NATIONAL COVID-19 CLINICAL EVIDENCE TASKFORCE

# Remdesivir methods brief

When the guidelines were first published in April 2020, the Taskforce only recommended the use of antivirals in the context of randomised trials. In version 4 (8 May 2020), following publication of the first trial of remdesivir for COVID-19 [1], we made a specific recommendation only to use remdesivir in the context of a randomised trial. In version 7 (4 June 2020), following publication of additional data from the same trial, we made a conditional recommendation to consider using remdesivir outside of a trial setting.

Following publication of the SOLIDARITY trial [2], the Taskforce updated its recommendation in version 27 (29 October 2020) to a conditional recommendation supporting the use of remdesivir in hospitalised adults with moderate to severe COVID-19 who do not require ventilation, and a recommendation against the use of remdesivir in adults who do require ventilation (invasive or non-invasive mechanical ventilation or extracorporeal membrane oxygenation (ECMO)).

Shortly thereafter, and based on analyses of the same data, the World Health Organization (WHO) released a conditional recommendation against the use of remdesivir in hospitalised patients, irrespective of disease severity. To address this discrepancy, the Taskforce considered it prudent to review its current recommendations relating to use of remdesivir and the decisions that led to them.

The main reasons for the differences between recommendations made by WHO and the Taskforce relate to different approaches to the analysis of the data and variations in the GRADE 'Evidence to Decision' assessment. In the following sections we discuss the differences between the recommendation made by WHO and our recommendations. We first describe differences in analyses and then differences in the Evidence to Decision assessment.

## Analyses

WHO used the *Instrument for assessing the Credibility of Effect Modification Analyses (ICEMAN)* (see https://www.iceman.help/overview) to determine whether the delineation of subgroups based on disease severity was credible.

Based on this analysis, WHO chose not to separate data by disease severity and instead formulated their recommendation using overall mortality data (see https://www.who.int/publications/i/item/therapeutics-and-covid-19-living-guideline).

The Taskforce also used the ICEMAN tool to assess the credibility of disease severity subgroups used in our current remdesivir recommendations. With additional input from clinicians, virologists and immunologists, the Taskforce considers it plausible that remdesivir has differential effects in patients with COVID-19 depending on severity of illness. As a result, the overall ICEMAN assessment indicated that the delineation of subgroups based on disease severity is credible.

Based on this assessment, the Disease Modifying Treatment and Chemoprophylaxis Panel and Guidelines Leadership Group of the Taskforce concluded that we should continue to provide a conditional recommendation supporting the use of remdesivir in hospitalised adults with moderate to severe COVID-19 who do not require ventilation , and a recommendation against the use of remdesivir in adults who do require ventilation (invasive or non-invasive mechanical ventilation or ECMO). [See page 5 for details of the Taskforce ICEMAN assessment.]

In addition to differences in the assessment of the credibility of recommendations by disease severity subgroups, WHO and the Taskforce use slightly different analysis approaches to the same trials as a basis for the recommendations [1,2]. WHO conducted a network meta-analysis using odds ratios, whereas the Taskforce maintains pairwise comparisons using a random effects model to determine the risk ratio. The differences between the results of these two approaches are however small and would probably not lead to different recommendations on their own.

## Evidence to Decision

### Benefits and harms

WHO determined that there was a lack of evidence demonstrating an improvement in patient-relevant outcomes when using remdesivir, such as mortality, need for mechanical ventilation, and time to clinical improvement. The Taskforce determined that using remdesivir in adults hospitalised with moderate to severe COVID-19 who do not require ventilation probably reduces the risk of death compared with standard care. Conversely, the Taskforce determined that using remdesivir in adults hospitalised with COVID-19 who require ventilation (invasive or non-invasive mechanical ventilation or ECMO) may increase the risk of death compared with standard care.

Although evidence for other patient-relevant outcomes was not reported separately based on disease severity, the Taskforce considers mortality to be the outcome of primary importance when developing recommendations.

### Certainty of the evidence

For important patient-relevant outcomes, WHO downgraded certainty based on risk of bias (due to deviation from intended intervention) and imprecision (due to wide confidence intervals). The Taskforce downgraded certainty based on imprecision only (due to wide confidence intervals), as risk of bias was not considered to be of significant concern with regards to mortality, even though the trials were open label, meaning that it was known which treatment arm a patient was assigned to. The reasons for this are that mortality is considered an objective outcome and the deviations from intended treatment, such as levels of crossover, were small. In the largest of the trials (SOLIDARITY), for example, only 2% of the control group also received remdesivir, the use of other antivirals, corticosteroids and biologics were reported and balanced between groups. It was also not judged plausible that patients would be treated significantly differently just because they were known to receive or not receive remdesivir.

