treatment-related AE	[redacted]	[redacted]	112 (48.1)	50 (43.5)
Grade 3 or 4 AE	[redacted]	[redacted]	[redacted]	[redacted]
Grade 3/4 AEs				
Alanine aminotransferase increased	[redacted]	[redacted]	NR	NR
Ankle fracture	[redacted]	[redacted]	NR	NR
Aspartate aminotransferase increased[a]	[redacted]	[redacted]	NR	NR
Bile duct stone[c]	[redacted]	[redacted]	NR	NR
Gamma-glutamyltransferase increased[a]	[redacted]	[redacted]	NR	NR
Haemorrhoids	[redacted]	[redacted]	NR	NR
Joint dislocation[b]	[redacted]	[redacted]	NR	NR
Pulmonary pain	[redacted]	[redacted]	NR	NR
Neutropaenia	[redacted]	[redacted]	NR	NR
≥1 treatment-related SAE	NR	NR	NR	NR
Deaths	NR	NR	NR	NR
Discontinuation due to AEs	NR	NR	1	NR
Common AEs[†]				
Headache	24 (11.9)	12 (12.0)	60 (25.8)	23 (20.0)
Fatigue	23 (11.4)	10 (10.0)	44 (18.9)	16 (13.9)
Insomnia	NR	NR	[redacted]	[redacted]
Nausea	[redacted]	[redacted]	32 (13.7)	15 (13.0)
Oropharingeal pain	[redacted]	[redacted]	NR	NR
Nasopharyngitis	NR	NR	[redacted]	[redacted]
Upper respiratory infection	NR	NR	[redacted]	[redacted]
Irritability	NR	NR	NR	NR
Cough	NR	NR	NR	NR
Pruritus	[redacted]	[redacted]	NR	NR
Dyspepsia	NR	NR	NR	NR
Back pain	NR	NR	NR	NR
Asthenia	[redacted]	[redacted]	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]	[redacted]	[redacted]
Dizziness	[redacted]	[redacted]	NR	NR
Constipation	NR	NR	NR	NR

Table 4.4: Overview of AEs (EXPEDITION-4)

	EXPEDITION-4, n (%) (N=104)
Any AE	74 (71.2)
Any DAA-related AE[a,b]	[redacted]
An AE Grade ≥3	[redacted]
Any DAA-related AE Grade ≥3[a,b]	[redacted]
Any SAE	25 (24.0)
Any DAA-related SAE[a,b]	0
Discontinuation of study drug due to:	
Any AE	4 (3.8)
Any DAA-related AE[a,b]	[redacted]
Any fatal AE	[redacted]
All deaths[c]	1 (1.0)

Source: CS Appendix F, Table 206, page 165
[a]DAAs = GLE, PIB, or G/P; [b]Investigator assessment; [c]Includes nontreatment-emergent deaths
AE = adverse event; DAA = direct-acting antiviral agent; G/P = glecaprevir/pibrentasvir; GLE = glecaprevir; PIB = pibrentasvir; SAE = serious adverse event

Among patients in EXPEDITION-4, the most frequently reported (≥10.0% of patients) AEs were pruritus, fatigue, and nausea (see Table 4.13).

Table 4.5: Treatment-emergent adverse events reported in ≥ 5.0% of patients

MedDRA 19.0 Preferred Term	EXPEDITION-4, (N = 104), n (%)
Any adverse event	[redacted]
Pruritus	[redacted]
Fatigue	[redacted]
Nausea	[redacted]
Asthenia	[redacted]
Diarrhoea	[redacted]
Decreased appetite	[redacted]
Headache	[redacted]
Vomiting	[redacted]
Dizziness	[redacted]
Dyspnoea	[redacted]

Source: CSR, Table 25, page 138[59]
EXPEDITION-4: G/P, 300 mg/120 mg QD for 12 weeks
MedDRA = Medical Dictionary for Regulatory Activities; QD = once daily

Of the patients in EXPEDITION-4 experiencing DAA-related events (N=[redacted]), [redacted] ([redacted]%) had events of maximum severity of Grade 1 (mild), [redacted] ([redacted]%) had a maximum severity of Grade 2, and [redacted] ([redacted]%) had a maximum severity of Grade 3.

[redacted]

[redacted]

In CERTAIN-1, the primary efficacy analysis was the percentage of GT1-infected NC patients in the ITT population of sub-study 1 without Y93H polymorphisms who achieved SVR12. This was 99.1% (two-sided 95% CI 97.2% to 100.0%) following eight weeks of treatment with G/P, and 100% [redacted] following 12 weeks of treatment with OBV/PTV/RTV. Further results for this study are not reported in the company submission. The CSR shows that a SVR12 rate of [redacted] was achieved in HCV GT3-infected patients with compensated cirrhosis or without cirrhosis and with or without prior pegylated IFN/ribavirin experience who were treated with 12 weeks of G/P.[64] [redacted].

