Molnupiravir, remdesivir and tixagevimab plus cilgavimab for treating COVID-19 [ID6261]: EAG critique of the post-appeal evidence submitted for remdesivir

Produced by	Sheffield Centre of Health and Related Research (SCHARR), The University of Sheffield
Authors	Matt Stevenson, Professor of Health Technology Assessment, SCHARR, University of Sheffield, Sheffield, UK Andrew Metry, Research Associate, SCHARR, University of Sheffield, Sheffield, UK Abdullah Pandor, Senior Research Fellow, SCHARR, University of Sheffield, Sheffield, UK Shijie Ren, Senior Research Fellow, SCHARR, University of Sheffield, Sheffield, UK Mark Clowes, 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	20th of October 2023

Source of funding: This report was commissioned by the NIHR Evidence Synthesis Programme as project number 135564.

Declared competing interests of the authors

No author declares a conflict of interest.

1 Introduction

This document should be read in conjunction with the initial EAG report[1], erratum[2], and a subsequent EAG report[3] discussing additional analysis undertaken after NICE issued its Appraisal Consultation Document. These provide more details on the work which has been undertaken for the treatments assessed in-hospital for severe COVID-19, which was NICE ID4038.

The final draft guidance for ID4038[4] did not recommend the use of molnupiravir, remdesivir and tixagevimab plus cilgavimab and the manufacturers of these interventions appealed the decision. In order for expediency with respect to the remaining interventions in ID4038, NICE provided a new ID number, ID6261 for molnupiravir, remdesivir and tixagevimab plus cilgavimab, with these appraisals to be concluded after the appeal.

The appeal panel upheld the appeal on multiple appeal points made by the three companies with the evaluation returned to the appraisal committee who must '*take all reasonable steps to address the following issues before publishing final guidance.*' Full details of the appeal decision are available online.[5]

NICE has entered into discussions with all three companies. The first re-appraisal is that of remdesivir. Following discussions between NICE and Gilead (the manufacturers of remdesivir) an agreement was reached that was summarised by Gilead as follows: "*NICE has agreed that four aspects must be addressed, namely (1) the opportunity for Gilead to make a targeted evidence submission, (2) an opportunity for engagement with the evidence assessment group (EAG) on economic modelling for remdesivir, (3) the ability for Gilead to comment on the EAG report following model adaptation and (4) an agenda for the third appraisal committee meeting (ACM) which allows appropriate room for discussion of the relevant evidence for remdesivir*".[6]

Section 2 summarises the targeted evidence review (and multiple subsequent documents) submitted by the company in response to the Appeal decision to assess the clinical effectiveness of remdesivir in three subgroups of patients with severe COVID-19 that required hospitalisation. These subgroups were patients requiring low-flow supplemental oxygen; children, and immunocompromised patients. Section 3 provides the EAG critique of the clinical evidence submitted by Gilead and its search strategy and proposes alternative evidence sources that the EAG thinks may be more appropriate.

As the EAG was writing up its report, Gilead sent its own economic model to the EAG. There was insufficient time to critique the implementation of the model, but the EAG noted that when a comparison was made between incremental cost-effectiveness ratios (ICERs) in terms of cost per quality-adjusted

life years (QALYs) gained generated by the company's model and the EAG's model, with attempts to ensure comparable input parameters, that the ICER was moderately lower in the EAG's model. Given that the EAG's model had been scrutinised by multiple companies, had been discussed at previous committee meetings, and the EAG believes it has additional flexibility to that of the company's model the EAG has maintained the use of its model which may be favourable to the intervention.

Section 4 details the changes introduced to the model by the EAG to consider the new evidence and selected subgroups. Section 5 provides the cost-effectiveness results generated by the EAG. The EAG's model produces ICERs for remdesivir compared with standard of care (SoC). Section 6 provides a discussion on the results generated by the EAG.

It is unclear whether tocilizumab would be a comparator. The NICE final draft guidance for ID4038[4] stated that tocilizumab was an option for treating adults with COVID-19 who are having systemic corticosteroids and need supplemental oxygen or mechanical ventilation, and thus there is potential for adult patients receiving low-flow oxygen (LFO) to have tocilizumab. Discussions with NICE did not provide a definitive conclusion on whether tocilizumab was a comparator and therefore, following guidance from NICE, the EAG has provided the results comparing remdesivir with tocilizumab in appendices should the Appraisal Committee find these results informative. There is a confidential patient access scheme (PAS) for tocilizumab which, following NICE guidance, is not considered within the report. A confidential appendix incorporating the PAS for tocilizumab has been provided to the NICE appraisal committee.

2 A summary of the company's targeted submission

The company submitted new evidence on remdesivir for the treatment of COVID-19 in the form of a targeted evidence submission (TS) on the 6th of September 2023. The EAG deemed that there were insufficient details in the TS and relayed this to NICE who scheduled a meeting between the company, the EAG and NICE representatives. Following this additional evidence was provided by the company in stages. On the 15th of September 2023, the company provided a draft clinical systematic literature review (SLR) technical report and an extraction grid relating to hospitalised patients. On the 9th of October 2023, the company further provided the clinical rationale for selecting the subgroups on which it focussed and a bias assessment using NICE-preferred tools in the form of an extraction grid.

Given the report deadline of the 20th of October 2023, the EAG could not follow best practice and independently undertake a systematic review and meta-analysis of the clinical evidence relevant to the decision problem. Following discussions with NICE, a pragmatic, alternative approach was undertaken relying on a brief critique of the company's TS (including additional supporting information) of the clinical evidence.

2.1 Critique of the company's targeted submission

Although remdesivir (Veklury) is indicated for the treatment of coronavirus disease 2019 (COVID-19) in:

- adults and paediatric patients (at least 4 weeks of age and weighing at least 3 kg) with pneumonia requiring supplemental oxygen (low- or high-flow oxygen or other non-invasive ventilation at start of treatment)
- adults and paediatric patients (weighing at least 40 kg) who do not require supplemental oxygen and who are at increased risk of progressing to severe COVID-19

the scope of the TS focused on populations in which the company considered remdesivir to be most effective. These populations included patients requiring LFO, children and immunocompromised patients. A definition of each of these patient populations, as provided by the company in the TS is reproduced in Table 1.

Table 1: Definition of relevant patient populations for remdesivir (reproduced from Table 3, TS)

Patient population	Definition
Low-flow oxygen (LFO)	Patients requiring oxygen delivered by a simple face mask or nasal canula at a flow rate usually up to 15 litres/min as per the NICE COVID-19 rapid guidelines[7]
Children	The paediatric patient population includes: • paediatric patients (at least 4 weeks of age and weighing at least 3 kg) with pneumonia requiring supplemental oxygen (low- or high-flow oxygen or other non-invasive ventilation at start of treatment) • paediatric patients (weighing at least 40 kg) who do not require supplemental oxygen and who are at increased risk of progressing to severe COVID-19 as outlined in the summary of product characteristics (SmPC) for remdesivir[8]
Immunocompromised patients	Patients who have a weakened immune system due to a particular health condition or patients who are on medication or treatment that suppresses their immune system

Following a request from NICE, the company provided a detailed rationale for the selected subgroups.[9] The EAG has briefly summarised the company's rationale.

- The LFO subgroup was considered as a distinct and readily defined population[7, 10-13] and the European Society of Clinical Microbiology and Infectious Diseases Guidelines[11, 12] conditionally recommend remdesivir for use in hospitalised patients requiring no or LFO but not in patients requiring high-flow oxygen.
- In paediatric patients, remdesivir is the only available licensed treatment option for COVID-19 and there is inequity of access to comprehensive clinical care for this group.
- The immunocompromised subgroup was considered to experience worse clinical outcomes with COVID-19 than the general population and comprise less than 1% of the UK population, but account for a large proportion of those hospitalised with, of dying from, COVID-19. In addition, nirmatrelvir and ritonavir (Paxlovid) is the only recommended antiviral and is not appropriate for all immunocompromised patients (including immunocompromised patients requiring supplemental oxygen).

As noted in the addendum to the TS (page 6), the company states that a '*dedicated systematic literature review (SLR) for patients receiving LFO was not feasible due to time constraints, Gilead leveraged existing SLRs conducted for inpatients with COVID-19... The technical reports of the clinical and economic SLR contain all relevant information required and expected of a high-quality systematic search, including a full description of the identification of studies, search strategy, search terms used and study selection criteria. Furthermore, the SLRs reported a PRISMA flow chart for the identified studies, a summary of the included clinical studies as well as a risk of bias assessment using the York Centre for Reviews and Dissemination checklist.[14] To complement the SLRs which focus on the inpatient sector, Gilead has conducted additional targeted searches for LFO, immunocompromised and paediatric patients specifically. These searches were conducted using Google scholar and leveraged search terms derived from the PICO framework, targeting LFO, immunocompromised and paediatric patients specifically.[15]*'

Whilst the EAG acknowledges the limitations and time constraints to undertake a full systematic review following the Appeal decision, the review methods, and processes in the TS (and accompanying technical report) are neither fully transparent nor reproducible, and the strengths and limitations of the company's review process are not fully acknowledged. For example, in the absence of clear and explicit review eligibility criteria in the TS (including supporting information), it is unclear how the company's broader systematic review of COVID-19 treatments in the inpatient setting was used and informed the TS, which focused on a subgroup of patients requiring LFO, children and immunocompromised patients; the advantages and disadvantages of using Google Scholar as a standalone source for the TS (the EAG notes that the use of Google Scholar as a standalone source for systematic review searches is not usually recommended or considered a replacement for traditional academic citation databases);[16-18] and how many (and which) primary studies met the review inclusion criteria (including a table of excluded studies with reasons) for the TS. The EAG further notes that the broader review only included primary studies (interventional and observational studies) and excluded existing systematic reviews and (network) meta-analyses (Company's Clinical SLR Technical report, Table 7, page 18-19) – a critique of the search strategy for the broader review is contained in Appendix 1. In contrast, the TS appears to have included existing systematic reviews and (network) meta-analyses and other studies (TS, Table 4, page 17; Figure 6, page 18). It is unclear why the selection of study designs was notably different between the broader review and the TS.

Although no narrative or statistical synthesis of the results was undertaken or reported in the broader review, the TS summarised the results of selected systematic reviews and primary studies. These included data on mortality, clinical improvement, time to discharge, recovery, hospital readmission, progression to invasive mechanical ventilation or death, and long COVID syndrome for patients requiring LFO, for children and for immunocompromised patients (TS, page 17-28). The EAG was

unable to undertake independent quality assessments of all included / reported studies, due to the multiple submissions and varied timing of the company's additional supporting information and the deadline for this report.

In the subsequent subsections, the EAG critique has been limited to key data inputs in the economic model for remdesivir in the LFO population, namely: mortality; clinical improvement; and time to hospital discharge. No critique of the evidence has been provided for the paediatric and immunocompromised populations (see TS, page 25-28 for details of supporting evidence) as the TS (page 38) states that, *'the evidence for the paediatric patient population receiving remdesivir does consist of non-comparative, single arm trials. Given this lack of comparative data, deriving incremental cost-effectiveness estimates against a SOC comparator are not feasible*.' No comment in the TS was made for the immunocompromised population where ICERs were also not provided.

