

Manuel David Gil-Sierra^{1,2}
Maria del Pilar Briceño-
Casado³
Emilio Jesús Alegre-Del Rey²
Marina Sánchez-Hidalgo¹

Efficacy of early use of remdesivir: a systematic review of subgroup analysis

¹Department of Pharmacology, Facultad de Farmacia, Universidad de Sevilla, Facultad de Farmacia, Sevilla, Spain.

²Pharmacy Department, Hospital Universitario Puerto Real, Cadiz, Spain.

³Pharmacy Department, Hospital Universitario de Jerez de la Frontera, Jerez de la Frontera, Spain.

Article history

Received: 3 November 2021; Revision Requested: 7 January 2022; Revision Received: 30 January 2022; Accepted: 2 February 2022; Published: 17 March 2022

ABSTRACT

Introduction. A possible benefit has been suggested for early treatment of severe coronavirus disease 2019 (COVID-19) with remdesivir. The efficacy of this drug is controversial and could significantly influence the efficiency in healthcare systems. The objective is the methodological interpretation of subgroup analyzes according to starting of remdesivir treatment with respect to symptom onset of COVID-19.

Methods. A search in Pubmed® database was performed. Randomized clinical trials (RCTs) with subgroup analysis regarding early and late use of remdesivir were selected. All endpoints were assessed using two methodologies. First methodology considered statistical interaction, pre-specification, biological plausibility, and consistency of results. Second methodology was a validated tool with preliminary questions to discard subset analysis without relevant minimum conditions, and a checklist with recommendations for applicability.

Results. A total of 54 results were found and five RCTs were selected. According first methodology, consistent heterogeneity was only found in time to clinical improvement and better clinical status score at day 15 for patients with severe COVID-19 and <7 days of symptoms. About second methodology, these results about early use of remdesivir may be applied to clinical practice with caution.

Conclusions. We developed a systematic search and application of an established methodology for interpretation of subgroup analysis about early use of remdesivir. Results in severe COVID-19 suggested that early use of remdesivir provides a greater benefit in <7 days of symptoms for time to clinical improvement and better clinical status score at day 15. Future studies could use 7-day cut-off of symptoms to evaluate remdesivir.

Keywords: COVID-19; remdesivir; clinical decision-making; subgroup analysis; drug evaluation.

Correspondence:
Manuel David Gil-Sierra,
Pharmacy Department, Hospital Universitario Puerto Real. Carretera. N-IV Km. 665, 11510
Puerto Real, Cadiz, Spain.
E-mail: mangilsie@yahoo.com.

Eficacia del uso temprano de remdesivir: una revisión sistemática de análisis de subgrupos

RESUMEN

Introduction. Se ha sugerido un posible beneficio para el tratamiento temprano de la enfermedad grave por coronavirus 2019 (COVID-19) con remdesivir. La eficacia de este fármaco es controvertida y podría influir significativamente en la eficiencia de los sistemas sanitarios. El objetivo es la interpretación metodológica de los análisis de subgrupos según el inicio del tratamiento con remdesivir respecto al inicio de los síntomas de la COVID-19.

Material y métodos. Se realizó una búsqueda en la base de datos Pubmed®. Se seleccionaron ensayos clínicos aleatorizados (ECA) con análisis de subgrupos respecto al uso temprano y tardío de remdesivir. Todas las variables se evaluaron mediante dos metodologías. La primera metodología consideró la interacción estadística, pre-especificación, la plausibilidad biológica y la consistencia de los resultados. La segunda metodología fue una herramienta validada con preguntas preliminares para descartar el análisis de subgrupos sin condiciones mínimas relevantes, y una lista de verificación con recomendaciones de aplicabilidad.

Resultados. Se encontraron un total de 54 resultados y se seleccionaron cinco ECA. Según la primera metodología, sólo se encontró heterogeneidad consistente en el tiempo hasta la mejora clínica y la mejor puntuación del estado clínico en el día 15 para los pacientes con COVID-19 grave y <7 días de síntomas. Sobre la segunda metodología, estos resultados sobre el uso temprano de remdesivir pueden aplicarse a la práctica clínica con precaución.