### Preference and values

WHO judged that most patients would be reluctant to use remdesivir due to the uncertainty regarding its effects on mortality and the other prioritised outcomes. This was particularly so as any beneficial effects of remdesivir, if they do exist, are likely to be small and the possibility of important harm remains. The panel acknowledged, however, that values and preferences are likely to vary, and there will be patients and clinicians who choose to use remdesivir, given the evidence has not excluded the possibility of benefit.

The Taskforce believes that as there is an observed mortality benefit in hospitalised patients who do not require ventilation, most patients in this disease severity category would likely opt for treatment with remdesivir. This is backed by the Taskforce Consumer Panel who believes that most informed patients would agree with the recommendation and choose this treatment. Conversely, based on an observed increase in mortality associated with use of remdesivir in hospitalised patients who require ventilation (invasive or non-invasive mechanical ventilation or ECMO), these patients would be reluctant to use remdesivir. The Taskforce Consumer Panel believes that as there is clear evidence demonstrating harms of remdesivir in patients requiring ventilation, informed patients would not choose this treatment.

### Resource implications, feasibility, equity and other considerations

WHO considers that a greater certainty of important benefits is required when recommending novel therapies. They highlight the absence of cost-effectiveness analyses, limited global availability and the requirement of intravenous

administration of remdesivir. In addition, concerns were raised over opportunity costs and the risk of diverting resources away from best supportive care or treatments with demonstrated benefit, such as corticosteroids.

Recommendations developed within the Taskforce are done so within the context of the Australian healthcare system. As a result, although global availability of remdesivir remains an issue, there are fewer concerns regarding opportunity costs in Australia compared with many other healthcare systems that fall within the operational sphere of WHO.

## Conclusions

There are several factors that contribute to differences in recommendations between WHO and the Taskforce:

- The Taskforce considers that delineation of subgroups based on disease severity is credible, and that the observed differences in effect on mortality between subgroups is plausible.
- The Taskforce highlights the differential effect of remdesivir on mortality within these subgroups (i.e. absolute effect estimate of 25 fewer deaths per 1000 in hospitalised patients who do not require ventilation, and absolute effect estimate of 50 more deaths per 1000 in hospitalised patients who require ventilation (invasive or non-invasive mechanical ventilation or ECMO)).
- The Taskforce is more certain that the effect of remdesivir on mortality is closer to the true effect than WHO, as certainty was not downgraded due to risk of bias (certainty for mortality outcomes within the Taskforce is moderate, compared with low certainty within WHO).
- As the Taskforce develops recommendations specific to the Australian healthcare context, there are fewer resource limitations and barriers to actioning these recommendations than in many countries within the operational sphere of WHO.

## References


1. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, et al. ACTT-1 Study Group Members. Remdesivir for the treatment of Covid-19: final report. *N Engl J Med*. 2020 Nov 5;383(19):1813-1826. doi: 10.1056/NEJMoa2007764.
2. WHO Solidarity Trial Consortium, Pan H, Peto R, Henao-Restrepo AM, et al. Repurposed antiviral drugs for Covid-19: interim WHO Solidarity Trial results. *N Engl J Med*. 2020 Dec 2. doi: 10.1056/NEJMoa2023184.

**ICEMAN (Instrument for assessing the Credibility of Effect Modification Analyses) assessment of the illness severity subgroups in remdesivir**

ICEMAN[1] is a new tool developed to assess the credibility of subgroup analyses both in trials and meta-analyses. We are using it to assess whether it is credible that there are illness severity subgroups when it comes to treating COVID-19 patients with remdesivir.