2.1.1 LFO population and mortality

As noted earlier, it is unclear how the company's broader systematic review was used to inform the TS and which primary studies were potentially eligible for inclusion. Nevertheless, the TS (page 17) and TS addendum (page 6) identified three potentially relevant systematic reviews (network meta-analyses of RCTs)[19-21] which provided data on relevant primary studies and mortality outcomes in hospitalised adult COVID-19 patients receiving LFO. A summary of each systematic review is provided in Table 2 and an assessment of methodological quality is provided in Appendix 2.

The company selected the 28-day mortality data from the Huang *et al.*[19] review to inform the base case of the EAG economic model, as the '*…paper was published most recently, used a risk ratio as the outcome measure – which aligns with the EAG model – and reports a result that falls in between the results reported by both Beckerman et al. and Amstutz et al., therefore representing a more balanced outcome for assessment in the face of uncertainty'* [TS addendum, page 7]. In addition, using the AMSTAR-2 critical appraisal tool for systematic reviews,[22] the company considered the Huang *et al.*[19] review '*…to have more than one weakness but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review* [TS addendum, Bias assessment using NICE preferred tools, Excel Spreadsheet]'. In contrast, the company considered the Beckerman *et al.*[20] review to have a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest. However, the EAG notes that purely based on the details provided in Appendix 2 which provides a summary of the company's AMSTAR-2 ratings for included systematic reviews, the company's assessment gradings appear to look similar for Beckerman *et al*.[20] and Huang *et al*.[19] across most critical domains (question 2, 4, 7, 9 and 15), except questions 11 and 13, which suggest both studies may have a one potential critical flaw.

Table 2: Summary of systematic reviews informing mortality outcomes (adapted, TS addendum, Table 1, page 6 and Table 2, page 8)

Author, year	LFO definition	Included studies	Analysis type	Data search	Population details	Mortality outcomes	
Huang *et al.*[19]	Four category ordinal scale: (1) not requiring supplemental oxygen; (2) requiring supplemental low-flow oxygen; (3) requiring non-invasive ventilation or high-flow oxygen; (4) requiring invasive mechanical ventilation or extracorporeal membrane oxygenation (ECMO).	Ali *et al.*[23] Beigel *et al.*[24] Wang *et al.*[25]	Aggregate	January 2020 to February 2023	Unvaccinated[f]	Remdesivir: 56/695[d] Control: 90/634[d]	Risk ratio: 0.59 (95% CI: 0.43, 0.80)
Beckerman *et al.*[20]	Low-flow oxygen defined as either hospitalized and requiring any supplemental oxygen or hospitalized requiring low-flow supplemental oxygen, depending on the study	Beigel *et al.* [24a] Spinner *et al.*[26 a] (Kalil *et al.* [27])[a, b]	Aggregate	Up until April 2021	Unvaccinated[f]	Remdesivir: 21/560d, [e] Best Supportive Care: 29/239d, [e]	Risk ratio: 0.24 (95% CrI: 0.11, 0.48)
Amstutz *et al.*[21]	WHO ordinal scale levels (no distinction between no, and low flow, oxygen)	Beigel *et al.*[24] Wang *et al.*[25] Spinner *et al.*[26] Ali *et al.*[23] SOLIDARITY[28 c] DisCoVeRy [29, 30]	Individual patient level data (10,480 patients)	Up until April 2022	Unvaccinated	Remdesivir: 409/4473[d] No Remdesivir: 465/4159[d]	Adjusted odds ratio: 0.80 (95% CI: 0.70, 0.93) (Analysis includes patients with no or low-flow oxygen requirements as a single patient population)

a: Based on list of study presented in table 3 of the Beckerman *et al.* paper; b: Results reported separately for remdesivir + baricitinib; c: SOLIDARITY data cited individually in the Amstutz paper, including FIN-SOLIDARITY, NOR-SOLIDARITY and additional WHO-SOLIDARITY; d: Event/ total; e: Sample size data reported in this table reflects the later mortality assessment; f: Study does not report distinctively that it assesses an unvaccinated population, but no vaccination can be assumed
CI – confidence interval; CrI – credible interval

The EAG prefers to use the individual patient data (IPD) meta-analysis results, conducted by Amstutz *et al*.[21], to better inform the base case of the EAG economic model. An IPD meta-analysis approach has advantages over a standard meta-analysis based on aggregate data by: increasing the quantity and quality of the data available; standardising outcome and subgroup definitions across trials; maximising power to assess the heterogeneity of the treatment effect across subgroups, and by allowing adjustment for baseline differences.[31, 32] The company's critical appraisal, using the AMSTAR-2 tool,[22] also considered this systematic review to have more than one weakness but no critical flaws and considered it to provide an accurate summary of the results of the included studies (TS addendum, Bias assessment using NICE preferred tools, Excel Spreadsheet).

Although the Amstutz *et al.* review[21] included the broadest set of studies, summarising results from 8 RCTs (6 separate trials), while Huang *et al.*[19] and Beckerman *et al.*[20] summarised results of 3 and 2 trials respectively (see Table 2), the company states that '*Amstutz et al. was not recommended as a base case input for 28-day mortality as it focused on a slightly different patient population, i.e. patients with no oxygen or LFO requirements*' (TS addendum, page 11). However, the EAG notes that Amstutz *et al*.[21] undertook a sensitivity analysis, which investigated oxygenation in more detail (Amstutz *et al.*[21] Appendix Figure S8, page 36 – summarised in Appendix 3), and found that '*patients who were receiving no oxygen at baseline derived a similar relative benefit (adjusted odds ratio [aOR] 0·86, 95% CI: 0·53–1·39 with and 0·77, 0·34–1·74 without additional WHO Solidarity data) to patients receiving low-flow oxygen (aOR 0·79, 0·68–0·92 with and aOR 0·59, 0·43–0·82 without additional WHO Solidarity data).*' As Amstutz *et al*.[21] did not show a significant difference between the no oxygen and the LFO groups and the NICE rapid guideline[7] (p100) stated that the '*for the WHO-SOLIDARITY trial, the panel agreed to include people having supplemental oxygen in the meta-analyses for people having low-flow or no oxygen at baseline*' the EAG used the results from the LFO and no oxygen groups combined, to reduce the uncertainty in the estimate of the efficacy of remdesivir. However, the EAG has also run analyses excluding data from SOLIDARITY[28] and used data only for patients requiring LFO. Odds ratios were transformed into hazard ratios (HRs) as described in Section 4.

2.1.2 LFO population and clinical improvement

The TS (page 19 and TS addendum page 11) appears to have identified and included one potential study by Garibaldi *et al*.[33] to inform the clinical improvement endpoint for LFO patients in the EAG model. This retrospective, multicentre comparative effectiveness study from the US, examined the effectiveness of remdesivir administration in hospitalised COVID-19 patients between the 23rd of February 2020 and the 11th of February 2021. The primary outcome was time to clinical improvement from the first day of remdesivir treatment (defined as a 2-point decrease in the 8-point WHO severity score or discharged alive from the hospital without worsening of the WHO severity score within 28 days). Remdesivir recipients were matched to controls using time-dependent propensity scores and cox

proportional hazards regression models were applied to estimate the treatment effect on the outcomes of interest. Of the 20,966 matched individuals receiving LFO (10,314 patients received remdesivir and 10,652 matched controls) remdesivir recipients were statistically significantly more likely to achieve clinical improvement by 28 days (adjusted HR 1.23, 95% CI: 1.19–1.27; median of 6 days for remdesivir compared to 7 days in controls).[33]

Although the company provided an assessment using the criteria reported in the NICE real-world evidence framework[34] and a tabulated summary of the methods used to minimise the risk of bias in the study by Garibaldi *et al*.[33] (reproduced in Appendix 4), a full critique of the strengths and limitations of this study was not adequately discussed. Garibaldi *et al*. [33] highlighted a number of limitations, most notably being unable to match approximately half of the remdesivir patients, unmeasured confounders and that the study was conducted prior to the widespread use of vaccines and the emergence of variants such as Delta and Omicron, which could impact generalisability. Despite these limitations, the company selected Garibaldi *et al*.[33] for the clinical improvement outcome due to the large sample size of the study (TS addendum, page 12).

Moreover, the TS addendum (page 12) states that *'It should be noted that Beckerman et al. report results for a similar outcome, which they label "recovery", defined as "either recovery from COVID-19 or discharge from hospital".[20] Given the similarity to the clinical improvement outcome, the outcome from Beckerman et al. might also be considered as additional evidence. Regardless, both Garibaldi et al. and Beckerman et al. report similar results, thus indicating high consistency across the two different studies (aHR 1.23, 95% CI 1.19, 1.27; RR 1.17, 95% CI 1.09, 1.28*)'. The company's TS does not provide sufficient details of these reviews, including: the individual RCTs that provided data for the recovery endpoint, including meta-analysis results; the quality, strengths and limitations of this evidence; and why this evidence was not considered relevant. However, as mentioned earlier, the EAG notes that the company considered the Beckerman *et al*. review[20] to have a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest.

The EAG has conducted analyses with, and without, a positive impact on remdesivir in terms of clinical improvement. When a positive impact was assumed, data from Covid-NMA[35] was used as previously assumed by the EAG.

However, as in the original modelling,[3] the EAG assumed no benefit in clinical improvement for remdesivir when an improvement in time to discharge was assumed as the ACTT-1 values incorporated clinical improvement as the time to discharge relative risk (RR) was for discharge or National Early

Warning Score ≤2 for 24 hours. Therefore, a RR of unity was assumed for clinical improvement in all 3 efficacy scenarios to reduce the possibility of double counting.

2.1.3 LFO population and time to hospital discharge

The TS (page 19-21 and TS addendum page 12) appears to have identified 2 RCTs that reported data on time to discharge from hospital: the ACTT-1 study[24] (n=1062), and Spinner *et al*.[26] (n=584; which only provided time to discharge curves in a supplementary analysis). The TS provided limited details of these studies (no quality appraisals appear to have been conducted for the Spinner *et al*. study[26]) and stated (page 20) that '*It should be noted that neither the ACTT-1 nor the results from Spinner et al.[26] for the TTD outcome were analysed for a patient population receiving low-flow oxygen, which is the patient population in which remdesivir is most effective*.' In summary, the company selected the time to hospital discharge data from the ACTT-1 study[24] (n=1062) to inform the EAG model, primarily due to larger sample size. No further rationale was provided, and a full critique of the strengths and limitations of this study was not adequately discussed. In the ACTT-1 study,[24] patients in the remdesivir group had a shorter time to discharge or to a National Early Warning Score of 2 or lower than those in the placebo group (median, 8 days vs. 12 days; HR, 1.27; 95% CI: 1.10-1.46). The EAG notes that the National Early Warning Score includes six physiological measures with total scores ranging from 0 to 20, with higher scores indicating greater clinical risk. It is unclear how, if at all, the National Early Warning Score is currently being used to safely discharge patients from UK hospitals. The company also notes that it '*…is aware of the committee's preference to exclude TTD effects for all treatments following the last two appraisal committee meetings and is conscious that TTD effects for remdesivir might not be considered by the committee to be aligned to previous recommendations made for tocilizumab, Paxlovid and Sotrovimab'*.