Conclusiones. Se desarrolló una búsqueda sistemática y la aplicación de una metodología establecida para la interpretación del análisis de subgrupos sobre el uso temprano de remdesivir. Los resultados en la COVID-19 grave sugirieron que el uso temprano de remdesivir proporciona un mayor beneficio en <7 días de síntomas para el tiempo de mejora clínica y mejor puntuación del estado clínico en el día 15. Los estudios futuros podrían utilizar el corte de 7 días de síntomas para evaluar el remdesivir.

Palabras clave: COVID-19; remdesivir; toma de decisiones clínicas; análisis de subgrupos; evaluación de fármacos.

INTRODUCTION

Coronavirus disease 2019 (COVID-19) has affected 200 million people and it has led to more than four million deaths worldwide [1]. Different drugs were used experimentally to treat this infection due to emergency situation and many studies were developed [2,3]. However, only vaccines and corticosteroids showed a clear relevant benefit [4-7].

Remdesivir is a prodrug with *in vitro* activity against severe acute respiratory syndrome coronavirus 2. Clinical results of these antiviral were evaluated in randomized clinical trials (RCT). Initially, Wang *et al* assessed remdesivir *versus* placebo including patients from Wuhan [8]. No statistically significant difference was observed for any endpoint analyzed in this RCT, such as time to clinical improvement and mortality at day 28. Despite this, a possible benefit in time to clinical improvement was suggested in time to clinical improvement for early-treated COVID-19 regarding a subgroup data.

Subgroup analysis compares the effect of a health intervention in different fractions of population. Interpretation of subgroup analysis presents limitations, so it should be conducted prudently and with strict methodological criteria [9]. Additional determinations and redistributions of patients for each characteristic subdividing population increase the possibility of obtaining differences by chance [10]. Distribution of population into subgroups reduces the statistical power of an analysis and the ability to detect differences between subgroups. Therefore, α and β errors are increased in subgroup analysis. Thus, misinterpretations often occur. Some of them include intragroup results assessment of hypothetical benefit for an intervention in a subgroup of patients without statistical difference with respect to complementary subgroup [11].

The commercialization of remdesivir and the health crisis due to COVID-19 triggered a barrage of hypotheses and claims. One of these hypotheses was directed to the timing of the antiviral use, associating the early treatment of remdesivir with greater benefit for patients [8,12]. However, was it statistically shown? About what endpoints? What is the real cut-off to differentiate early from late use of remdesivir? The answers to these questions are controversial and the use of the drug could significantly influence the efficiency in health-care systems. A detailed assessment of this issue could prevent future mistakes in similar emergency situations. The objective of this study is the methodological interpretation of subgroup analyzes according to remdesivir treatment initiation with respect to symptom onset in patients with COVID-19, using established methodologies.

METHODS

Literature search. PICOS strategy was used to formulate the research question: population, intervention, comparator, outcome and study design. COVID-19 patients were included as population. Remdesivir treatment (or its combinations) was intervention. Any comparator was included. Outcome was sub-

group analysis regarding the symptom onset for any endpoint. RCT was the selected study design.

Preferred Reporting Items for a Systematic Review and Meta-analysis (PRISMA) guidelines was applied in the systematic search. The review was conducted in Pubmed® database up to August 20, 2021. Search strategy was performed using screening tool "Clinical Queries/Narrow": (*remdesivir AND covid*) AND (*Therapy/Narrow[filter]*). Subsequently, a citation tracking was conducted.

Screening and selection of studies. Two investigators developed the search independently. Disagreements were resolved by discussion. Titles and abstracts of review records were checked to identify and discard studies without the established inclusion criteria. Full text of search results was examined in eligibility process. RCTs about COVID-19 with a subgroup analysis according to use of remdesivir with respect to symptom onset were selected. Early use of remdesivir was defined as the start of treatment before a certain day from symptom onset, which was detailed in each study. Late use of remdesivir occurred after that cut-off. Records obtained from the review in a language other than Spanish or English were excluded.

Data extraction. RCT included were analyzed in order of publication date. The following data were collected: authors, publication date, population, intervention and comparator therapies, sample sizes, endpoints, number of days subdividing the global population and efficacy of treatments.