The fixed-dose combination of G/P 300 mg/120 mg QD administered for eight and 12 weeks was well tolerated by Japanese patients including those without cirrhosis, with compensated cirrhosis, and with severe renal impairment, including those on dialysis. A similar safety profile was observed between HCV GT1-infected, DAA treatment-naïve, Japanese patients treated with either G/P 300 mg/120 mg QD administered for eight weeks or OBV/PTV/RTV QD for 12 weeks. Overall, among patients treated with G/P, the most common (≥ 5.0% of patients) TEAEs were nasopharyngitis, pruritus, and headache.[64]

4.5.2 CERTAIN-2

The CERTAIN-2 trial (NCT02723084) was a Phase III, randomised, open-label, multicentre study to evaluate the efficacy of G/P in Japanese NC adults with chronic GT2 HCV infection.[64, 67-69] The objectives of the study were to determine the safety and efficacy of G/P treatment.

GT2-infected NC DAA-TN patients were randomised at a 2:1 ratio to receive G/P (300 mg/120 mg) for eight weeks or SOF + RBV for 12 weeks. 136 patients were enrolled. The primary efficacy endpoint tested the non-inferiority of the SVR12 rate in the eight-week G/P arm to the 12-week SOF + RBV arm. The secondary efficacy endpoints were in line with CERTAIN-1.

In CERTAIN-2, the SVR rate among GT2-infected DAA-TN patients without cirrhosis 12 weeks after treatment with G/P for eight weeks was 97.8% (two-sided 95% CI 94.7% to 100.0%), and 93.5% [redacted] with SOF + RBV for 12 weeks. Further results for this study are not reported in the company submission.

The fixed dose combination of G/P 300 mg/120 mg QD administered for eight weeks was well tolerated by Japanese patients with HCV GT2 infection without cirrhosis. Patients treated with G/P treatment had fewer overall TEAEs and TEAEs related to treatment compared to SOF + RBV treatment. Patients treated with SOF + RBV had higher rates of anemia, hyperbilirubinemia, and hyperuricemia. Overall among patients treated with G/P, the most common (≥ 5% of patients) TEAEs were nasopharyngitis, headache, and malaise. No TEAE related to treatment was reported in > 5% of patients treated with G/P. The most common (≥ 5% of patients) TEAEs reported among patients receiving SOF + RBV were anemia, blood bilirubin increased, malaise, nasopharyngitis, nausea, stomatitis, and hyperuricemia. TEAEs related to SOF + RBV reported in > 5% of patients included anemia and blood bilirubin increased. The higher rates of these events related to SOF + RBV are likely due to the effect of RBV.[69]

4.6 Conclusions of the clinical effectiveness section

The conclusion from the G/P studies is that G/P has high SVR rates in all genotypes. In addition, G/P has a relative favourable safety and tolerability profile. However, patient numbers for most GT4, GT5 and GT6 populations are very low, often less than 10 patients in each subgroup. Only four out of the 24 subgroups included more than 100 patients (GT1/TN/NC, GT1/TE/NC, GT2/TN/NC and GT3/TN/NC). Therefore, the uncertainty around SVR rates in most subgroups is considerable.

In line with previous approaches accepted by NICE,[176] the company did not include onward transmission and the probability of re-infection in their cost effectiveness model. The ERG agrees with the company that modelling onward transmission would not fit into a common Markov model. However, re-infection probabilities have been excluded from the model without any proper justification. The company claims (on page 145 in the CS) that including onward transmission in the model is likely to result in lower ICERs for active treatments,[2] in particular, for those that are most effective and for which onward transmission would be most reduced. In contrast, re-infection is likely to result in higher ICERs for active treatments since patients who achieved SVR would be in risk of advancing to more severe health states without the possibility of re-achieving SVR (given that subsequent therapies are not included in the model). The company also refers to Madin-Warburton et al. 2016 where it is shown that "*there is a net positive impact on cost effectiveness in a dynamic transmission model for treatment of HCV infection of incorporating both re-infection and onward transmission*".[179] Based on these, the company concluded (on page 145 in the CS) that their model "*may represent a conservative approach that under-estimates the cost effectiveness of active treatments including G/P*".[2] While this conclusion might be correct, the ERG considers that it is not possible to determine the extent to what this approach is indeed conservative or not. In response to the clarification letter, the company performed a scenario analysis showing for one subgroup that the addition of these reinfection probabilities had only minimal impact on the results.

5.2.3 Population

The patient population considered in the company's economic analyses was adults with CHC. Results are presented for 26 different subgroups, which are characterised by HCV genotype, treatment history and fibrosis status. There are six different HCV genotypes (GT1-GT6), each with different characteristics (see also Section 2 of this report). Treatment history distinguishes between treatment-naïve and treatment-experienced patients where the latter are defined as patients who have not adequately responded to prior IFN/RBV-based treatment with or without SOF. This is in line with the clinical trial programme of G/P (see Section B.2 in the CS).[2] Fibrosis status considers non-cirrhotic patients (i.e. patients with METAVIR score F0-F3) and patients with compensated cirrhosis (i.e. patients with METAVIR score F4). Analyses for IFN-ineligible versus IFN-eligible patients are conducted for GT2 treatment-naïve patients only. However, it should be noted that the only differences between the IFN-eligible and IFN-ineligible patients are the comparators considered for the economic analyses, i.e. the SVR or AE rates are not adjusted according to IFN-eligibility. Furthermore, GT1a and GT1b subgroups are not differentiated in the company's model. A summary of the subgroups included in the CS is presented in Table 5.3.