The EAG has conducted analyses with, and without, a positive impact on remdesivir in terms of time to hospital discharge. When a positive impact was assumed, the EAG used data from ACTT-1,[24] as did the company.

3 The EAG's critique of the company's search strategies and selected clinical evidence

For a critique of the use of Google Scholar for obtaining estimates of clinical efficacy, see Section 2. A critique of the broader search strategy is contained in Appendix 1.

The systematic reviews of economic evidence (comprising reviews of cost-effectiveness, health-related quality of life and cost and resource use) were originally conducted in July 2020 (then updated in May 2021 and December 2022). Databases covered MEDLINE and EMBASE plus international HTA websites, conference proceedings and registries of cost-effectiveness and utility studies. While the ERG usually recommends searching EconLit for the purpose of economic SLRs, this is not essential. Study design terms are based on the unvalidated (but widely used) Scottish Intercollegiate Guidelines Network filters, with terms added.

Unusually, the company have included terms relating to the interventions of interest in all three of the reviews. For reviews of cost or utility data (HRQoL), guidance recommends excluding interventions and using population terms only, with the addition of an appropriate filter.

The ERG speculates that the decision to include intervention terms may have been taken for practical reasons (due to the high prevalence of COVID-19 during the acute phase of the pandemic) although this could perhaps have been addressed more effectively for the review of HRQoL evidence by limiting the population to inpatients.

The net result of this approach is that papers containing useful data about the costs of standard of care (e.g. hospitalisation) or other drugs (outside of the company's scope) will be missing from the models proposed in the TS. The impact of this is unknown although the EAG expects that this will not be a significant limitation given that the EAG model has been scrutinised by many stakeholders and costs associated with hospital care were amended followign stakeholder comments.

4 Amendments to the EAG's model and the analyses undertaken

Section 4 is subdivided into the amendments required within the model in order to assess the subgroups put forward by the company and the analyses which were undertaken.

4.1 Amendments made to the EAG's model

The model required changing to take into consideration the fact that the company was positioning remdesivir only for patients receiving LFO. This meant that all patients were placed at ordinal scale 5 rather than divided amongst ordinal scales 5 to 7 as modelled previously.

Patients receiving LFO are less severe than patients receiving high-flow oxygen or mechanical invasive ventilation and the underlying mortality rate used previously needed to be changed to take this into account. The company suggested a value of 10% at 15 days for patients on LFO, based on 14% mortality in SOLIDARITY[28] at day 15, however the EAG used an alternative value of 14.0% at 28 days (432 deaths out of 3076 patients who needed oxygen but without ventilation who did not receive remdesivir as reported at Amstutz *et al*.[21])

The 28-day mortality data from Amstutz *et al*.[21] and Huang *et al*.[19] were reported as ORs and RRs, however the EAG's model uses HRs. To estimate HRs, the goal seek function of Excel was used to calculate the HRs that would generate the same clinical outcomes as reported in Amstutz *et al*.[21] and Huang *et al*.[19]. Table 3 provides the mortality data from the studies and the corresponding HRs calculated by the EAG. The ORs and RRs reported are midpoints with 95% CIs, with the values used for the mean, mean-low and low efficacy scenarios calculated by the EAG. These scenarios, and the rationale for choosing them, are described in Section 4.2.

Table 3: ORs/RRs of the mortality data used in the EAG's model with the corresponding calculated HRs time to death

Source for mortality data relating to remdesivir	ORs/RRs calculated by the EAG or reported in study	HR calculated by the EAG
Amstutz *et al.*[21] including SOLIDARITY[28] (mean efficacy)	0.792	0.817
Amstutz *et al.*[21] including SOLIDARITY[28] (low efficacy)	0.919	0.930
Amstutz *et al.*[21] including SOLIDARITY[28] (mean-low efficacy)	0.856	0.865
Amstutz *et al.*[21] excluding SOLIDARITY[28] (mean efficacy)	0.598	0.635
Amstutz *et al.*[21] excluding SOLIDARITY[28] (low efficacy)	0.817	0.839
Amstutz *et al.*[21] excluding SOLIDARITY[28] (mean-low efficacy)	0.707	0.723
Huang *et al.*[19] (mean efficacy)	0.597	0.559
Huang *et al.*[19] (low efficacy)	0.800	0.773
Huang *et al.*[19] (mean-low efficacy)	0.699	0.682

OR – odds ratio; RR – relative risk; HR – hazard ratio.

The company did not provide ICERs for children and immunocompromised patients, but the EAG has provided exploratory analyses assuming that only patients receiving LFO are considered. This subgroup was chosen as the EAG was aware that the European Society of Clinical Microbiology and Infectious Diseases Guidelines[11, 12] conditionally recommend remdesivir for use in hospitalised patients requiring no or low-flow oxygen, but not in patients requiring high-flow oxygen. The analyses undertaken by the EAG have assumed that the efficacy values used in LFO patients are generalisable to children and the immunocompromised which may not be correct. Further, these analyses are populated with some values identified from non-systematic reviews, however, the EAG believes that these analyses will be informative to the Appraisal Committee.

For children, the average age of hospitalised patients was arbitrarily reduced to 15 years. The underlying probability of death at 28 days was set to two alternate values. The first value was that reported in Ward *et al.*[36] of 48 deaths from 10,540 hospitalisations within 28 days (0.45%) with the second value being that associated with Wilde *et al.*[37] of 55 deaths at any time during the study period from 29,230 patients with a first SARS-CoV-2-related hospitalisation (0.19%). The average length of stay for children was considerably shorter than for adults with Wilde *et al.*[37] reporting a median length of stay of 2 days, with an interquartile range of 1 to 4 days. Due to the structure of the model, which adjusted the Kaplan-Meier plot from the control arm of the RECOVERY study[38] 100% of children patients with need of supplemental oxygen were assumed to be discharged at 28 days resulting in an average length of stay of around 5 days, which was the minimum length of stay that could be modelled.

For immunocompromised patients, the EAG identified a paper that provided data on the outcomes of immunocompromised patients during the Omicron SARS-CoV-2 variant.[39] This reported that from 4585 patients broadly-defined as immunocompromised there were 4585 hospitalisations and 1145 deaths resulting in 24.98% of hospitalisations resulting in death. This percentage was broadly similar for patients stringently-defined as immunocompromised and whether or not the patient had three doses of a COVID-19 vaccine. The EAG notes that the definition of death included patients who did not die in hospital and may therefore overestimate the probability of deaths following hospitalisation, but the extent of the overestimation is unknown. Due to the absence of data the average age of hospitalised patients and average length of stay was left unchanged from previous modelling.

4.2 The scenarios undertaken

The EAG has run 27 scenarios all of which assume a positive impact of remdesivir on mortality; these are provided in Table 4. The values assumed for tocilizumab in each scenario are contained in Appendix 5. Appendices 6 and 7 contain the comparative results between remdesivir and tocilizumab.

Scenarios 1-9 assume no differences in either clinical improvement or time to discharge; Scenarios 10-18 assume differences in clinical improvement; Scenarios 19-27 assuming differences in time to discharge but not in clinical improvement due to the risk of double-counting in ACTT-1.[24] Scenarios 25 and 26 most closely resemble that of the company's base case with the differences being that the company uses Garibaldi *et al*.[33] for clinical improvement the EAG analysis assumes no clinical improvement benefit and that in the mean scenario the company assumes the midpoint value whereas the EAG has used the calculated mean from the distribution.

Each block of nine scenarios are the combinations of three mortality estimates for remdesivir (Amstutz *et al*.[21] with SOLIDARITY[28], Amstutz *et al*.[21] without SOLIDARITY[28] and Huang *et al.*[19]) and three assumed efficacies levels (mean, low and mean-low). The efficacy values related to mortality for these studies are shown in Table 3.

The mean efficacy value was the expected mean from the specified distribution (calculated by the EAG) whilst the low efficacy value used the more unfavourable 95% confidence limit. As the ICERs for remdesivir for adult patients receiving LFO were below £20,000 using the mean values, analyses using the more favourable 95% confidence limit were not undertaken, and instead mean-low efficacy analyses were run which averaged the value from the mean and low scenarios. This approach was deemed by the EAG to provide useful granularity between the mean and low scenarios and not result in data overload for the Appraisal Committee. The rationale for exploring worse mortality benefit from that observed in the studies is due to the change in circumstances since the studies were conducted which

include changes in: the SARS-CoV-2 variant in circulation; the vaccination status of patients; the prior infection status of patients; and improvements in SoC across time.

Previously, the EAG capped values in the low efficacy scenarios when it was assumed there was no benefit for mortality in order that the treatments evaluated do not, on balance, harm patients. That is, at the very worst, the treatments would produce identical QALYs to SoC. However, as the HRs used for the risk of mortality for remdesivir are all below 1, no capping was applied as the EAG believes it plausible that other aspects such as time to discharge and clinical improvement could be worse as a by-product of preventing death.

Table 4: Parameter values used in the EAG's analyses

Scenario	Study used for remdesivir	Efficacy scenario	Remdesivir parameters*
1	Amstutz *et al.*[21†]	Mean	0.817, unity, unity
2	Amstutz *et al.*[21†]	Low	0.930, unity, unity
3	Amstutz *et al.*[21†]	Mean-Low	0.865, unity, unity
4	Amstutz *et al.*[21]	Mean	0.635, unity, unity
5	Amstutz *et al.*[21]	Low	0.839, unity, unity
6	Amstutz *et al.*[21]	Mean-Low	0.723, unity, unity
7	Huang *et al.*[19]	Mean	0.559, unity, unity
8	Huang *et al.*[19]	Low	0.773, unity, unity
9	Huang *et al.*[19]	Mean-Low	0.682, unity, unity
10	Amstutz *et al.*[21†]	Mean	0.817, 1.040, unity
11	Amstutz *et al.*[21†]	Low	0.930, 0.990, unity
12	Amstutz *et al.*[21†]	Mean-Low	0.865, 1.015, unity
13	Amstutz *et al.*[21]	Mean	0.635, 1.040, unity
14	Amstutz *et al.*[21]	Low	0.839, 0.990, unity
15	Amstutz *et al.*[21]	Mean-Low	0.723, 1.015, unity
16	Huang *et al.*[19]	Mean	0.559, 1.040, unity
17	Huang *et al.*[19]	Low	0.773, 0.990, unity
18	Huang *et al.*[19]	Mean-Low	0.682, 1.015, unity
19	Amstutz *et al.*[21†]	Mean	0.817, unity, 1.270
20	Amstutz *et al.*[21†]	Low	0.930, unity, 1.100
21	Amstutz *et al.*[21†]	Mean-Low	0.865, unity, 1.187
22	Amstutz *et al.*[21]	Mean	0.635, unity, 1.270
23	Amstutz *et al.*[21]	Low	0.839, unity, 1.100
24	Amstutz *et al.*[21]	Mean-Low	0.723, unity, 1.187
25	Huang *et al.*[19]	Mean	0.559, unity, 1.270
26	Huang *et al.*[19]	Low	0.773, unity, 1.100
27	Huang *et al.*[19]	Mean-Low	0.682, unity, 1.187

† Including data from SOLIDARITY[28]

*Parameter values are: hazard ratio for time to death; relative risk for clinical improvement; hazard ratio for time to discharge

5 The results generated by the EAG

Results are presented for the three patient subgroups: adults requiring LFO; children (also assumed to require LFO) and immunocompromised patients (also assumed to require LFO). The EAG highlights that time to discharge is much more influential on the ICERs and patient outcomes than changes in clinical improvement. As such, there are only small differences between the results obtained in Scenarios 1 to 9 and those obtained in Scenarios 10 to 18.