Data analysis. Evaluation criteria of subgroup analysis from Sun *et al* were applied [9]: statistical interaction, pre-specification of subgroups, biological plausibility supporting the observed effect and consistency of subgroup results with other studies or within the same trial. Statistical interaction assesses whether difference among different subgroups is compatible with chance using the probability of interaction $p(i)$. Considering the limitations of subset data, heterogeneity is defined as a statistically significant difference among subgroups [$p(i)<0.1$] [13]. Estimation of $p(i)$ was obtained with calculators using relative risk values, odds ratio, hazard ratio and confidence intervals [14-19]. If insufficient data were provided, $p(i)$ was evaluated using graphs when it was possible. Pre-specification of subgroups avoids the consideration of really non-existent differences caused by multiplicity. Biological plausibility evaluates the existence of hypotheses that justify differences between subgroups. Consistency assessed the agreement among subgroup data of similar studies or endpoints within the same RCT (internal consistency).

After that, a validated tool to assess the applicability of subgroup analysis was used [20]. It is a systematic methodology with two parts: preliminary questions to discard the assessment of subgroup results without relevant minimum conditions, and a checklist. Preliminary questions assess the evidence level of study, relevance of the endpoint evaluated, existence of a difference in effect [$p(i)<0.1$] among subgroups and temporal sequence between subpopulation generating

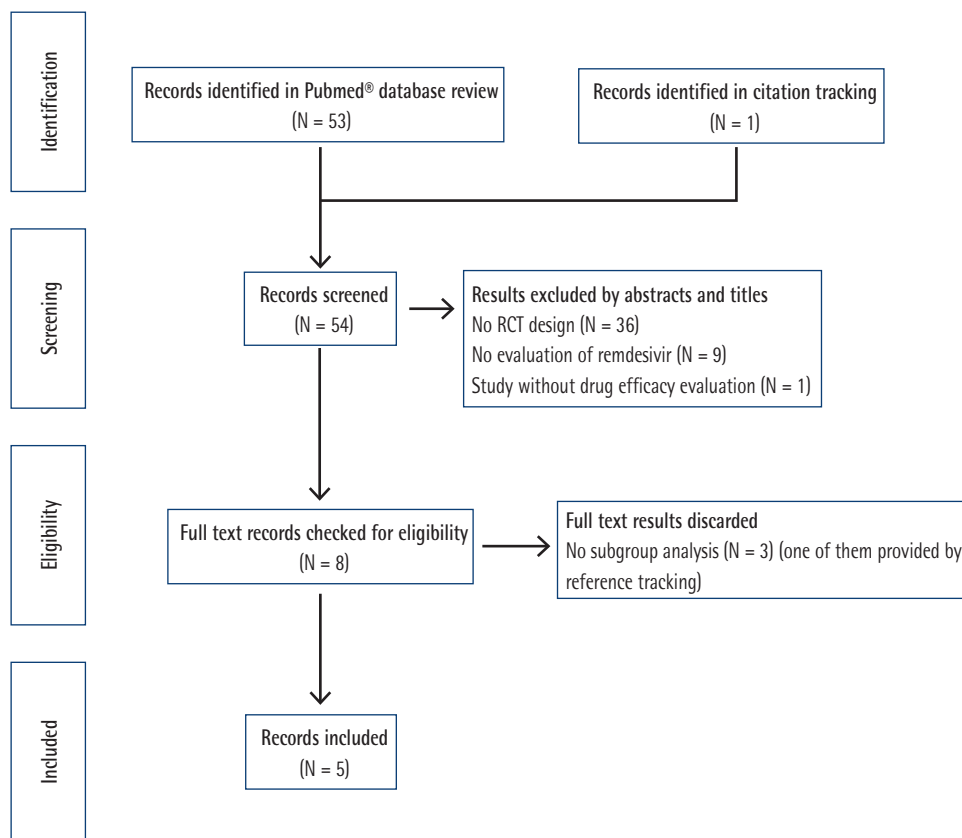


Figure 1 | Review of literature

factor and health intervention. A negative answer to any of preliminary questions discards subsequent questions, applicability of checklist and subgroup analysis. Checklist assigns an evaluation and score to a series of criteria for the interpretation of subgroup analysis: statistical association including $p(i)$, pre-specification of subgroups, sample size, number of factors considered, and overall outcome of study; biological plausibility and consistency. Higher scores are related to higher reliability of each criterion: probable (+3 points), possible (+2), doubtful (0) and null (-3). Overall score is associated with a recommendation for applicability of results in clinical practice. Higher overall scores correspond to greater applicability and recommendations for consideration in clinical decision-making: 9-7 points (probable recommendation) mean applicability of subgroup results; 5-6 points (possible recommendation) indicate prudent applicability; 3-4 points (doubtful recommendation) mean avoid applicability with exceptions; and <3 points (null recommendation) indicate no application of subgroup data.