Table 5.6: Population subgroups considered in the company's economic analyses

HCV genotype	Treatment-naïve		Treatment-experienced	
	Non-cirrhotic	Compensated cirrhosis	Non-cirrhotic	Compensated cirrhosis
GT1	✓	✓	✓	✓
GT2	IFN-eligible: ✓ IFN-ineligible: ✓	IFN-eligible: ✓ IFN-ineligible: ✓	✓	✓
GT3	✓	✓	✓	✓
GT4	✓	✓	✓	✓
GT5	✓	✓	✓	✓

The company did not consider any treatment continuation rules for G/P or any relevant comparators. Although NICE guidance recommends SOF + DCV for GT3 NC patients with significant fibrosis only, the company took a pragmatic approach and included this treatment as a comparator for all GT3 NC patients.

5.2.5 Perspective, time horizon and discounting

The cost effectiveness analyses performed by the company adopted the perspective of the NHS/PSS. A discount rate of 3.5% was applied for both costs and utilities. A 70-year time horizon with an annual cycle length was assumed in the cost effectiveness model.

5.2.6 Treatment effectiveness and extrapolation

Treatment effectiveness parameters for the model were derived from the trial data described throughout Section 4 of this report. As explained in Section 5.2.2, two main types of transition probabilities can be distinguished in the model: SVR rates and natural disease progression transition probabilities. These are discussed in more detail below.

Sustained virologic response rates

SVR rates were obtained from clinical trial data. These were used to estimate the transition probabilities from baseline health states (mild fibrosis, moderate fibrosis or CC) to the corresponding "recovered" health state after successful treatment. In particular, the SVR rates (defined as HCV RNA <LLOQ) observed at 12 weeks after the end of treatment on the ITT population (denoted by SVR12) from the company and comparator clinical trials were used directly in the model. These are presented in Table 4.16 of this report. SVR rates are further stratified by fibrosis severity (NC [F0–F3] and CC [F4]) and HCV genotype (GT1 to GT6). Since in most of cases available data did not report different SVR rates for mild (F0-F1) and moderate (F2-F3) fibrosis, the available NC SVR rate was applied for both the mild and moderate fibrosis health states. Only for SOF/LDV in GT1 TN patients, SVR rates were obtained separately for patients with mild and moderate fibrosis.

ERG comment: The model uses the SVR12 rates obtained in RCTs with the various treatment options as model input for treatment effectiveness. As also discussed in Section 4 of this report the main concern is that data for SVR12 were taken from single arms. Therefore, the comparisons for SVR12 rates between G/P and comparators rely on a naïve indirect comparison. The company does not present any information about the comparability of populations between G/P studies and comparator studies; and about how response and adverse events for comparator studies were selected and whether all possible sources were used. The limitations of this input data necessarily lead to non-robust cost effectiveness outcomes.

In addition, patient numbers for most GT4, GT5 and GT6 populations are very low, often less than 10 patients in each subgroup. Only four out of the 24 subgroups included more than 100 patients (GT1/TN/NC, GT1/TE/NC, GT2/TN/NC and GT3/TN/NC). Therefore, the uncertainty around SVR rates in most subgroups is considerable.

Natural disease progression transition probabilities

Natural disease progression transition probabilities were derived from the literature. These were categorised in four different groups: fibrosis progression, non-fibrosis progression, liver transplantation and liver-related mortality. A brief description of each category and a summary of the annual transition probabilities used in the economic model are given below.

ERG comment: Using utilities derived from the literature[153] is consistent with the approach used in previous STAs.[25, 26, 195, 197] However, it also means that in this STA, as well as some of the previous STAs, utilities derived from RCTs have not been taken into account in the base-case. In the CS it is argued that UK patients represented only a small percentage of the total enrolled patient sample in the various G/P RCTs and that it was therefore felt that these utilities would not be representative of the UK patients suffering with CHC. A similar justification was given in the STA of EBR/GZR.[147] However, the ERG questions to what extend utility values published in 2006 (originating from EQ-5D questionnaires completed in 2002),[153] i.e. before the DAA-era, can be seen as representative of UK patients currently suffering with CHC.

As the RCT-based utilities are higher than those observed in Wright et al. 2006,[153] with smaller differences between F0-F1, F2-F3, and F4, and smaller differences between states with and without a SVR, it is relevant to assess the impact of changing the source of the health state utility values. This scenario analysis has been provided in the CS, and the results are presented in Section 5.3. There it can be seen that these RCT utility values lead to a higher number of QALYs per treatment, without really altering the conclusions regarding cost effectiveness.