Below you will find some quick instructions on how the tool is used and you will find the assessments and overall rating made by the Taskforce. You can find a paper on ICEMAN here

**Quick Instructions**

- Synonyms for effect modification include subgroup effect, interaction, and moderation
- The instrument applies to a *single proposed effect modification at a time*; complete one form per each outcome, time-point, effect measure, and effect modifier
- Response options on the left indicate definitely or probably reduced, response options on the right probably or definitely increased credibility
- Completely unclear goes under probably reduced credibility
- It is helpful to provide a supporting comment or quotation under each question
- Whether an effect modification is patient-important is *not* part of the credibility assessment
- The manual provides more detailed instructions and examples

**Preliminary considerations**

Study reference(s): **24 Nov 2020 Remdesivir for COVID-19.rm5 (our analysis, see forest plot below)**

If available, protocol reference(s): **Not published**

State a single outcome and, if applicable, time-point of interest (e.g. mortality at 1 year follow-up): **Mortality at day 28**

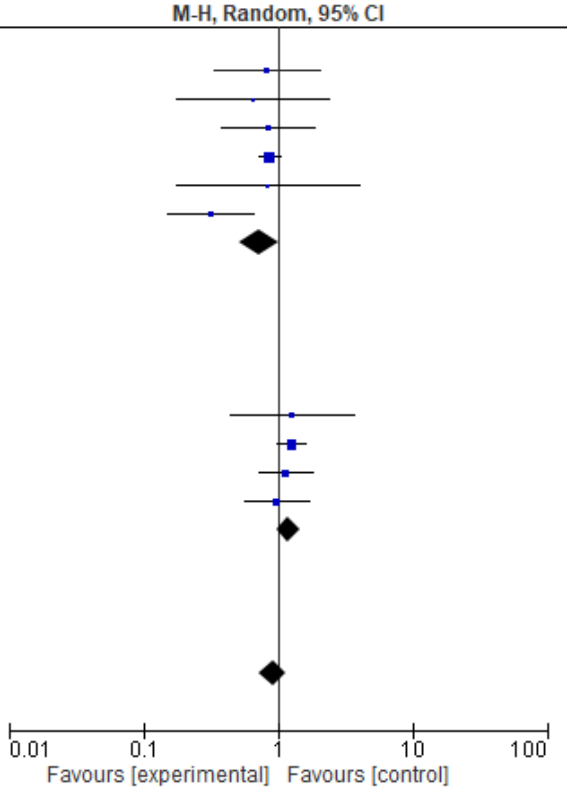
State a single effect measure of interest (e.g. relative or absolute risk difference): **RR, random effects**

State a single potential effect modifier of interest (e.g. age or comorbidity): **Severity of illness measured via baseline need for respiratory support**

Was the potential effect modifier measured before randomization? **[ x ]** yes, continue [ ] no, stop here and refer to manual for further instructions