Risk of bias evaluation. RCT results may be affected by the prognosis of recruited patients. Disease severity, coexisting comorbidities and baseline score on ordinal scales about

oxygen use of patients of RCTs were checked. Sample sizes of subgroups were evaluated.

RESULTS

Characteristics of studies. Fifty-three results were found in the systematic search. A RCT comparing 5-day and 10-day regimens of remdesivir without a subgroup analysis was found in citation tracking [21]. Exclusion criteria had the following distribution: 36 studies presented a design different from RCT design, 9 evaluated a drug other than remdesivir, 3 without subgroup analysis and one without efficacy consideration. Of the total of 54 records, 5 RCTs were included. Figure 1 illustrates the literature review regarding the PRISMA protocol and Table 1 shows data from included studies.

Subgroup analysis. Wang *et al* published a subgroup analysis regarding the effect of early and late use of remdesivir (with a cut-off for 10 days from symptom onset) for the following endpoints: time to clinical improvement, mortality at day 28 and viral RNA load on upper respiratory tract [8]. Criteria for interpretation of subgroup analysis described in Sun *et al* were applied [9]. $P(i)$ was not detailed in the trial

Table 1 Data from randomized clinical trials included in study

Authors	Online publication date	Population	Intervention	Comparator	Trial sample size	Endpoints with subgroup analysis	Subgroups according to days from symptom onset to treatment	Sample size of the early remdesivir use subgroup ^a	Sample size of the late remdesivir use subgroup ^a	Efficacy for endpoints in global population (95% CI)	Efficacy in subgroup with the early use of remdesivir (95% CI)	Efficacy in subgroup with the late use of remdesivir (95% CI)	
Wang et al. [8]	April 29, 2020	Patients with severe COVID-19	Remdesivir 200 mg on day 1 and remdesivir 100 mg on days 2-10	Placebo	237	Time to clinical improvement	≤10 days and >10 days	118	115	Hazard Ratio: 1.23 (0.87 to 1.75)	Hazard Ratio: 1.52 (0.95 to 2.43)	Hazard Ratio: 1.07 (0.63 to 1.83)	
						Mortality at day 28	≤10 days and >10 days	118	115	Rate differences: 1.1% (-8.1 to 10.3)	Rate differences: -3.6% (-16.2 to 8.9)	Rate differences: 4.6% (-8.2 to 17.4)	
						Viral RNA load on upper respiratory tract	≤10 days and >10 days	64	60	No data of viral load -Log10 copies/mL (graphs only)	No data of viral load -Log10 copies/mL (graphs only)	No data of viral load -Log10 copies/mL (graphs only)	
Spinner et al. [23]	August 21, 2020	Patients with moderate COVID-19	Remdesivir 200 mg on day 1 and remdesivir 100 mg on day 2-10 (5-days and 10-days course)	Standard of care	596	Clinical status on study day 11	5-days course: <9 days and ≥9 days	No data	No data	No data of difference in proportions (graphs only)	No data of difference in proportions (graphs only)	No data of difference in proportions (graphs only)	
							10-days course: <9 days and ≥9 days	No data	No data	No data of difference in proportions (graphs only)	No data of difference in proportions (graphs only)	No data of difference in proportions (graphs only)	
Beigel et al. (ACT-1 Study Group, final report) [24]	October 9, 2020	Patients with severe COVID-19	Remdesivir 200 mg on day 1 and remdesivir 100 mg on days 2-10	Placebo	1062	Time to clinical improvement	≤10 days and >10 days	676	383	Rate ratio: 1.29 (1.12-1.49)	Rate ratio: 1.37 (1.14-1.64)	Rate ratio: 1.20 (0.94-1.52)	
							≤9 days and >9 days	582	477	Rate ratio: 1.29 (1.12-1.49)	Rate ratio: 1.32 (1.09 to 1.61)	Rate ratio: 1.29 (1.04 to 1.59)	
							1st Quartile: <7 days ^b	282	777	Rate ratio: 1.29 (1.12-1.49)	Rate ratio: 1.92 (1.41 to 2.60)	-	
							2nd Quartile: 7 to ≤ 9 days ^b	300	-	Rate ratio: 1.29 (1.12-1.49)	Rate ratio: 0.99 (0.76 to 1.28)	-	
							3th Quartile: 10 to ≤ 12 days ^b	221	-	Rate ratio: 1.29 (1.12-1.49)	Rate ratio: 1.45 (1.07 to 1.98)	-	
							4th Quartile: ≥13 days ^b	803	256	Rate ratio: 1.29 (1.12-1.49)	Rate ratio: 1.15 (0.86 to 1.54)	-	
							Better (Lower) clinical status score at day 15	≤10 days and >10 days	676	383	Odds ratio: 1.6 (1.3 to 1.9)	Odds ratio: 1.7 (1.3 to 2.2)	Odds ratio: 1.3 (0.9 to 1.9)
								≤9 days and >9 days	582	477	Odds ratio: 1.6 (1.3 to 1.9)	Odds ratio: 1.7 (1.2 to 2.2)	Odds ratio: 1.4 (1.0 to 1.9)
								1st Quartile: <7 days ^b	282	777	Odds ratio: 1.6 (1.3 to 1.9)	Odds ratio: 2.7 (1.8 to 4.2)	-
								2nd Quartile: 7 to ≤ 9 days ^b	300	-	Odds ratio: 1.6 (1.3 to 1.9)	Odds ratio: 1.1 (0.8 to 1.7)	-
								3th Quartile: 10 to ≤ 12 days ^b	221	-	Odds ratio: 1.6 (1.3 to 1.9)	Odds ratio: 1.4 (0.9 to 2.3)	-
4th Quartile: ≥13 days ^b	803	256	Odds ratio: 1.6 (1.3 to 1.9)	Odds ratio: 1.3 (0.9 to 2.1)	-								