From the RCT-based utility values as presented in Table 117 from the CS,[2] it can be seen that the difference in utility of a health state with or without SVR ranges from 0.025 to 0.029, substantially lower than the increment of 0.05 applied by the company.[153] This raises the question if the utility gain observed in Wright et al. 2006 can still be considered as a valid estimate.[153] The ERG therefore requested in their clarification letter (question B11) that the company would perform a scenario analysis with the SVR-gain set to 0, as an extreme scenario. [13] Although the company explained how to do such scenario analysis in the electronic model, they did not provide the results of that scenario analysis. Hence, the ERG ran the scenario and its results are presented in Section 5.3, showing only a minimal impact on the results.

The impact of receiving treatment on health-related quality of life was taken into account in the company model using utility increments and decrements. Note that these changes in utility were only applied while patients are on treatment but not through the whole model's time horizon. Conceptually, the ERG agrees with this approach as it takes into account both the impact of a quick response to treatment and the impact of adverse events. However, most of these adjustment estimates were based on the same studies as the estimates of SVR rates and AE rates, implying that all comments regarding those (see Section 4.6) apply here as well. Therefore, the ERG requested in their clarification letter (question B11) that the company would perform a (worst case) scenario analysis in which no utility adjustments would be applied.[13] However, the company opted not to provide the results of such analysis and instead only described which changes had to be made to run the analysis. In Section 5.3 the results of the scenario analysis as run by the ERG are presented, showing only a small impact on the results.

5.2.9 Resources and costs

In the CS the costs for the clinical management of CHC are made up of two main components: 1) Health state costs and 2) treatment-related costs.

Health state costs capture the average medical costs in a specific health state. Costs include those associated with the management of progressive liver disease (in patients who do not respond to treatment) and with post-treatment surveillance following treatment cessation and achievement of SVR.

Treatment-related costs consist of drug acquisition costs multiplied by the mean treatment duration from trials, costs associated with on-treatment monitoring for response, and costs of treating adverse events to treatment.

Observational data regarding resource use for adverse events would be needed to reduce the uncertainty that currently exists. However, from the lack of mentioning of AE costs in the tornado diagrams reporting the DSA (CS Appendix L.1.3) it can be deducted that even when adverse event costs are altered by 50%, they have an almost negligible impact on the results.[16]

5.2.10 Cost effectiveness results

Cost effectiveness results were presented incrementally including all relevant comparators for the different subgroups considered in the analyses. Subgroups were characterised by genotype (GT1 – GT6), treatment history (treatment-naïve or treatment-experienced) and cirrhosis status (non-cirrhotic or compensated cirrhosis). Furthermore, GT2 treatment-naïve patients were also subdivided by IFN-eligibility. This resulted in 26 subgroups in total as reported in Table 5.3 in Section 5.2.3.

Base-case incremental cost effectiveness analysis results

The results summarised in this section are sourced from Appendix B14 in the clarification responses.[17] These were provided by the company after it was discovered during the clarification phase (Question B14 in the clarification letter[17]), that the results reported in the CS did not match those obtained from the submitted economic model. In these analyses, list prices were used for G/P and all comparators.

Table 5.17 below provides an overview of the (list price) base-case cost effectiveness results per subgroup. In the CS, results often refer to both the £20,000 and £30,000 cost per QALY threshold, which might be leading to some confusion, given the vast amounts of results that need to be presented. Given the high level of uncertainty associated with some of the efficacy input parameters of the model (due the small sample sizes on which they are based), the ERG chose to describe the cost effectiveness results in this section based on the £20,000 threshold.

It was observed that, for non-cirrhotic patients, G/P was always cost effective except for the following two subgroups: GT2 treatment-naïve IFN-eligible patients and GT3 treatment-experienced patients. For the subgroups of non-cirrhotic patients for which G/P was cost effective, the relevant comparator was always no treatment, which resulted in very low ICERs (always below £5,000 per QALY). For patients with compensated cirrhosis, G/P was only cost effective for GT1 treatment-naïve and GT3 treatment-naïve patients. For seven of the 13 subgroups where G/P was not considered cost effective as it was dominated, G/P could be considered as approximately equivalent (same QALYs at only slightly higher costs, i.e. max £200) to the most cost-effective comparator, SOF/VEL. This is indicated with shaded cells in Table 5.17. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was nearly equivalent to SOF/VEL, the most cost effective comparator. In the remaining six subgroups, G/P was clearly not cost effective (ICER above cost effectiveness threshold).

For patients with compensated cirrhosis, the results for G/P were the same as in GT4.

5.2.11 Sensitivity analyses

Sensitivity analyses were undertaken in the 26 patient subgroups described in Section 5.2.3 of this report. Due to the large number of subgroups and comparators within each subgroup, the company judged it unfeasible to perform PSA/DSA for all treatment comparisons in all patient subgroups (cf. pp. 217 and 219 in the CS).[2] Thus, for each subgroup a comparison of G/P to a single comparator treatment was chosen. The comparator was selected as the one against which G/P had the lowest incremental net monetary benefit when valuing a QALY at £20,000. The comparators used by the company in the PSA/DSA are summarised per subgroup in Table 5.18.