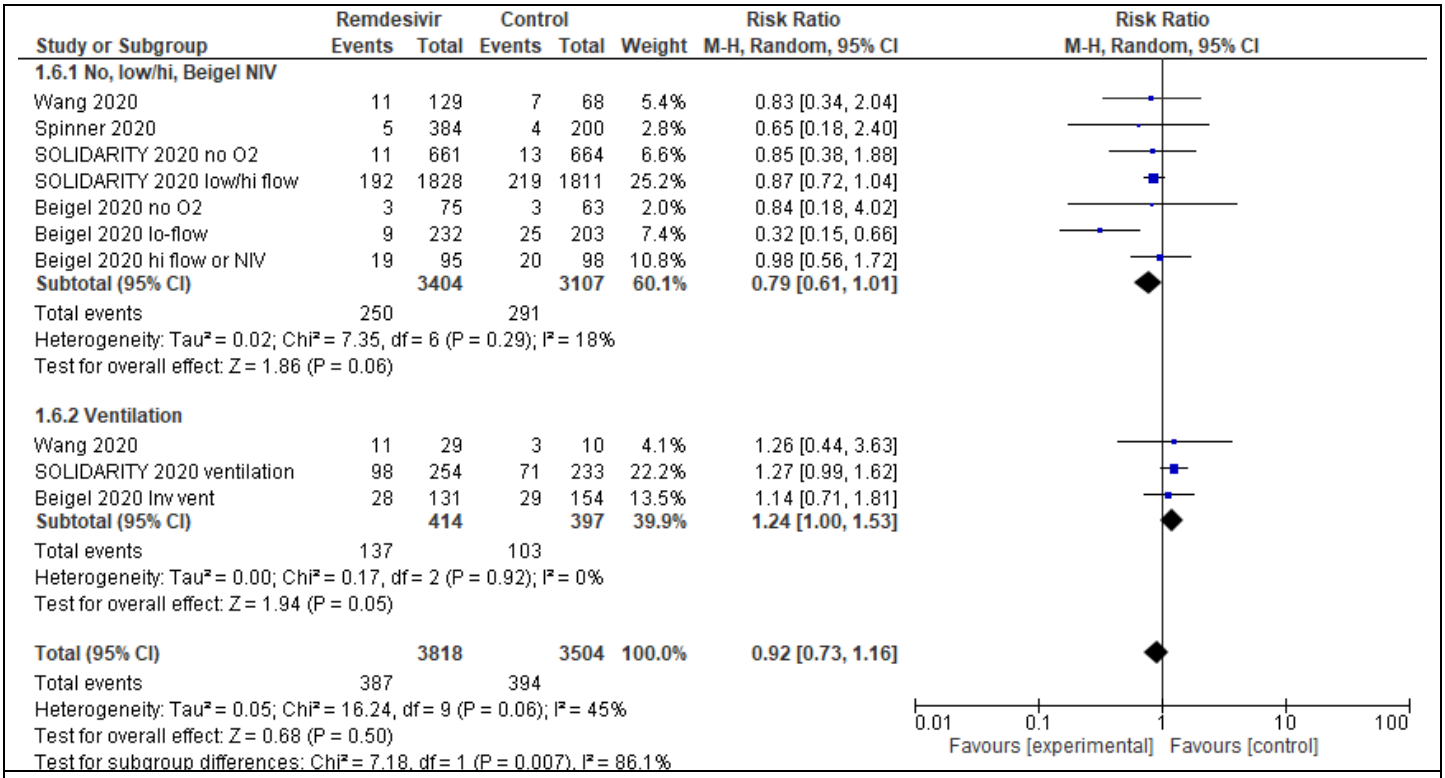
[1] Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses
Stefan Schandelmaier, Matthias Briel, Ravi Varadhan, Christopher H. Schmid, Niveditha Devasenapathy, Rodney A. Hayward, Joel Gagnier, Michael Borenstein, Geert J.M.G. van der Heijden, Issa J. Dahabreh, Xin Sun, Willi Sauerbrei, Michael Walsh, John P.A. Ioannidis, Lehana Thabane, Gordon H. Guyatt
CMAJ Aug 2020, 192 (32) E901-E906; DOI: 10.1503/cmaj.200077

| Study or Subgroup | Remdesivir Events | Total | Control Events | Total | Weight | Risk Ratio M-H, Random, 95% CI |
|---|---|---|---|---|---|---|
| **1.5.1 No, low/hi** | | | | | | |
| Wang 2020 | 11 | 129 | 7 | 68 | 5.4% | 0.83 [0.34, 2.04] |
| Spinner 2020 | 5 | 384 | 4 | 200 | 2.8% | 0.65 [0.18, 2.40] |
| SOLIDARITY 2020 no O2 | 11 | 661 | 13 | 664 | 6.6% | 0.85 [0.38, 1.88] |
| SOLIDARITY 2020 low/hi flow | 192 | 1828 | 219 | 1811 | 25.2% | 0.87 [0.72, 1.04] |
| Beigel 2020 no O2 | 3 | 75 | 3 | 63 | 2.0% | 0.84 [0.18, 4.02] |
| Beigel 2020 lo-flow | 9 | 232 | 25 | 203 | 7.4% | 0.32 [0.15, 0.66] |
| **Subtotal (95% CI)** | | **3309** | | **3009** | **49.3%** | **0.72 [0.52, 1.01]** |
| Total events | 231 | | 271 | | | |
| Heterogeneity: Tau² = 0.05; Chi² = 6.98, df = 5 (P = 0.22); I² = 28% | | | | | | |
| Test for overall effect: Z = 1.88 (P = 0.06) | | | | | | |
| **1.5.2 Ventilation** | | | | | | |
| Wang 2020 | 11 | 29 | 3 | 10 | 4.1% | 1.26 [0.44, 3.63] |
| SOLIDARITY 2020 ventilation | 98 | 254 | 71 | 233 | 22.2% | 1.27 [0.99, 1.62] |
| Beigel 2020 Inv vent | 28 | 131 | 29 | 154 | 13.5% | 1.14 [0.71, 1.81] |
| Beigel 2020 hi flow or NIV | 19 | 95 | 20 | 98 | 10.8% | 0.98 [0.56, 1.72] |
| **Subtotal (95% CI)** | | **509** | | **495** | **50.7%** | **1.20 [0.98, 1.47]** |
| Total events | 156 | | 123 | | | |
| Heterogeneity: Tau² = 0.00; Chi² = 0.75, df = 3 (P = 0.86); I² = 0% | | | | | | |
| Test for overall effect: Z = 1.79 (P = 0.07) | | | | | | |
| **Total (95% CI)** | | **3818** | | **3504** | **100.0%** | **0.92 [0.73, 1.16]** |
| Total events | 387 | | 394 | | | |
| Heterogeneity: Tau² = 0.05; Chi² = 16.24, df = 9 (P = 0.06); I² = 45% | | | | | | |
| Test for overall effect: Z = 0.68 (P = 0.50) | | | | | | |
| Test for subgroup differences: Chi² = 6.42, df = 1 (P = 0.01), I² = 84.4% | | | | | | |