Table 1		Data from randomized clinical trials included in study (cont.)										
Authors	Online publication date	Population	Intervention	Comparator	Trial sample size	Endpoints with subgroup analysis	Subgroups according to days from symptom onset to treatment	Sample size of the early remdesivir use subgroup ^a	Sample size of the late remdesivir use subgroup ^a	Efficacy for endpoints in global population (95% CI)	Efficacy in subgroup with the early use of remdesivir (95% CI)	Efficacy in subgroup with the late use of remdesivir (95% CI)
Kalil et al. [25]	January 5, 2021	Patients with severe and moderate COVID-19	Remdesivir 200 mg on day 1 and remdesivir 100 mg on days 2-10 + baricitinib 4 mg for 14 days	Placebo + Remdesivir 200 mg on day 1 and remdesivir 100 mg on days 2-10	1033	Time to clinical improvement	≤10 days and >10 days	764	253	Rate ratio: 1.16 (1.01-1.32)	Rate ratio: 1.13 (CI 0.97-1.32)	Rate ratio: 1.27 (0.97-1.67)
Barratt-Due et al. [26]	July 13, 2021	Patients with severe and moderate COVID-19	Remdesivir 200 mg on day 1 and remdesivir 100 mg until day 9	Standard of care	185	Oropharyngeal viral clearance	<7 days and ≥7 days	No data	No data	Difference in daily decrease rate: 0.113 (-0.001 to 0.227)	Difference in daily decrease rate: 0.19 (0.03 to 0.36)	Difference in daily decrease rate: 0.02 (-0.15 to 0.19)

^aThe early use of remdesivir was associated with the subgroup of patients who received remdesivir with the lowest number of days between the onset of symptoms and the start of treatment. The late use of remdesivir is the subgroup of patients with the highest number of days in this period of time.

^bQuartiles about duration of symptoms prior to enrollment of patients in ACTT-1 Study Group (final report) only presented one efficacy data. Complementary subgroup in first and fourth quartiles was estimated from trial data.