Table 5.7: Comparators used for PSA/DSA analyses

HCV genotype	Treatment-naïve		Treatment-experienced	
	Non-cirrhotic	Compensated cirrhosis	Non-cirrhotic	Compensated cirrhosis
GT1	SOF/LDV	EBR/GZR	OBV/PTV/RTV + DSV	SOF/VEL
GT2	IFN-eligible: peg-IFN + RBV IFN-ineligible: SOF + RBV	IFN-eligible: SOF/VEL IFN-ineligible: SOF/VEL	SOF/VEL	SOF/VEL
GT3	SOF/VEL	SOF/VEL	SOF + peg-IFN + RBV	SOF/VEL
GT4	OBV/PTV/RTV	OBV/PTV/RTV	OBV/PTV/RTV	OBV/PTV/RTV
GT5	SOF/VEL	SOF/VEL	SOF/VEL	SOF/VEL
GT6	SOF/VEL	SOF/VEL	SOF/VEL	SOF/VEL

Source: Table 113 in the CS.[2]
DSA = deterministic sensitivity analysis; DSV = dasabuvir; EBR = elbasvir; GT = genotype; GZR = grazoprevir; IFN = interferon; LDV = ledipasvir; OBV = ombitasvir; PSA = probabilistic sensitivity analysis; PTV = paritaprevir; peg-IFN = pegylated IFN; RBV = ribavirin; RTV = ritonavir; SOF = sofosbuvir; VEL = velpatasvir

ERG comment: The ERG considers that choosing a single comparator in a PSA is methodologically incorrect and the interpretation of the results can be potentially misleading. In general, when more than two treatments have a positive cost effectiveness probability at a certain cost effectiveness threshold, restricting the analysis to two treatments only is likely to overestimate the cost effectiveness probability of the most cost effective treatment. Therefore, PSA with multiple comparators should have been performed.

Probabilistic sensitivity analysis

The company distinguished between treatment-specific and non-treatment specific input parameters. The first group included SVR rates, AE rates and treatment-related utility change. Treatment-specific input parameters were varied when possible using the 95% confidence intervals observed in the clinical trials. This was the case for SVR and AE rates, which were assumed to follow a Beta distribution, with the input parameters given by the trial subgroup sample size and percentage of patients achieving SVR or with an AE in that subgroup. SVR rates were summarised in Table 4.16 and AE rates in Table 5.9 and 5.10. Due to the lack of data, only for G/P was the treatment-related utility change (see Table 5.12)

of subgroups included in the economic analyses, the adjustments that needs to be made for each of them (e.g. selecting the appropriate comparators) and the lack of time, the ERG considered that the aspects mentioned above could have been corrected in the model to facilitate its validation and to avoid an unnecessary burden on the ERG.

5.3 Exploratory and sensitivity analyses undertaken by the ERG

In this section, the ERG conducted additional scenario analyses on the company base-case to explore the uncertainty around the assumptions taken in the company's base-case analysis. The ERG refrained from setting a preferred base-case, despite the concerns about the uncertainty surrounding SVR rates for the intervention and its comparators, which are caused by small sample sizes for some groups (e.g. n=2) as well as the method used to compare the effectiveness between treatments (naïve indirect comparison). The impact of including uncertainty appropriately for 100% SVR rates and 0% AE rates was already addressed in section 5.2.11. Instead of setting a preferred base-case, the ERG conducted a number of exploratory scenario analyses.

The following exploratory scenarios were conducted:

- No utility gain in SVR
- No treatment effect in utility
- Age based utility decrement
- Alternative transition probability inputs for fibrosis states
- Non-zero re-infection rates

5.3.1 Scenario-1: No utility gain in SVR

In this scenario, it was assumed that after SVR, there is no additional gain in health utility, whereas in the base-case a utility gain of 0.05 was assumed. In this scenario, it was assumed that after SVR, there is no utility gain, whilst in the base-case a utility gain of 0.05 was assumed. The removal of this utility gain has no impact on the ranking of G/P regarding cost effectiveness (yes or no in a subgroup), total costs, and total QALYs; these remain the same as presented in Table 5.17.

5.3.2 Scenario-2: No treatment specific health utility change

In this scenario, it was assumed that there is no treatment-related health utility change whilst on treatment. In the base-case, the values given in Table 5.12 were applied. Removing these utility adjustments had only an impact on the QALY ranking for GT4, GT5 and GT6, for TE NC patients. It had no impact on the ranking of G/P regarding cost effectiveness and total costs.