| **Credibility Assessment** | | | |
|---|---|---|---|
| **1: Is the analysis of effect modification based on comparison within rather than between trials?** | | | |
| [ ] Completely between | [ ] Mostly between or unclear | [ x] Mostly within | [ ] Completely within |
| *Subgroup analysis or meta-regression comparing overall effects of each individual trial. This is typical for aggregate data meta- analysis.* | *Subgroup analysis or meta-regression with most information coming from overall effects, but some trials providing within-trial subgroup information* | *Most trials providing within-trial subgroup information; or individual participant data analysis that combines within and between trial information* | *All trials providing within-trial subgroup information or individual participant data; and the analysis separates within from between trial information, e.g. meta-analysis of interactions* |
| Comment: One trial (Spinner 2020) only included patients that did not require mechanical ventilation. Spinner 2020 only weights 3% in the total analysis | | | |
| **2: For within-trial comparisons, is the effect modification similar from trial to trial?** [ ] Not applicable: no or one within-RCT comparison | | | |
| [ ] Definitely not similar | [ ] Probably not similar or unclear | [ ] Mostly similar | [ x] Definitely similar |
| *Effect modification reported for two or more trials and clearly different directions* | *Effect modification not reported for individual trials or too imprecise to tell* | *Effect modification reported for two or more trials, mostly similar in direction, but considerable differences in magnitude* | *Effect modification reported for two or more trials, similar in direction, only some differences in magnitude* |
| Comment: The effects are similar in the subgroups. Beigel 2020 High flow or NIV is interesting because need for respiratory support is an ordinal scale and the way Beigel 2020 reported the data was slightly different from the other trials. Beigel 2020 reported patients receiving high flow and noninvasive ventilation together. The other trials separated those. This means that this particular group of patients sits in the middle regarding need for respiratory support and the effect of remdesivir in this group also sits in the middle. Under item 8 of this tool, we show a sensitivity analysis demonstrating the subgroup effects remain regardless of how this particular group of patients is classified. | | | |
| **3: For between-trial comparisons, is the number of trials large?** [x ] Not applicable: no between RCT comparison | | | |
| [ ] Very small | [ ] Rather small or unclear | [ ] Rather large | [ ] Large |
| *1 or 2 or in smallest subgroup; 5 or less in continuous meta-regression* | *3-4 in smallest subgroup; 6-10 in continuous meta-regression* | *5-9 in smallest subgroup; 11 to 15 in continuous meta-regression* | *10 or more in smallest subgroup; more than 15 in continuous meta- regression* |
| Comment: Spinner only contributed 3% to the overall analysis | | | |
| **4: Was the direction of effect modification correctly hypothesized a priori?** | | | |
| [ ] Definitely no | [ x] Probably no or unclear | [ ] Probably yes | [ ] Definitely yes |
| *Clearly post-hoc or results inconsistent with hypothesized direction or biologically very implausible* | *Vague hypothesis or hypothesized direction unclear* | *No prior protocol available but unequivocal statement of a priori hypothesis with correct direction of effect modification* | *Prior protocol available and includes correct specification of direction of effect modification, e.g. based on a biologic rationale* |
| Comment: Due to the novelty of the COVID-19 disease, knowledge has evolved rapidly over time. The relevant Taskforce panel currently holds the view that if a patient requires ventilation, there is likely to be substantial contribution to their clinical condition from host immune response, and it is biologically plausible that antivirals would have reduced efficacy. However, the direction of effect modification was not explicitly discussed either by the Taskforce Guidelines Team or the relevant panel prior to the initiation of the relevant systematic review or publication of the first remdesivir trial results. | | | |