Table 2		Summary of interpretation for subgroup analysis from studies.									
Methodology	Criteria	Wang et al. [8] Endpoints			Spinner et al. [23] Endpoints		Beigel et al (ACTT-1 Study Group) [24] Endpoints for 7 days of symptoms cut-off		Kalil et al. [25] Endpoints	Barratt-Due et al. [26] Endpoints	
		Time to clinical improvement	Mortality at day 28	Viral load on upper respiratory tract	Clinical status on study day 11		Time to clinical improvement	Better clinical status score at day 15	Time to clinical improvement	Oropharyngeal viral clearance	
					5-day course of remdesivir	10-day course of remdesivir					
Sun et al. [9]	Statistical interaction	No	No	Insufficient data	No	Yes	Yes	Yes	No	No	
	Pre-specification of analysis	Undefined	Undefined	Undefined	No	No	Yes	Yes	Yes	Yes	
	Biological plausibility	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
	Consistency of subgroup results	Results with internal consistency of no benefit	Results with internal consistency of no benefit	Insufficient data	No	No	Yes (internal consistency of greater benefit in early use of remdesivir)	Yes (internal consistency of greater benefit in early use of remdesivir)	Yes (consistency of lack of differences between early and late use of remdesivir)	No	
Validated tool (Gil-Sierra et al.) [20]	Preliminary questions	The study shows the highest level of evidence with subset analysis	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
		Clear clinical relevance of considered endpoint or primary surrogate outcome of study	Yes	Yes	No	Yes	Yes	Yes	Yes	No	
		Existence of difference in effect between the subgroups for the factor evaluated (p(i) < 0.1)	No	No	Not applied	No	Yes	Yes	Yes	No	Not applied
		Determining factor of subgroup analysis was present prior to health intervention	Not applied	Not applied	Not applied	Not applied	Yes	Yes	Yes	Not applied	Not applied
	Checklist		Not applied	Not applied	Not applied	Not applied	Applied	Applied	Applied	Not applied	Not applied
	Statistical association (score)						Null (-3 points)	Probable (+3 points)	Probable (+3 points)		
	Biological plausibility (score)						Probable (+3 points)	Probable (+3 points)	Probable (+3 points)		
	Consistency (score)						Null (-3 points)	Doubtful (0 points)	Doubtful (0 points)		
	Recommendation (sum)						Null (-3 points): "Subgroup analysis should not be considered"	Possible (6 points): "Subgroup results analysis may be applied to clinical practice with caution"	Possible (6 points): "Subgroup results analysis may be applied to clinical practice with caution"		

Table 3 Data about benefit-related factors of patients in studies.

Prognosis-related factors of patients	Wang et al. [8]	Spinner et al. ^a [23]	Beigel et al. (ACTT-1 Study Group) [24]	Kalil et al. [25]	Barratt-Due et al. [26]
Disease severity (%)					Moderate and severe
Moderate	0%	100%	9.9%	31.7%	Insufficient data
Severe	100%	0%	90.1%	68.3%	Insufficient data
Coexisting comorbidities (%)	71%	Insufficient data	81.7%-82.2%	87.1%-81.7%	Insufficient data
Hypertension	46%-38%	44%-43%-41%	50.6%-50.9%	51%-52%	36.6%-24.6%
Diabetes	25%-21%	44%-37%-38%	32%-31%	40%-36%	22%-15.8%
Cardiovascular disease	9%-3% ^b	58%-58%-54%	18%-16%	21%-21%	14.6%-21.1%
Obesity	Insufficient data	Insufficient data	46%-45%	58%-53%	28.9%-18.4%
Baseline score on ordinal scales about oxygen use (%)					
Hospitalized, requiring supplemental oxygen	82%-83%	12%-16%-19%	42.9%-39%	55.9%-53.3%	Insufficient data
Hospitalized, receiving non-invasive ventilation or high-flow oxygen devices	18%-12%	0%	17.6%-18.8%	20%-21.8%	Insufficient data
Hospitalized, receiving invasive mechanical ventilation or extracorporeal membrane oxygenation	0%-1%	0%	24.2%-29.6%	10.5%-11%	Insufficient data
Ordinal scale	Six-category scale	Seven-category scale	Eight-category scale	Eight-category scale	Ten-category scale

Data with one result details the percentage of global population in study. Data with two results represents the percentages collected from intervention and control arms (% of patients in intervention arm-% of patients in control arm). ^aSpinner et al presented three arms [23]: 10-day course of remdesivir, 5-day course of remdesivir and control arm (% of patients in 10-day remdesivir arm-% of patients in 5-day remdesivir arm-% of patients in control arm). ^bOnly coronary heart disease.

for any endpoint. It was calculated and no heterogeneity was found between subgroups of early and late use of remdesivir for time to clinical improvement [$p(i)=0.33$] and mortality at day 28 [$p(i)=0.37$]. Insufficient data were provided to calculate $p(i)$ between subgroups for viral RNA load on upper respiratory tract. Wang *et al* commented that no statistically significant difference was found for viral RNA load in upper respiratory tract between control and intervention arms in early and late use of remdesivir without mentioning $p(i)$. Pre-specification of subgroup analyses was not defined in study protocol for any outcome. Biological plausibility supporting the early use of remdesivir in patients with COVID-19 appeared reasonable. This hypothesis was based on experience with viruses such as influenza or other pathogens [12,22]. Wang *et al* study was the first RCT about use of remdesivir in COVID-19 and consistency with previous trials could not be assessed. However, absence of heterogeneity between subgroups for time to clinical improvement and mortality at day 28 showed internal consistency of subgroup analysis [8].