Table 5.8: G/P cost effectiveness per subgroup, without a treatment-related utility adjustment (based on list price deterministic full incremental results)

HCV genotype	Treatment-naïve		Treatment-experienced	
	Non-cirrhotic	Compensated cirrhosis	Non-cirrhotic	Compensated cirrhosis
GT1	same as Table 5.17	same as Table 5.17	same as Table 5.17	same as Table 5.17
GT2	IFN-ineligible: same as Table 5.17	IFN-eligible: same as Table 5.17	same as Table 5.17	same as Table 5.17
	IFN-ineligible: same as Table 5.17	IFN-ineligible: same as Table 5.17		
GT3	same as Table 5.17	same as Table 5.17	same as Table 5.17	same as Table 5.17

HCV genotype	Treatment-naïve		Treatment-experienced	
	Non-cirrhotic	Compensated cirrhosis	Non-cirrhotic	Compensated cirrhosis
GT4	same as Table 5.17	same as Table 5.17	G/P cost effective 2nd lowest total costs *highest QALYs (together with SOF/VEL, EBR/GZR and OBV/PTV/RTV + DSV ± RBV)*	G/P not cost effective 4th lowest total costs highest QALYs (together with SOF/VEL and *LDV/SOF*)
GT5	same as Table 5.17	same as Table 5.17	G/P cost effective 2nd lowest total costs *highest QALYs (together with SOF/VEL)*	same as Table 5.17
GT6	same as Table 5.17	same as Table 5.17	G/P cost effective 2nd lowest total costs *highest QALYs (together with SOF/VEL)*	same as Table 5.17

Source: Electronic model.[204]

GT = genotype; IFN = interferon; G/P = glecaprevir/pibrentasvir (300 mg/120 mg); QALY = quality-adjusted life year; SOF = sofosbuvir; VEL = velpatasvir; DSV = dasabuvir; EBR = elbasvir; GZR = grazoprevir; LDV = ledipasvir; OBV = ombitasvir; PTV = paritaprevir; RTV = ritonavir; RBV = ribavirin;

5.3.3 Scenario-3: Age-based utility decrement

In this scenario, age based utility decrements derived from Ara and Brazier 2010[210] were applied. In the base-case, no age based utility decrements were applied. The addition of these age based utility decrements has no impact on the ranking of G/P regarding cost effectiveness, total costs, and total QALYs; these remain the same as presented in Table 5.17.

5.3.4 Scenario-4: Alternative transition probabilities for the fibrosis states

In this scenario, alternative transition probabilities from Grischenko et al. 2009 were applied for the transitions between the fibrosis states.[178] In the base-case transition probabilities from Thein et al. 2008 were used.[158] When compared with the base-case results, the addition of these alternative transition probabilities has no impact on the ranking of G/P regarding cost effectiveness, total costs, and total QALYs; these remain the same as presented in Table 5.17.

5.3.5 Scenario-5: Non-zero re-infection rates

In this scenario alternative probabilities for re-infection from SVR states were incorporated. This scenario was performed by the company in response to the clarification letter for one subgroup, and was repeated by the ERG for all subgroups. The re-infection probability estimate of 0.0033 from Simmons et al. 2016[211] was assumed. In the base-case re-infection probability was assumed to be zero.

5.4 Conclusions of the cost effectiveness section

The ERG considered that the economic model described in the CS meets the NICE reference case to a reasonable extent. While the economic model is in line with the decision problem formulated by the company, it is only partially in line with the scope. Intervention and comparators included in the company's economic analysis were also included in the scope. However, other comparators listed in the NICE scope [1) DCV in combination with SOF, with or without RBV (for specific people with GT1 or GT4; as recommended by NICE); 2) IFN with RBV (for GT1–6; except in GT2 non-cirrhotic treatment-naïve patients); 3) SOF in combination with RBV, with or without IFN (for specific people with GT1 and GT4; as recommended by NICE)] were not included in the company's cost effectiveness analysis because, according to the company, these are not used in current NHS practice. Furthermore, despite being included in the scope, the company did not perform subgroup analyses for patients who are co-infected with HIV and post-liver transplantation. The subgroup of patients who are intolerant to or ineligible for interferon treatment were only considered for GT2 TN patients.

The ERG assessment indicated that the model was presented and reported appropriately except for the sensitivity analyses. The company developed a de novo cost effectiveness model to assess the cost effectiveness of G/P compared to nine different comparators: BSC-watchful waiting, DCV/SOF, DCV/SOF/RBV, EBR/GZR, LDV/SOF, OBV/PTV/DSV+DSV ± RBV, PR, SOF/PR, SOF/RBV and SOF/VEL.

The cost effectiveness analyses performed by the company are in line with previous STAs for HCV treatments. The population considered in the cost effectiveness analyses was sub-divided into 26 different subgroups, where patients were stratified by genotypes (GT1, GT2, GT3, GT4, GT5 and GT6), treatment experience (treatment-naïve and treatment-experienced patients), cirrhosis status (cirrhotic and non-cirrhotic patients) and IFN-eligibility (only for GT2 TN patients).

The cost effectiveness model developed for this submission was a Markov model which consists of 13 health states. Non-cirrhotic patients start from states F0-F3, and cirrhotic patients start from F4. All treatment related outcomes (achieving SVR, treatment related adverse events and discontinuation) occur within the first year of the model. Patients who do not achieve SVR are at risk of progressing to more severe states. Patients who reached F4 can progress to DC and HCC states, which may lead to liver transplantation and liver related death. Liver transplantation state was divided into two categories (first year and later years).