| **5: Does a test for interaction suggest that chance is an unlikely explanation of the apparent effect modification?** (consider irrespective of number of effect modifiers) | | | |
|---|---|---|---|
| [ ] Chance a very likely explanation | [ ] Chance a likely explanation or unclear | [x ] Chance may not explain | [ x] Chance an unlikely explanation |
| *Interaction or meta-regression p- value >0.05* | *Interaction or meta-regression p- value ≤0.05 and >0.01, or no test of interaction reported and not computable* | *Interaction or meta-regression p- value ≤0.01 and >0.005* | *Interaction or meta-regression p- value ≤0.005* |
| Comment: The test for subgroup differences was p=0.01. | | | |
| **6: Did the authors test only a small number of effect modifiers or consider the number in their statistical analysis?** | | | |
| [ ] Definitely no | [ ] Probably no or unclear | [ ] Probably yes | [ x] Definitely yes |
| *Explicitly exploratory analysis or large number of effect modifiers tested (e.g. greater than 10) and multiplicity not considered in analysis* | *No mention of number or 4-10 effect modifiers tested and number not considered in analysis* | *No protocol available but unequivocal statement of 3 or fewer effect modifiers tested* | *Protocol available and 3 or fewer effect modifiers tested or number considered in analysis* |
| Comment: Baseline severity was the only subgroup considered | | | |
| **7: Did the authors use a random effects model?** | | | |
| [ ] Definitely no | [ ] Probably no or unclear | [ ] Probably yes | [ x] Definitely yes |
| *Fixed (or common) effect or fixed effects model explicitly stated* | *Probably fixed effect(s) model* | *Probably random (or mixed) effects* | *Random (or mixed) effects explicitly stated* |
| Comment: Random effects used | | | |
| **8: If the effect modifier is a continuous variable, were arbitrary cut points avoided?** [x ] not applicable: not continuous but ordinal | | | |
| [ ] Definitely no | [ ] Probably no or unclear | [ x] Probably yes | [ ] Definitely yes |
| *Analysis based on exploratory cut point(s), e.g. picking cut point associated with highest interaction p-value* | *Analysis based on cut point(s) of unclear origin* | *Analysis based on pre-specified cut point(s), e.g. suggested by prior RCT* | *Analysis based on the full continuum, e.g. assuming a linear or logarithmic relationship* |
| Comment: Need for respiratory support is measured on an ordinal scale and the categories on this scale are not arbitrary. We performed a sensitivity analysis, moving the Beigel subgroup receiving high flow or NIV at baseline into the less severe group. The subgroup difference remained statistically significant (see below). | | | |




**9 Optional: Are there any additional considerations that may increase or decrease credibility?** (manual section 3.9) [x ] not applicable

| | [ ] yes, probably decrease | [ ] yes, probably increase | |
|---|---|---|---|

Comments: We did not identify additional considerations.

**10: How would you rate the overall credibility of the proposed effect modification?**

The overall rating should be driven by the items that decrease credibility. The following provides a sensible strategy:

- All responses definitely or probably decrease credibility or unclear -> very low
- Two or more responses definitely decrease credibility -> maximum usually low even if all other responses satisfy credibility criteria
- One response definitely decreases credibility -> maximum usually moderate even if all other responses satisfy credibility criteria
- Two responses probably decrease credibility -> maximum usually moderate even if all other responses satisfy credibility criteria
- No response options definitely or probably decrease credibility -> high very likely

Place a mark on the continuous line (e.g. hit "x" in electronic version)

| **Very low credibility** | **Low credibility** | **Moderate credibility** | **High credibility** |
|---|---|---|---|
| Very likely no effect modification modification<br>Use overall effect for each each subgroup | Likely no effect modification<br>Use overall effect for each subgroup but note remaining uncertainty | Likely effect modification<br>Use separate effects for each subgroup but note remaining | Very likely effect<br>Use separate effects for subgroup uncertainty |

Comment: Item 4 (whether the direction of effect modification was correctly hypothesized a priori) was the only item to probably decrease credibility. We acknowledge that much is still to be learned regarding the biology of COVID-19 and require further data to acheive high certainty regarding the effect of remdesivir on mortality.