The comparative results between tocilizumab and remdesivir are contained in Appendix 6 and Appendix 7. These results use the list price for tocilizumab; results with the PAS discount applied for tocilizumab are reported in a confidential appendix.

5.1 ICERs estimated by the EAG for remdesivir when treating adult patients requiring LFO

Figure 1 shows the incremental net monetary benefit (NMBs) values for remdesivir when compared to SoC for treating patients requiring LFO at an ICER threshold of £20,000, whilst Figure 2 presents these values at the £30,000 threshold. The ICERs for remdesivir compared with SoC are reported in Appendix 6.

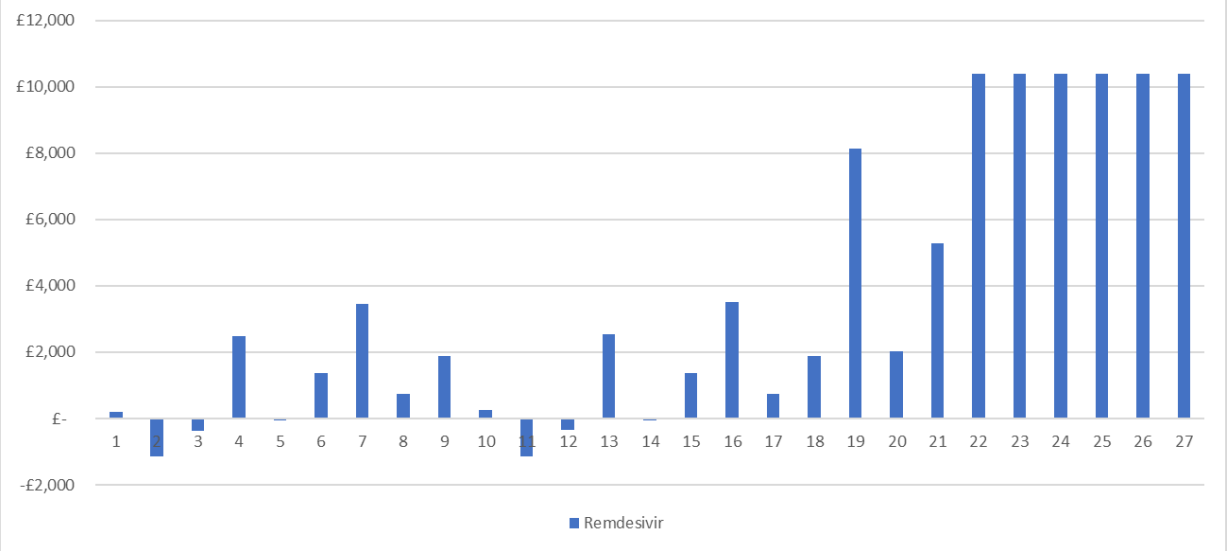

Figure 1: Incremental NMB results for adults receiving LFO at an ICER threshold of £20,000

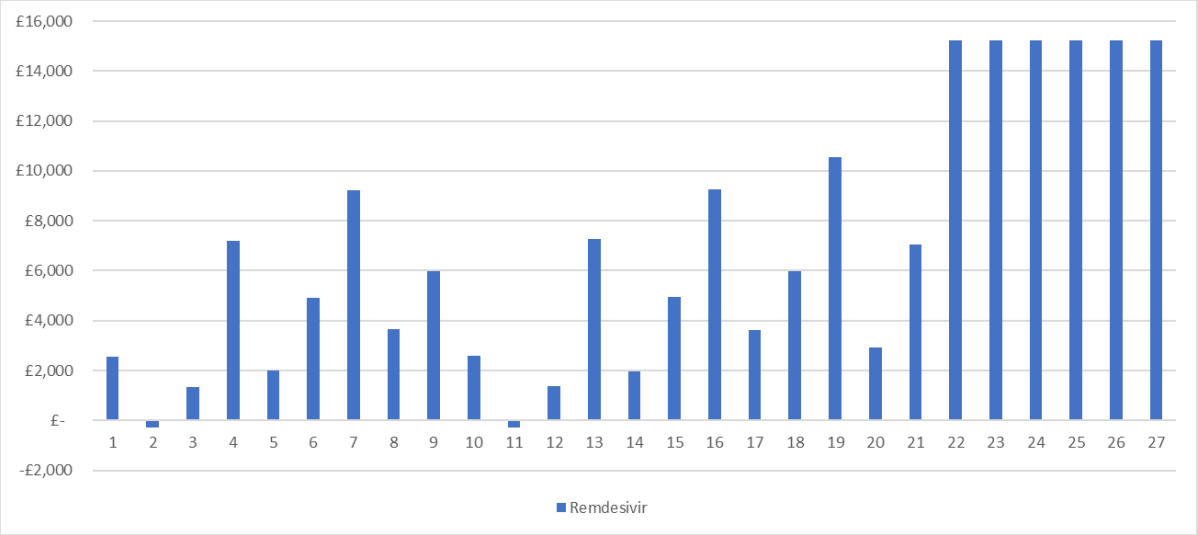

Figure 2: Incremental NMB results for adults receiving LFO at an ICER threshold of £30,000

In the EAG analyses, remdesivir has an ICER compared with SoC above £30,000 in Scenarios 2 and 11, these scenarios both assume low efficacy taken from Amstutz *et al*.[21] when data from SOLIDARITY[28] was included, and that remdesivir has no benefit on time to discharge.

Scenarios 3, 5, 12 and 14 are estimated to have ICERs compared with SoC above £20,000. Scenarios 3 and 12 both assume mean-low efficacy taken from Amstutz *et al*.[21] when data from SOLIDARITY[28] was included and that remdesivir has no benefit on time to discharge. Scenarios 5 and 14 both assume low efficacy taken from Amstutz *et al*.[21] when data from SOLIDARITY[28] was excluded and that remdesivir has no benefit on time to discharge.

5.2 Exploratory ICERs estimated for children

Figure 3 shows the NMBs values for remdesivir when compared to SoC for treating children requiring LFO at an ICER threshold of £20,000, whilst Figure 4 presents these values at the £30,000 threshold. The ICERs are reported in Appendix 8. Both figures present NMBs using probability of death taken from in Ward *et al*.[36] (0.45%) and from Wilde *et al*.[37] (0.19%).

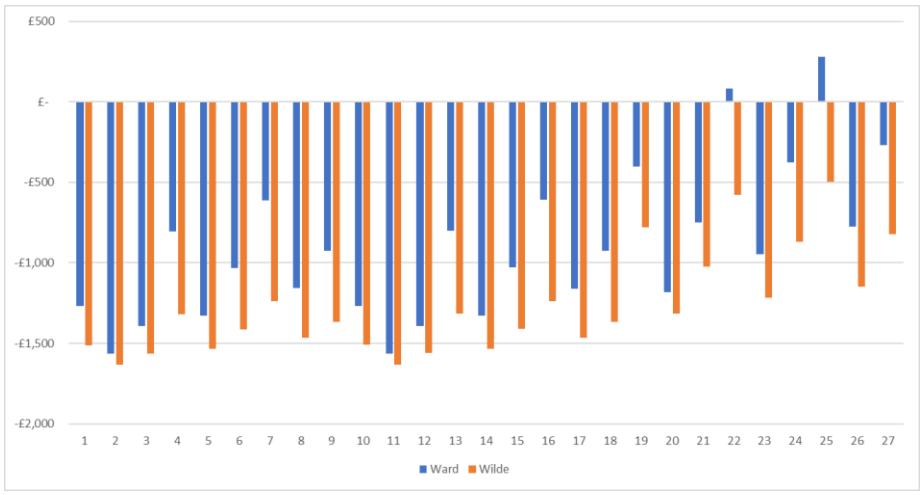

Figure 3: Incremental NMB results for children receiving LFO at an ICER threshold of £20,000

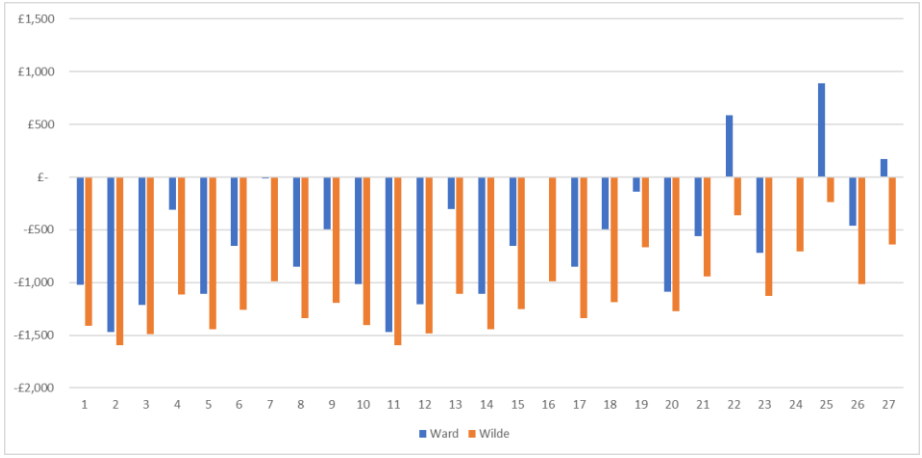

Figure 4: Incremental NMB results for children receiving LFO at an ICER threshold of £30,000

No ICERs were below £30,000 when the probability of death from Wilde *et al*.[37] was used. When the probability of death from Ward *et al*.[36] was used the ICER for remdesivir was below £20,000 in two scenarios (22 and 25) and scenario 27 had an ICER below £30,000. All three scenarios assumed that remdesivir had a beneficial impact of time to discharge; Scenarios 22 and 25 assumed mean efficacies from Amstutz *et al*.[21] and Huang *et al*.[19] respectively, whereas Scenario 27 uses the mean-low estimate from Huang *et al*.[19]

5.3 Exploratory ICERs estimated for immunocompromised patients.

Figure 5 shows the incremental net monetary benefit (NMBs) values for remdesivir when compared to SoC for treating patients requiring LFO at an ICER threshold of £20,000, whilst Figure 6 presents these values at the £30,000 threshold. The ICERs are reported in Appendix 7.