The model uses health state based utilities from the literature (utilities that were used in Wright et al. 2006[153] and Ratcliffe et al. 2002[161]) in line with previous STAs for HCV treatments. A utility increment of 0.05 due to SVR is applied based on Shepherd et al. 2007[154] and Hartwell et al. 2011[155]. Treatment-related health utility changes were applied to adjust for the impact on HRQoL of treatment, e.g. due to adverse events.

List prices were used as treatment costs for G/P and the comparator treatments in the cost effectiveness analysis. Health state costs (disease management costs based on disease stage) and other costs for adverse events were based on literature, expert opinion, UK reference costs and previous appraisals for HCV (especially TA430).

It should be noted that while the current model structure does not allow for sequential treatments, in clinical practice, patients who do not achieve SVR (who do not respond to the therapy or discontinue due to adverse events) or who were re-infected after SVR may receive further lines of treatments.

Onward transmission was not included in the economic model. Incorporating onward transmission would require a dynamic transmission model to capture an ongoing risk of infection for individuals in a population, and therefore could not be incorporated into the current modelling framework. Similarly, the company assumed a zero-reinfection probability after reaching SVR and assumed that no natural recovery takes place, despite contrary evidence reported in the clinical literature. However, a scenario analysis by the company showed that the addition of these reinfection probabilities has only minimal impact on the results.

Treatment effectiveness was modelled as the probability of achieving SVR. Other treatment-specific parameters included adverse event rates, treatment duration, and treatment-related utility adjustments. All these parameter estimates were based on naïve indirect comparison of clinical trials assessing the efficacy of G/P and its comparators in the relevant subgroups. The ERG has concerns on the plausibility of this approach, which is not in line with the evidence synthesis best practices and susceptible to bias. Furthermore, some of the SVR rates were derived from very small sample sizes or the effectiveness in a subgroup was assumed to hold in another subgroup. Since SVR probability is the main driver of costs and effectiveness, all these assumptions create a substantial uncertainty on the cost effectiveness of G/P.

Furthermore, it was not clear to the ERG why age-dependent transition probabilities were not updated every year.

The health state utilities from RCTs could have been used by the company in their cost effectiveness analysis instead of the utilities from the literature. The ERG questions to what extend utility values published in 2006 (originating from EQ-5D questionnaires completed in 2002),[153] i.e. before the DAA-era, can be seen as representative of UK patients currently suffering with CHC. Similarly, the RCT-based utility values show a difference in utility with or without SVR ranging from 0.025 to 0.029, substantially lower than the increment of 0.05 applied by the company based on Wright et al. 2006[153] thus raising doubt about the validity of the latter value.

The impact of receiving treatment on QoL during treatment was taken into account in the company model using utility increments and decrements. However, most of these adjustment estimates were based on the same studies as the estimates of SVR rates and AE rates, implying that all comments regarding those (see Section 4.6) apply here as well.

The ERG was unsure about the completeness of the health state cost estimates used in the model, as items such as GP visits and home care costs are not included.

The base-case cost effectiveness results showed that, for non-cirrhotic patients, G/P was always cost effective except for the following two subgroups: GT2 treatment-naïve IFN-eligible patients and GT3 treatment-experienced patients. For the subgroups of non-cirrhotic patients for which G/P was cost effective, the relevant comparator was always no treatment, which resulted in very low ICERs (always below £5,000 per QALY). For patients with compensated cirrhosis, G/P was only cost effective for GT1 treatment-naïve and GT3 treatment-naïve patients. For seven of the 13 subgroups where G/P was not considered cost effective, as it was dominated, G/P could be considered as approximately equivalent (same QALYs at only slightly higher costs, i.e. max £200) to the most cost-effective comparator, SOF/VEL. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was nearly equivalent to SOF/VEL, the most cost effective comparator. In the remaining six subgroups, G/P was clearly not cost effective (ICER above cost effectiveness threshold).

7. OVERALL CONCLUSIONS

7.1 Statement of principal findings

Eighty-one publications (reporting on 79 studies) were identified as meeting the eligibility criteria. Fourteen of these publications, representing seven studies, involved G/P. In addition, information on four further clinical studies of G/P in patients with CHC are included in the company submission. These studies were conducted in special populations of patients with HIV co-infection, renal impairment, failure on prior DAAs and a post-transplant population. Finally, the company mentions two trials in Japanese patients with CHC: CERTAIN-1 and CERTAIN-2. These trials are only minimally discussed in the CS and not included in the economic model because "*these two trials were conducted entirely in Japanese patients*" which "*precludes their generalisability to the UK patient population and subsequently their use in the economic model*", according to the company.[2] Apart from these two trials in Japanese patients, none of the included studies presented comparative data for the licensed treatment duration of G/P with any of the comparators.

The G/P studies included patients with all genotypes; treatment-naïve and experienced patient populations; and patients with 'no cirrhosis and compensated cirrhosis'.

When split by cirrhosis status and previous treatment (naïve or experienced), SVR rates were consistently above 90% for all genotypes, except for GT2/TE/NC ([redacted] in SURVEYOR-II, Part 4; but [redacted] in SURVEYOR-II, Parts 1 and 2), GT3/TE/CC ([redacted] in SURVEYOR-II, Part 2; but [redacted] in SURVEYOR-II, Part 3) and GT6/TN/NC ([redacted] in SURVEYOR-II, Part 4).