Subsequently, the validated tool about applicability of subgroup analysis was used [20]. Some of preliminary questions were answered negatively in the three endpoints so checklist was not developed. Subgroup analysis regarding time of drug administration for time to clinical improvement and mortality at day 28 presented no differences of effect among subgroups [$p(i)>0.1$]. Insufficient data were provided for heterogeneity estimation for viral RNA load on upper respiratory tract. Likewise, it is a surrogate endpoint with little clinical relevance. Table 2 shows a summary of the interpretation for

subgroup analysis described in Wang *et al* [8].

Spinner *et al* provided a trial with subgroup analysis based on early and late use of remdesivir (with a cut-off for 9 day) for clinical status on study day 11, for both 5-day and 10-day courses of remdesivir [23]. At first, criteria for subgroup analysis interpretation of Sun *et al* were used [9]. $P(i)$ value was not described for clinical status on day 11 and no exact data were reported. However, forest plot in subset analysis showed almost total overlap between subgroups of early and late use of remdesivir for 5-day course [$p(i)>0.2$], while overlap was less than 50% for 10-day course. Difference in proportions and 95% confidence intervals (95%CI) were extracted from graphical representations of forest plot for 10-day scheme. Difference in proportions in early use of remdesivir (<9 days) was 12.6 (95%CI, -0.6 to 5.3) and late use of remdesivir (≥ 9 days) was -4 (95%CI, -18.6 to 10). $P(i)$ between these subgroups was estimated as 0.0393. Subgroup analysis about time of remdesivir use with respect symptom onset was not prespecified (*post hoc* analysis). Biological plausibility supported early use of remdesivir, in accordance with previous experience with viruses and other microorganisms [12,22]. There was internal inconsistency between subgroup results of 5-day and 10-day courses of treatment within the same trial [23]. One remdesivir scheme proved a statistically significant difference between subgroups according to timing of remdesivir use and the other scheme did not. Different cut-off was evaluated in Wang *et al* (10 days), which showed no differences in effect between subgroups [8]. No comparisons can be performed among subgroups with different time periods.

Validated tool about applicability of subgroup analysis was applied [20]. For 5-day course of remdesivir, preliminary question regarding the existence of differences in effect between subgroups was answered negatively for clinical status on study day 11 [$p(i) > 0.2$]. For 10-day course of remdesivir, all preliminary questions were answered positively. When applying the checklist, statistical association presented "null" evaluation (-3 points) due to lack of pre-specification of the considered factor (*post hoc* analysis). Biological plausibility had "probable" assessment (+3 points) and consistency was "null" (-3 points). The sum of these scores (-3 points) was associated with "null" recommendation for applicability in clinical practice of subset results. Therefore, results of early use subgroup (<9 days) from 10-day regimen of remdesivir should not be considered for decision-making in moderate COVID-19. A summary of interpretation about subgroup analyzes described in Spinner *et al* can be consulted in Table 2.