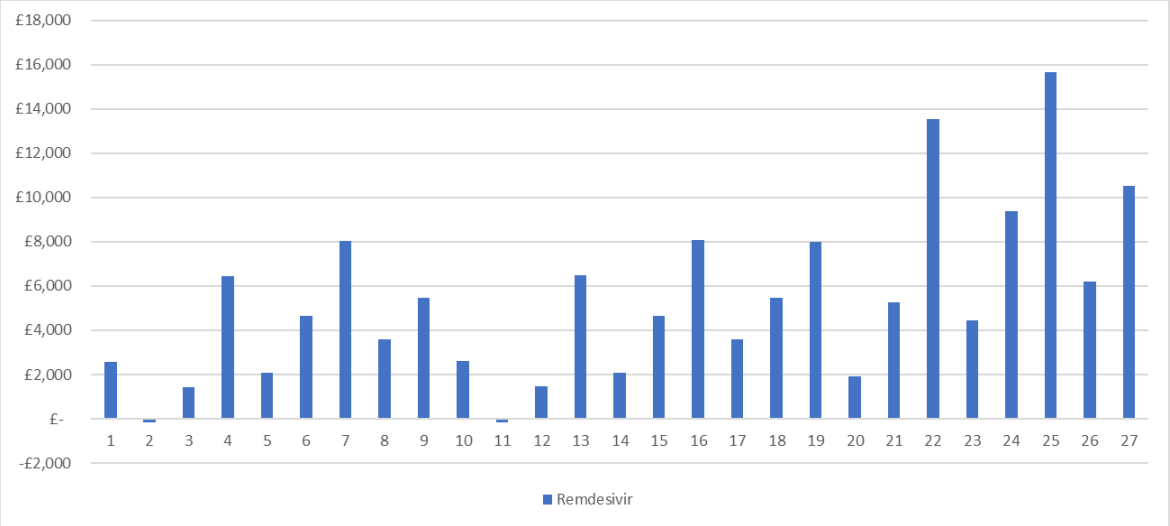

Figure 5: Incremental NMB results for immunocompromised patients receiving LFO at an ICER threshold of £20,000

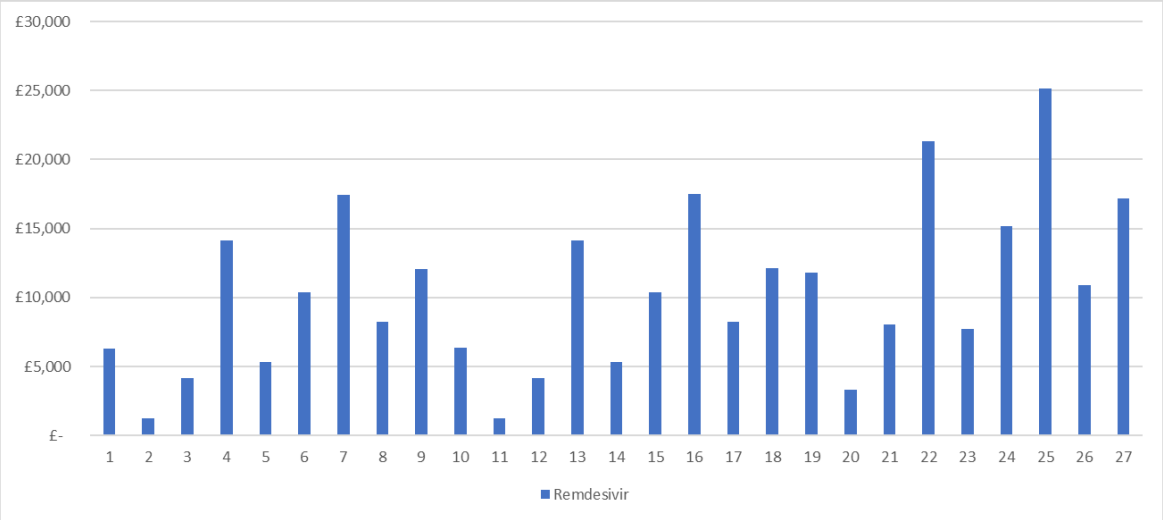

Figure 6: Incremental NMB results for immunocompromised patients receiving LFO at an ICER threshold of £30,000

Remdesivir had an ICER below £30,000 in all scenarios. Two scenarios (2 and 11) had ICERs slightly above £20,000. Both scenarios assumed low efficacy from Amstutz *et al*.[21] with data from SOLIDARITY[28] included for mortality and no impact on time to discharge.

6 Discussion of the results generated by the EAG

The EAG has provided 27 scenarios for each of the three subgroups which produce a wide range in the ICER. Considering remdesivir compared with SoC only the ICER in adult patients requiring LFO ranged from dominating to £33,130; for children requiring LFO the ICERs ranged from £15,413 to £183,524; and for adult patients requiring LFO who are immunocompromised the ICERs ranged from dominating to £21,225. Results for children and immunocompromised adult patients have additional uncertainty due to the necessity of assuming that the treatment effects associated with adult patients requiring LFO are generalisable to these groups.

Key drivers in the ICERs are: which study should provide the estimate of mortality benefit associated with remdesivir; whether the mean estimate of effect should be used or a lower estimate; and whether any benefit in time to discharge should be assumed.

The EAG believes that the Amstutz *et al*.[21] paper provides the best estimate as it included the broadest set of studies and used individual patient data. There is uncertainty over whether data from SOLIDARITY[28] should be included as this also included patients not requiring supplemental oxygen, although the EAG notes that the SOLIDARITY data were used in the NICE rapid guideline.[7] If Amstutz *et al*.[21] is used for the source of mortality benefit then this generates ICERs that are less favourable to remdesivir, being most unfavourable when data from SOLIDARITY[28] are included.

The level of reduction in benefit associated with remdesivir due to changes in the SARS-CoV-2 variant in circulation; the vaccination status of patients; the prior infection status of patients; and improvements in SoC across time is uncertain and has been left for Appraisal Committee discussion. The EAG has aimed to provide sufficient data points such that the Committee has a good idea of the ICER associated with its preferred decision. When less favourable assumptions are made the ICERs increase.

Similarly, whether or not remdesivir provides a benefit in time to discharge has been left for Appraisal Committee discussion. The EAG notes that the final draft guidance for ID4038 states that the Committee concluded that it was reasonable to remove these treatment benefits (Section 3.2.3).[4] Assuming that remdesivir does not improve time to discharge increases the ICERs.

If tocilizumab were considered a comparator, then the comparison of remdesivir and tocilizumab is complex as the intervention with the highest NMB varies depending on the scenario chosen and it is plausible that the Appraisal Committee prefer different scenarios for each intervention. As stated, the tocilizumab results do not include the confidential PAS.

7 References

1. Metry A, Pandor A, Ren S, Shippam A, Clowes M, Dark P, *et al.* Therapeutics for people with COVID-19. An economic evaluation. University of Sheffield; 2022.
2. Metry A, Stevenson M. Therapeutics for people with COVID-19. An economic evaluation. University of Sheffield; 2022.
3. Metry A, Pandor A, Ren S, Stevenson M. Therapeutics for people with COVID-19. An economic evaluation: EAG additional analysis post NICE Appraisal Consultation Document. In; 2023.
4. National Institute for Health & Care Excellence (NICE). Therapeutics for people with COVID-19. *Final draft guidance* 2023.
5. Nice. Molnupiravir, remdesivir and tixagevimab plus cilgavimab for treating COVID-19 [ID6261] - final judgement by the appeal panel; 2023.
6. Sciences G. ID6261 COVID-19 Gilead targeted submission post appeal. In; 2023.
7. NICE. COVID-19 rapid guideline: Managing COVID-19; 2023.
8. EMC. Veklury 100 mg powder for concentrate for solution for infusion. 2023. https://www.medicines.org.uk/emc/product/11597/smpc#gref (Accessed
9. Gilead. Response to NICE request. In; 2023.
10. NIH. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines - July 2023; 2023.
11. Bartoletti M, Azap O, Barac A, Bussini L, Ergonul O, Krause R, *et al.* European society of clinical microbiology and infectious diseases guidelines for coronavirus disease 2019: an update on treatment of patients with mild/moderate disease. *Clin Microbiol Infect* 2022;28:1578-90.
12. Bartoletti M, Azap O, Barac A, Bussini L, Ergonul O, Krause R, *et al.* ESCMID COVID-19 living guidelines: drug treatment and clinical management. *Clin Microbiol Infect* 2022;28:222-38.
13. World Health Organisation (WHO. Clinical management of COVID-19: Living guideline, 18 August 2023. In; 2023.
14. Tacconelli E. Systematic reviews: CRD's guidance for undertaking reviews in health care. *The Lancet Infectious Diseases* 2010;10:226.
15. Amir-Behghadami M, Janati A. Population, Intervention, Comparison, Outcomes and Study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. *Emergency Medicine Journal* 2020.
16. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Res Synth Methods* 2020;11:181-217.
17. Haddaway NR, Collins AM, Coughlin D, Kirk S. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLoS One* 2015;10:e0138237.
18. Bramer WM, Giustini D, Kramer BM, Anderson P. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. *Syst Rev* 2013;2:115.
19. Huang C, Lu T-L, Lin L. Remdesivir Treatment Lacks the Effect on Mortality Reduction in Hospitalized Adult COVID-19 Patients Who Required High-Flow Supplemental Oxygen or Invasive Mechanical Ventilation. *Medicina* 2023;59:1027.
20. Beckerman R, Gori A, Jeyakumar S, Malin JJ, Paredes R, Póvoa P, *et al.* Remdesivir for the treatment of patients hospitalized with COVID-19 receiving supplemental oxygen: a targeted literature review and meta-analysis. *Scientific Reports* 2022;12:9622.
21. Amstutz A, Speich B, Mentré F, Rueegg CS, Belhadi D, Assoumou L, *et al.* Effects of remdesivir in patients hospitalised with COVID-19: a systematic review and individual patient data meta-analysis of randomised controlled trials. *The Lancet Respiratory Medicine* 2023.
22. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, *et al.* AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.