[redacted] In studies without a comparator, many treatment arms [redacted]

According to the company, G/P has a favourable safety profile that was similar to placebo and SOF/DCV, and that was similar across durations of eight, 12, and 16 weeks. G/P was well tolerated across a broad and diverse population of patients, including patients with CC, HIV co-infection, and CKD Stage 4/5. Common study adverse drug reactions (ADRs) occurring in ≥5% of patients were headache, fatigue, and nausea. Adverse drug reactions were mostly Grade 1 (mild) in severity. Serious ADRs and ADRs leading to premature study drug discontinuation were rare (≤0.1%).

The results of the company's base-case showed that, for non-cirrhotic patients, G/P was always cost effective except for the following two subgroups: GT2 treatment-naïve IFN-eligible patients and GT3 treatment-experienced patients. For the subgroups of non-cirrhotic patients for which G/P was cost effective, the relevant comparator was always no treatment, which resulted in very low ICERs (always below £5,000 per QALY). For patients with compensated cirrhosis, G/P was only cost effective for GT1 treatment-naïve and GT3 treatment-naïve patients. For seven of the 13 subgroups where G/P was not considered cost effective, as it was dominated, G/P could be considered as approximately equivalent (same QALYs at only slightly higher costs, i.e. max £200) to the most cost-effective comparator, SOF/VEL. Thus, in summary, at a cost effectiveness threshold of £20,000 per QALY gained, G/P was considered cost effective in 13 of 26 subgroups. In seven of the 13 subgroups where G/P was not cost effective, G/P was nearly equivalent to SOF/VEL, the most cost effective comparator. In the remaining six subgroups, G/P was clearly not cost effective (ICER above cost effectiveness threshold).

Probabilistic results were reported by the company as the probability that G/P is cost effective against one single comparator for each subgroup at £20,000 and £30,000 thresholds. However, the ERG showed that including all comparators in the PSA could substantially alter the probability that G/P would be cost effective. The result of the deterministic sensitivity analyses showed that in general the ICER was most sensitive to changes in SVR rates. Two scenario analyses conducted by the company first demonstrated how the cost effectiveness of G/P might change after the CMU price agreement (when comparators from other companies were based on list prices). Second, it was shown that using trial based utilities increased total QALY estimates compared to the base-case when literature based utilities were used as input, without really altering the conclusions from the base-case analyses.

The ERG did not present an alternative base-case, since it was not clear that any alternative base-case assumptions regarding point estimates and structure would be properly justified, as in most situations the assumptions made by the company were reasonable and in line with previous appraisals. Instead of setting a preferred base-case, the ERG conducted a number of exploratory scenario analyses.

In the scenario analyses assumptions surrounding the utility gain due to SVR, impact of the treatment on utility, impact of age on utility were challenged. In addition alternative inputs for transition probabilities in between fibrosis stages and re-infection rates were explored. Even though these scenarios changed the total costs and/or total QALYs estimates, the impact on incremental results were minimal. The cost effectiveness of G/P in each subgroup did not change, hence the cost effectiveness results of the base-case seem to be robust to changes in utility and treatment unrelated clinical model inputs.

The exploratory PSA analyses conducted by the ERG showed that that the inclusion of parameter uncertainty around all SVR and AE rates (which was not included in the company's base-case whenever rates were 100% or 0%) can have a major impact on the G/P cost effectiveness probability for certain subgroups; but also that this impact can go in either direction (although more frequently it decreases the G/P cost effectiveness probability). This was especially striking for GT5 TN NC patients, for whom the company might have overestimated the cost effectiveness probability of G/P by 66 percent. Furthermore, the ERG showed that the PSA outcomes were enormously scattered over the CE plane quadrants for a number of subgroups and illustrated the main limitation of presenting cost effectiveness probabilities only (as in the CS).

7.2 Strengths and limitations of the assessment

The conclusion from the G/P studies is that G/P has high SVR rates in all genotypes. In addition, G/P has a relatively favourable safety and tolerability profile. However, patient numbers for most GT4, GT5 and GT6 populations are very low, often less than 10 patients in each group. Only four out of the 24 subgroups included more than 100 patients (GT1/TN/NC, GT1/TE/NC, GT2/TN/NC and GT3/TN/NC). Therefore, the uncertainty around SVR rates in most subgroups is considerable.

The main concern regarding clinical effectiveness is that data (for SVR12 and AEs) for G/P and comparators were taken from single arms. Therefore, the comparisons for SVR12 rates and AEs between G/P and comparators relies on a naïve indirect comparison. The company does not present any information about the comparability of populations between G/P studies and comparator studies; and about how response and adverse events for comparator studies were selected and whether all possible sources were used. In most cases the sources for SVR rates and AEs for comparators are the same as in TA430 (Sofosbuvir-velpatasvir for treating chronic hepatitis C). Therefore, the same critique as for TA430 applies: these methods increase the chance of bias and cherry-picking.