23. Ali K, Azher T, Baqi M, Binnie A, Borgia S, Carrier FM, *et al.* Remdesivir for the treatment of patients in hospital with COVID-19 in Canada: a randomized controlled trial. *Cmaj* 2022;194:E242-E51.
24. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, *et al.* Remdesivir for the Treatment of Covid-19 — Final Report. *New England Journal of Medicine* 2020;383:1813-26.
25. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, *et al.* Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet* 2020;395:1569-78.
26. Spinner CD, Gottlieb RL, Criner GJ, Arribas López JR, Cattelan AM, Soriano Viladomiu A, *et al.* Effect of Remdesivir vs Standard Care on Clinical Status at 11 Days in Patients With Moderate COVID-19: A Randomized Clinical Trial. *JAMA* 2020;324:1048-57.
27. Kalil AC, Patterson TF, Mehta AK, Tomashek KM, Wolfe CR, Ghazaryan V, *et al.* Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19. *N Engl J Med* 2021;384:795-807.
28. Consortium WHOST. Remdesivir and three other drugs for hospitalised patients with COVID-19: final results of the WHO Solidarity randomised trial and updated meta-analyses. *The Lancet* 2022;399:1941-53.
29. Ader F, Bouscambert-Duchamp M, Hites M, Peiffer-Smadja N, Mentré F, Burdet C, *et al.* Final results of the DisCoVeRy trial of remdesivir for patients admitted to hospital with COVID-19. *The Lancet Infectious Diseases* 2022;22:764-5.
30. Ader F, Bouscambert-Duchamp M, Hites M, Peiffer-Smadja N, Poissy J, Belhadi D, *et al.* Remdesivir plus standard of care versus standard of care alone for the treatment of patients admitted to hospital with COVID-19 (DisCoVeRy): a phase 3, randomised, controlled, open-label trial. *The Lancet Infectious Diseases* 2022;22:209-21.
31. Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical trials (London, England)* 2005;2:209-17.
32. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, *et al.* Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. *PLoS Med* 2015;12:e1001855.
33. Garibaldi BT, Wang K, Robinson ML, Betz J, Caleb Alexander G, Andersen KM, *et al.* Real-world effectiveness of remdesivir in adults hospitalized with coronavirus disease 2019 (COVID-19): a retrospective, multicenter comparative effectiveness study. *Clinical Infectious Diseases* 2022;75:e516-e24.
34. National Institute for Health & Care Excellence (NICE). NICE real-world evidence framework. In; 2022.
35. Initiative CN. Living COVID-19 NMA. 2023. https://covid-nma.com/metacovid/ (Accessed
36. Ward JL, Harwood R, Kenny S, Cruz J, Clark M, Davis PJ, *et al.* Pediatric Hospitalizations and ICU Admissions Due to COVID-19 and Pediatric Inflammatory Multisystem Syndrome Temporally Associated With SARS-CoV-2 in England. *JAMA Pediatr* 2023;177:947-55.
37. Wilde H, Tomlinson C, Mateen BA, Selby D, Kanthimathinathan HK, Ramnarayan P, *et al.* Hospital admissions linked to SARS-CoV-2 infection in children and adolescents: cohort study of 3.2 million first ascertained infections in England. *Bmj* 2023;382:e073639.
38. Recovery Collaborative Group, Horby PW, Mafham M, Peto L, Campbell M, Pessoa-Amorim G, *et al.* Casirivimab and imdevimab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *The Lancet* 2022;399:665-76.
39. Evans RA, Dube S, Lu Y, Yates M, Arnetorp S, Barnes E, *et al.* Impact of COVID-19 on immunocompromised populations during the Omicron era: insights from the observational population-based INFORM study. *The Lancet Regional Health - Europe* 2023; https://doi.org/10.1016/j.lanepe.2023.100747:100747.

8 Appendices

8.1 *Appendix 1: Critique of broader search strategy to identify clinical evidence*

Searches for clinical effectiveness evidence were conducted in two phases, in January and December 2022 respectively. Databases included all the core sources required by NICE (MEDLINE; Embase; Cochrane Library) plus clinical trial registries and the proceedings of relevant conferences. The search strategies from both iterations of the search are well-designed, incorporating subject headings and free text terms for the population, intervention and comparators of interest. The strategies used differ slightly between the two iterations, with the December searches including some additional terms relating to interventions for outpatients (outside the scope of this review) and non-RCT evidence.

A search filter was used to identify RCT evidence, with the addition of terms to identify other eligible study types including real world and observational studies. The added terms are appropriate, however it is unclear whether any formal validation of this filter has ever taken place to measure its accuracy in the retrieval of these types of study. No search terms relating to systematic reviews and network meta-analyses were included, though these were eligible for inclusion at the title/abstract stage as a source of relevant studies (but subsequently excluded unless they contained primary data). If relevant reviews were intended to be retrieved as a means of identifying primary studies, it might have been prudent to search for them.

Searches were limited to evidence from 2019 onwards, which is appropriate given the disease area (first cases of COVID-19 were reported in late 2019). The ERG considers the clinical searches unlikely to have missed relevant primary studies eligible for inclusion.

8.2 *Appendix 2: Summary of the company's AMSTAR-2 ratings for included systematic reviews*

Table 5 : Summary of the company's AMSTAR-2 ratings for included systematic reviews (reproduced and adapted for presentation, Company Bias assessment using NICE preferred tools – extraction grid)

	Huang *et al.*[19]	**Beckerman *et al.*[20]**	**Amstutz *et al.*[21]**
Q1 Did the research questions and inclusion criteria for the review include the components of PICO?	**Y**	**Y**	**Y**
Q1 Notes	RCTs were eligible for inclusion if they directly compared the clinical effectiveness of remdesivir to a placebo in the treatment of hospitalized adult COVID-19 patients. Studies that had any one or more of the following outcomes were included: hospital mortality or 28-day mortality, and ordinal scale of the patients at the start of treatment.	The SLR included the population (patients hospitalized with COVID-19 requiring supplemental oxygen at baseline); the intervention (at least one arm of the trial must have been treated with remdesivir); the comparator (any); and the outcomes (mortality; recovery [defined as recovery from COVID-19 or discharge from hospital]; no longer requiring supplemental oxygen; progressing to non-invasive ventilation or mechanical ventilation).	Eligible studies were RCTs (unpublished or published, any format, in any language) that randomly assigned adult patients (aged ≥16 years) who were treated in hospital for COVID-19 to receive either remdesivir or no remdesivir (i.e., usual care as defined by the local context, with or without placebo). The primary outcome was mortality at 28 days
Q2 Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	**N**	**N**	**Y**
Q2 Notes	The authors did not include an explicit statement to establish that the review methods were determined prior to the initiation of review.	There is no explicit statement that review methods were established prior to the conduct of the review.	The study protocol is available on PROSPERO (CRD42021257134), Open Science Framework (https://osf.io/7a4wf), and in the appendix. It states the review question, search strategy, inclusion criteria and risk of bias assessment. Also, under the data-analysis section a synthesis plan

			is reported. To assess heterogeneity in interaction estimates across trials, forest plots were used.
Q3 Did the review authors explain their selection of the study designs for inclusion in the review?	**N**	**N**	**Y**
Q3 Notes	The authors did not explain their choice to only include RCTs.	Authors did not explain limiting their inclusion to only randomised controlled trials.	There are conflicting results in RCTs on patients treated with remdesivir in hospital for COVID-19, and so the focus on RCTs is justified
Q4 Did the review authors use a comprehensive literature search strategy?	**Partial Y**	**Partial Y**	**Y**
Q4 Notes	The authors detail that the search strategy included PubMed, Web of Science, and Cochrane Library databases searched from 1 January 2020 and 28 February 2023. The following search terms were used: “Remdesivir”, “Veklury”, “GS-5734”, “COVID-19”, “coronavirus" and “SARS-CoV-2.” The authors did not provide details on any additional searching (i.e., grey literature, trial registries, reference lists of included studies), nor on any consultation with experts in the field.	The authors searched at least two databases (MEDLINE (PubMed), medRxiv, EMBASE and Cochrane Trials), provided the search strategy (see supplement), and did not apply publication restrictions according to the publication (e.g., language).	Multiple databases searched (PubMed, Embase, the International Clinical Trials Registry Platform [ICTRP] from WHO, and medRxiv), search strategies given in appendix. No publication restrictions (unpublished and non-English studies included). To ensure literature saturation, reference lists of relevant reviews and original articles identified through the search were scanned. Finally, results with trials identified by other published or registered systematic searches as well as personal knowledge were included. The protocol was discussed with two patient representatives from Switzerland and two practising infectious disease specialists. Search conducted in 2022 and completed in 2023

Q5 Did the review authors perform study selection in duplicate?	**N**	**N**	**Y**
Q5 Notes	The publication did not specify that a dual-review approach for selection was used.	The number of individuals performing study selection was not reported.	Each title and abstract were assessed for potential eligibility by two independent reviewers. Each full text included was obtained and independently assessed by two further reviewers.
Q6 Did the review authors perform data extraction in duplicate?	**N**	**N**	**Y**
Q6 Notes	The publication did not specify that a dual approach was used for extraction.	Study selection was reported to be completed by one individual.	Two review authors independently extracted data on patient characteristics, randomization methods, interventions and outcomes by using a standardized pre-piloted data extraction form
Q7 Did the review authors provide a list of excluded studies and justify the exclusions?	**N**	**N**	**N**
Q7 Notes	The authors did not provide a list of excluded studies. Figure 1 shows a flow diagram of the study selection process and at which stage articles were excluded but does not provide a reasoning for the exclusion.	A list of potentially relevant studies which were excluded at full-text reviewer was not provided.	A list of excluded studies has not been provided. However, characteristics of randomized trials that could not be included in the individual patient data meta-analysis was provided, with a reason for its exclusion from the analysis.
Q8 Did the review authors describe the included studies in adequate detail?	**Y**	**Partial Y**	**Partial Y**
Q8 Notes	Table 2 provides a comprehensive list of characteristics of the included studies including the author, region, study period, number of patients, mean age of patients, other treatments for patients receiving remdesivir, and the median time of symptoms before first dose of	The study did describe the population, intervention, comparators, outcomes, and research designs sufficiently; although the standard of care arm is not well-defined, the authors acknowledge this is due to poor reporting in those studies.	In Table 2 the population and intervention have been described in detail, study setting, and time frame also given. However, the comparator was just reported as usual care and has not been reported in detail.

	remdesivir. Although study design was not explicitly stated in the chart, all included studies were RCTs.		
Q9 Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	**Y**	**Y (includes only RCTs)**	**Y**
Q9 Notes	The risk of bias for each trial was assessed using the Cochrane Risk of Bias Tool 2.0 for RCTs, which assesses unconcealed allocation, lack of blinding of patients and assessors, randomness of allocation sequence and selection of reported results.	This SLR utilized the RoB 2 checklist, which includes assessing if the allocation request was truly random (see bias arising from the randomisation process domain), if there was selection of the reported result from among multiple measurements or analyses of a specified outcome (see bias in selection of the reported result domain), if there was risk of bias from unconcealed allocation (see bias arising from the randomisation process domain), and if there was lack of blinding of patients and assessors when assessing outcomes (see bias in measurement of the outcome domain).	Bias was assessed using the Cochrane Risk of Bias 2 tool
Q10 Did the review authors report on the sources of funding for the studies included in the review?	**N**	**N**	**N**
Q10 Notes	The authors did not report on the sources of funding for the studies included in the review.	Source of funding for studies included in the review were not reported.	Funding of included studies not reported
Q11 If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	**Y**	**N**	**Y**
Q11 Notes	Statistical analysis was completed using RevMan 5, the Cochrane Review Manager tool. For continuous and categorical variables,	Methods for adjusting for heterogeneity within the meta-analysis are not reported by the study.	Justification for IPD meta-analysis reported in protocol and a full break down of techniques reported in the data analysis

	the relative risk (RR) and mean difference with a 95% confidence interval (CI) were calculated, respectively. Significant heterogeneity I2 between the studies was defined as an greater than 50% and a p value for the Q-test less than 0.10 for each study. When effects were thought to be homogenous, the fixed effects model was applied, and when they were heterogeneous, the random-effects model was applied.		section of the protocol, with a mixed effects logistic regression model used for the primary outcome
Q12 If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	**N**	**N**	**Y**
Q12 Notes	In Section 4.5 (Limitations) the authors note that two of the studies included in the meta-analysis had a high risk of bias. However, the authors did not investigate the possible impact of this bias on the results.	Some studies with high risk of bias were included in the meta-analysis, but there were no reported analyses investigating the impact of RoB on summary estimates of effect.	A scenario analysis was conducted of the meta-analysis to only include trials that were judged to have a low risk of bias for all outcomes.
Q13 Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	**N**	**Y**	**Y**
Q13 Notes	Although the authors note that two of the included studies have a high risk of bias, they do not discuss the impact of this on the results other than acknowledging it as a weakness of the review.	In the discussion, authors did briefly discuss the impact of RoB on the interpretation of the results.	A scenario analysis was conducted of the meta-analysis to only include trials that were judged to have a low risk of bias for all outcomes.
Q14 Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	**Y**	**Y**	**Y**
Q14 Notes	Th authors note that the included studies were heterogeneous due to	The authors report heterogeneity in the results for patients receiving high-flow oxygen and	Reported that forest plots would be used to assess heterogeneity, and then reported that they did

	different counties, populations, and study designs.	explain this difference may indicate that patients receiving low-flow oxygen benefit more greatly from remdesivir, or that this may be due to the smaller sample size of high-flow oxygen patients or the confounding effect of including patients on NIV in the high-flow oxygen population.	not find credible evidence for effect modification by age, presence of comorbidities, enrolment period, or corticosteroid use
Q15 If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	**N**	**N**	**N**
Q15 Notes	The authors did not detail the potential for publication bias.	Publication bias was not discussed, and its effect was not evaluated.	Although bias of studies was looked at using the RoB2 and sensitivity analysis that just included studies with low risk of bias, there was no discussion on the likelihood/magnitude of impact of publication bias
Q16 Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	**Y**	**Y**	**Y**
Q16 Notes	The authors declared no conflicts of interest.	Conflicts of interest were reported.	The authors reported no competing interests
Overall confidence level in review results	**Medium**	**Low**	**Medium**
Justification	The systematic review has more than one weakness but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review.	The majority of the questions' responses are no, with some partial yes responses, and few yes responses.	Overall good reporting with full appendices provided making the case for thorough strategies at each step. However, the study did not provide a list of excluded studies nor were the likelihood/magnitude of impact of publication bias discussed.

8.3 *Appendix 3: Data from Amstutz et al.*[21]

Table 6 : Data from Amstutz *et al* .[21] Sensitivity analyses on different subgroup definitions on the primary outcome of mortality at day 28 (adapted from Figure S8)

Subgroup	**Outcome variable: Mortality at day 28**				
	Total number (N)	**Remdesivir (n/N)**	**No remdesivir (n/N)**	**Adjusted odds ratio (95% CI)**	**Interaction p-value**
No oxygen, no ventilation at baseline	2357	34/1274	33/1083	0.86 (0.532, 1.394)	-
Oxygen, but no ventilation at baseline	6274	374/3198	432/3076	0.79 (0.68, 0.919)	0.505
High-flow oxygen or non-invasive ventilation at baseline	741	90/372	91/369	1.04 (0.712, 1.519)	0.764
Mechanical ventilation/ECMO at baseline	949	163/472	150/477	1.15 (0.862, 1.522)	0.439
Without additional WHO- SOLIDARITY[28] data*: No oxygen, no ventilation at baseline	857	14/525	12/332	0.77 (0.338, 1.74)	-
Without additional WHO- SOLIDARITY[28] data*: Oxygen, but no ventilation at baseline	2106	75/1094	114/1012	0.59 (0.431, 0.817)	0.514
Without additional WHO- SOLIDARITY[28] data*: High-flow oxygen or non-invasive ventilation at baseline	741	90/372	91/369	1.04 (0.712, 1.519)	0.539
Without additional WHO-SOLIDARITY[28] data*: Mechanical ventilation/ECMO at baseline	509	54/246	55/263	1.07 (0.695, 1.646)	0.523

ECMO, Extracorporeal membrane oxygenation *These subgroup analyses included data from CATCO, DisCoVeRy, NOR- SOLIDARITY, and FIN- SOLIDARITY, but excluded the additional WHO- SOLIDARITY trial data (n=6167)

8.4 Appendix 4: Summary of the company's NICE Real World Evidence ratings on the methods used by Garibaldi et al.[33]

Table 7 : Summary of the company's NICE Real World Evidence ratings on the methods used by Garibaldi et al.[33] to minimise the risk of bias (reproduced and adapted for presentation, Company Bias assessment using NICE preferred tools – extraction grid)

Study name	Q1 Selection bias at study entry	Q2 Selection bias at study exit	Q3 Addressing confounding	Q4 Detection bias	Q5 Measurement error and misclassification	Q6 Missing data	Q7 Reverse causation
Garibaldi *et al.* 2022[33]	Approximately half of the remdesivir patients were not able to be matched and were therefore excluded from the analysis, potentially introducing bias by selecting a smaller patient population. Symptom onset was not available in the dataset, so we were not able to examine whether or not the benefit of remdesivir differed based on timing of treatment. Because antiviral therapies are likely most effective early in the disease course, differential timing of treatment could bias outcomes toward specific groups.	The primary outcome was time to clinical improvement from the first day of remdesivir treatment or the matched day. Failure of clinical improvement was censored at the last day of follow-up or 28-days, whichever came first. The secondary outcome was time to death from the first day of remdesivir treatment or the matched day Patients who were discharged alive to "home" or "self-care" were censored at 28	The following factors were included in the regression models developed to address confounding: demographics, oxygen delivery device, vital signs, key laboratory data, comorbidities (including the Charleson comorbidity index) and COVID-19-specific medications (e.g., dexamethasone and tocilizumab). The standardized difference between matched cases and controls is presented in the table. The study also uses time-dependent propensity score matching, to create pairs of individuals, one treated with remdesivir and the	HCA Healthcare comprises over 2000 care sites including more than 180 acute-care facilities, and therefore there is substantial risk of bias being introduced as a result of variable clinical practice across hospitals and health systems, particularly for the primary outcome of time to clinical improvement. The authors do not specify how this potential for detection bias was mitigated.	Although the primary outcome of clinical improvement was defined as a 2-point decrease in the 8-point WHO severity score or discharge within 28 days, this assessment was at the discretion of the study physician, which may introduce bias. The secondary outcome was time to death from the first day of remdesivir treatment or the matched day, which is unlikely to be captured incorrectly.	For the laboratory results, missing values were imputed using the last observation carried forward if the last observation was within three days of the missing data, otherwise, using multiple imputation by chained equations (MICE) with a predictive mean matching method. Variables with more than 50% missingness were not included in the models. These precautions pre-	There is unlikely to be a risk from reverse causation

	To account for the variable timing of administration, time-dependent PS matching was utilized to create pairs of individuals, one patient treated with remdesivir and the other the most similar patient eligible for treatment at the time of remdesivir initiation but who did not receive remdesivir. In order to account for changes in the pandemic over time, an individual that received remdesivir prior to 1 October 2020 would be matched to a control patient hospitalized before 1 October 2020. To further mitigate time-related bias, a sensitivity analyses excluding patients hospitalized before 1 July 2020 was conducted, as the early months of the pandemic presented unique challenges to health systems that may have effected results.	days. Patients who were discharged to another healthcare facility without a known death date were censored at last follow-up. There is a low possibility of informative censoring in this study, as time-dependent propensity score matching would eliminate unequal dropouts between the cases and controls.	other the most similar patient eligible for treatment at the time of remdesivir initiation but who did not receive treatment. Notably, the study was conducted prior to the widespread use of vaccines and the emergence of variants such as Delta and Omicron, and therefore their potential for confounding was not investigated.			emptively could mitigate the risk of bias from missing data.	

	A patient who received a certain number of days of remdesivir treatment was matched to a control patient who stayed in the hospital at least that length of time (up to a maximum of five days) beyond the matching day. This time constraint on the matching prevents matching remdesivir patients to individuals would not have been considered candidates for remdesivir treatment as they were healthy enough to be discharged.						

8.5 Appendix 5: The assumed efficacy values for tocilizumab

The efficacy values used in the EAG analyses for are provided in Table 8. The HRs for preventing mortality and time to discharge and the RR for clinical improvement were those used in ID4038 which was sourced from COVID-NMA.[35] As with remdesivir, as the HRs used for the risk of mortality for tocilizumab are all below 1, no capping of parameter values at 1 was applied, as the EAG believes it plausible that other aspects such as time to discharge and clinical improvement could be worse as a by-product of preventing death.

For simplicity, the assumption for remdesivir that there was no clinical improvement when an impact on time to discharge was assumed, was also applied to tocilizumab. This is marginally unfavourable to tocilizumab which has a slight beneficial effect on clinical improvement.

Table 8: Parameter values used in the EAG's analyses for tocilizumab

Scenario number	Efficacy scenario	Tocilizumab parameters*
1, 4, 7	Mean	0.763, unity, unity
2, 5, 8	Low	0.900, unity, unity
3, 6, 9	Mean-Low	0.831, unity, unity
10, 13, 16	Mean	0.763, 1.050, unity
11, 14, 17	Low	0.900, 1.000, unity
12, 15, 18	Mean-Low	0.831, 1.025, unity
19, 22, 25	Mean	0.763, unity, 1.050
20, 23, 26	Low	0.900, unity, 0.880
21, 24, 27	Mean-Low	0.831, unity, 0.967

*Parameter values are: hazard ratio for time to death; relative risk for clinical improvement; hazard ratio for time to discharge

8.6 Appendix 6 ICERs generated by the EAG analyses for adults requiring LFO

Table 9: ICERs generated by the EAG analyses for adults requiring LFO

Scenario number	Remdesivir compared with SoC	Tocilizumab compared with SoC	Remdesivir compared with tocilizumab
1	£19,086	£13,605	Dominated
2	£33,001	£17,800	Dominated
3	£22,146	£14,856	Dominated
4	£14,771	£13,605	£16,847
5	£20,270	£17,800	£24,228
6	£16,169	£14,856	£18,154
7	£14,013	£13,605	£14,467
8	£17,425	£17,800	£17,138
9	£15,427	£14,856	£16,051
10	£18,877	£13,399	Dominated
11	£33,130	£17,800	Dominated
12	£22,042	£14,715	Dominated
13	£14,657	£13,399	£16,897
14	£20,328	£17,800	£24,380
15	£16,115	£14,715	£18,230
16	£13,916	£13,399	£14,490
17	£17,468	£17,800	£17,213
18	£15,379	£14,715	£16,104
19	Dominant	£7,895	£90,372[†]
20	Dominant	£52,896	£222,607[†]
21	Dominant	£20,069	£170,670[†]
22	Dominant	£7,895	Dominant
23	£5,046	£52,896	Dominant
24	£605	£20,069	Dominant
25	£560	£7,895	Dominant
26	£6,670	£52,896	Dominant
27	£1,907	£20,069	Dominant

SoC – Standard of care

[†]Located in Southwest quadrant of cost-effectiveness plane (i.e., remdesivir is cheaper and less efficacious than tocilizumab)

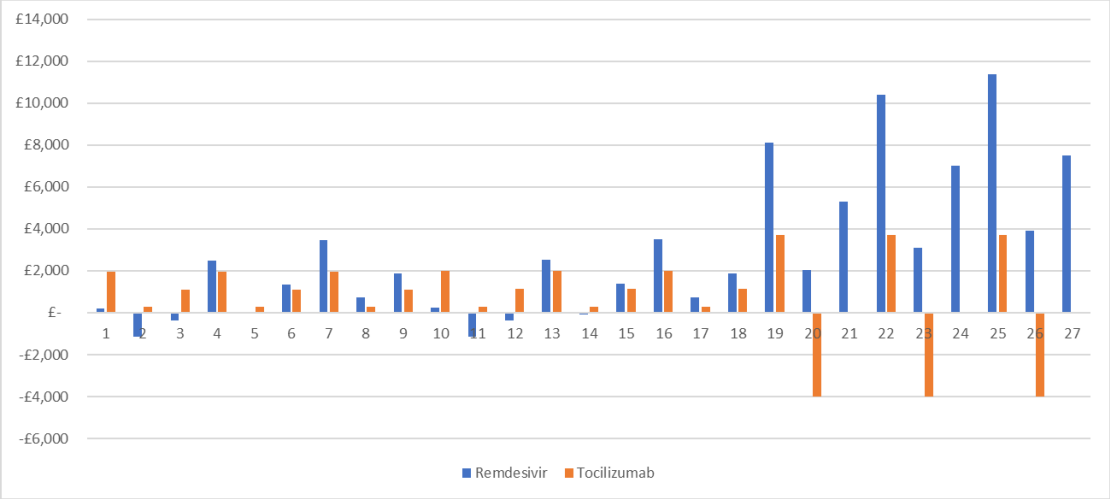

Figure 7: Incremental NMB results for adults receiving LFO at an ICER threshold of £20,000 when tocilizumab is considered a comparator

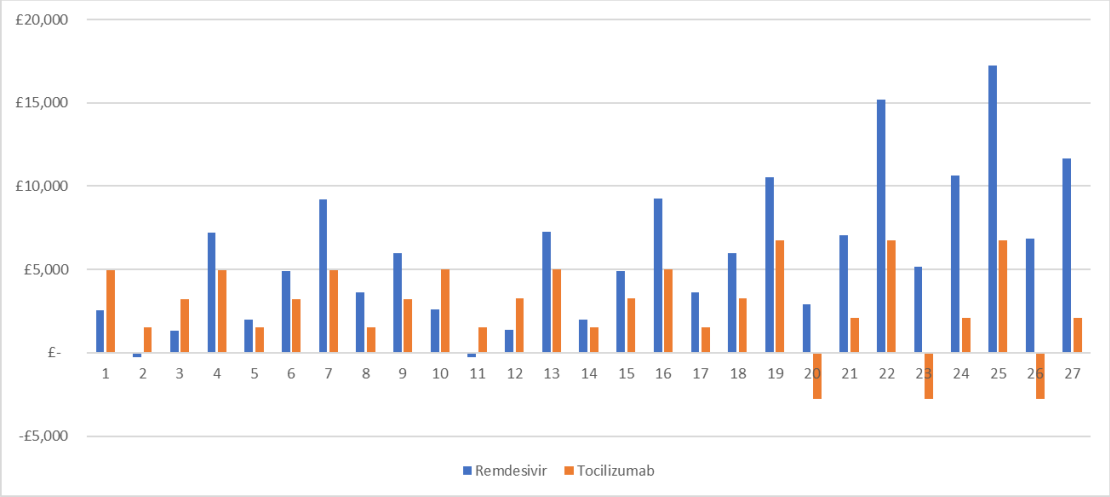

Figure 8: Incremental NMB results for adults receiving LFO at an ICER threshold of £30,000 when tocilizumab is considered a comparator

The comparison of remdesivir and tocilizumab is complex as the intervention with the highest NMB varies depending on the scenario chosen and it is plausible that the Appraisal Committee prefer different scenarios for each intervention. As stated, the tocilizumab results do not include the confidential PAS.

8.7 Appendix 7: ICERs generated by the EAG analyses for immunocompromised adult patients requiring LFO

Table 10: ICERs generated by the EAG analyses for immunocompromised adult patients requiring LFO

Scenario number	Remdesivir compared with SoC	Tocilizumab compared with SoC	Remdesivir compared with tocilizumab
1	£13,036	£9,993	£70†
2	£21,180	£11,770	Dominated
3	£14,670	£10,310	Dominated
4	£11,610	£9,993	£14,397
5	£13,642	£11,770	£16,599
6	£11,878	£10,310	£14,190
7	£11,456	£9,993	£13,017
8	£12,289	£11,770	£12,679
9	£11,735	£10,310	£13,248
10	£12,958	£9,913	Dominated
11	£21,225	£11,770	Dominated
12	£14,633	£10,258	Dominated
13	£11,561	£9,913	£14,402
14	£13,663	£11,770	£16,654
15	£11,857	£10,258	£14,213
16	£11,412	£9,913	£13,011
17	£12,305	£11,770	£12,707
18	£11,715	£10,258	£13,261
19	Dominant	£6,888	£34,658†
20	£6,403	£29,908	£90,599†
21	£901	£13,071	£65,536†
22	£2,529	£6,888	Dominant
23	£6,464	£29,908	Dominant
24	£3,739	£13,071	Dominant
25	£3,487	£6,888	Dominant
26	£6,730	£29,908	Dominant
27	£4,235	£13,071	Dominant

SoC – Standard of care

†Located in Southwest quadrant of cost-effectiveness plane (i.e., remdesivir is cheaper and less efficacious than tocilizumab)

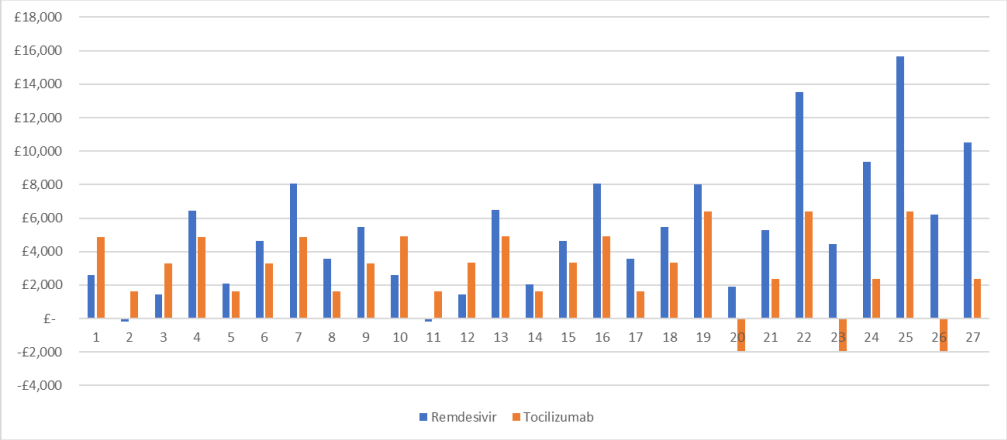

Figure 9: Incremental NMB results for immunocompromised adult patients receiving LFO at an ICER threshold of £20,000 when tocilizumab is considered a comparator

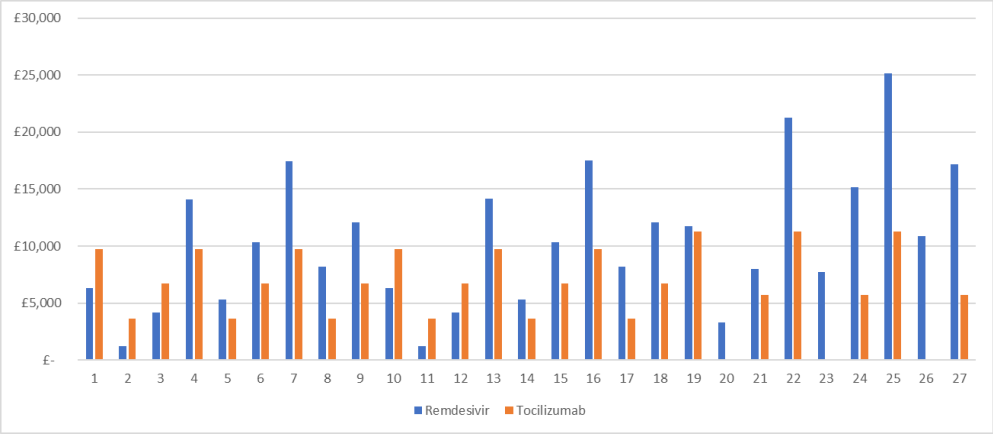

Figure 10: Incremental NMB results for immunocompromised adult patients receiving LFO at an ICER threshold of £30,000 when tocilizumab is considered a comparator

The comparison of remdesivir and tocilizumab is complex as the intervention with the highest NMB varies depending on the scenario chosen and it is plausible that the Appraisal Committee prefer different scenarios for each intervention. As stated, the tocilizumab results do not include the confidential PAS.

8.8 Appendix 8: ICERs generated by the EAG analyses for children requiring LFO

Table 11: ICERs generated by the EAG analyses for children requiring LFO

	Remdesivir compared with SoC	
Scenario number	Probability of death from Ward *et al*.[36] (0.45%)	Probability of death from Wilde *et al*.[37] (0.19%)
1	£70,761	£165,864
2	£183,430	£432,324
3	£95,485	£224,350
4	£36,161	£83,936
5	£80,320	£188,472
6	£47,301	£110,335
7	£30,193	£69,766
8	£57,381	£134,199
9	£41,374	£96,295
10	£70,622	£165,579
11	£183,524	£432,528
12	£95,413	£224,201
13	£36,091	£83,796
14	£80,359	£188,553
15	£47,266	£110,266
16	£30,134	£69,650
17	£57,408	£134,255
18	£41,344	£96,235
19	£35,438	£89,696
20	£138,770	£326,235
21	£59,118	£144,138
22	£18,417	£46,985
23	£62,168	£148,833
24	£29,801	£73,478
25	£15,413	£39,273
26	£44,629	£107,060
27	£26,117	£64,405