Beigel *et al* (final report of ACTT-1 Study Group) published a subgroup analysis regarding duration of symptoms prior to enrollment for two endpoints: time to clinical improvement and better clinical status score at day 15 [24]. Subset analyses evaluated these efficacy endpoints across following cut-offs: quartiles of days (<7 days, 7 to ≤ 9 days, 10 to ≤ 12 Days, >13 Days); ≤ 9 days *versus* >9, and ≤ 10 days *versus* >10. Criteria of Sun *et al* were used to analyze subset analysis [9]. No $p(i)$ value was reported for outcomes in different cut-offs. It was calculated and heterogeneity [$p(i) = 0.009$] was found in the first quartile of days (subgroup with <7 days of symptoms before enrollment *versus* ≥ 7 days) for time to clinical improvement. $P(i) < 0.05$ was also estimated in the first quartile for better clinical status score at day 15. However, a value of $P(i) > 0.2$ was quantified at any of other cut-offs for both endpoints. Subgroup analysis about duration of symptoms before to enrollment was pre-specified in trial protocol. Biological plausibility justified early use of remdesivir according to experience with other microorganisms [12, 22]. Internal consistency was found between subgroup results for time to clinical improvement and better clinical status score at day 15. Both endpoints presented statistically significant improvement [$p(i) < 0.05$] for patients with early treatment of remdesivir in <7 days of symptoms. On the other hand, results at cut-offs for 9 and 10 days were compared with previously published data [8, 23]. None of these studies showed consistent benefit for early use of remdesivir at the day 9 and 10 thresholds. The rest of subgroups (7 to ≤ 9 days, 10 to ≤ 12 Days and >13 Days) were not considered in previous studies and therefore could not be compared with Beigel *et al* [24].

The tool about applicability of subgroup analysis was used [20]. Preliminary questions were answered positively in the first quartile (<7 days of symptoms before enrollment *versus* ≥ 7 days) for time to clinical improvement and better clinical status score at day 15. When applying the checklist, statistical association showed "probable" consideration (+3 points). Biological plausibility also presented "probable" assessment (+3 points). Consistency had "doubtful" evaluation (0 points). The sum of scores (+6 points) was related with "possible" rec-

ommendation for applicability in clinical practice of subgroup results. Thus, a greater benefit of early use of remdesivir in <7 days of symptoms for time to clinical improvement and better clinical status score at day 15 may be applied to clinical practice with caution in severe COVID-19. Preliminary questions about the existence of differences in effect among subgroups were answered negatively for the rest of cut-offs in selected endpoints. A summary of subgroup analysis interpretation about 7 days of symptoms cut-off in final report of ACTT-1 Study Group is shown in Table 2.

Kalil *et al* compared baricitinib plus remdesivir *versus* remdesivir [25]. This RCT published a subgroup analysis regarding symptom onset (with a cut-off for 10 days) for time to clinical improvement. Subgroup analysis criteria of Sun *et al* were applied [9]. No $p(i)$ estimation was presented for the evaluated endpoint. $P(i)$ was calculated and no heterogeneity was observed between subgroups of early and late use of antiviral drug [$p(i) = 0.46$]. Subgroup analysis regarding to duration of symptoms was pre-specified in trial protocol. Biological plausibility justifying early use of remdesivir was based on previous experience with viruses and other microorganisms [12,22]. Prior studies on the use of remdesivir in COVID-19 containing subgroup analysis according to symptom onset (with a cut-off for 10 days) showed no heterogeneity of the antiviral effect [8,24]. These data could be consistent with subgroup results reported in Kalil *et al* [25].

Afterwards, validated clinical applicability tool for subgroup analysis was applied [20]. Preliminary question about the existence of differences in effect among subgroup according to early or late use of remdesivir was answered negatively [$p(i) > 0.1$]. Therefore, the rest of the tool was not applied. Table 2 describes a summary of interpretation about subgroup analysis presented in Kalil *et al*.

Barratt-Due *et al* presented a subset analysis about the time of use of remdesivir respect to symptom onset (with a cut-off for 7 days) for oropharyngeal viral clearance [26]. Sun *et al* criteria for interpretation of subgroup analysis were applied [9]. $P(i)$ was not reported in the trial for oropharyngeal viral clearance. Own estimations showed no heterogeneity between subsets of early and late use of remdesivir for the selected endpoint [$p(i) = 0.166$]. Pre-specification of subset analysis according to symptom onset was defined in protocol. Biological plausibility of early use of remdesivir for COVID-19 was justified by experience with other infectious pathogens [12,22]. The absence heterogeneity between subsets of early and late use of remdesivir in the cut-off for 7 days of Barratt-Due *et al* was no consistent compared to statistically significant differences observed in Beigel *et al* [24,26].

Thereupon, the validated tool about subset analysis was used [20]. Preliminary question on the relevance of the endpoint evaluated in selected subgroups was answered negatively, so the rest of the tool was not applied. A summary of interpretation about subset analysis reported in Barratt-Due *et al* is detailed in Table 2.

Risk of bias evaluation. Three RCTs included a popula-

tion with moderate and severe COVID-19, one RCT evaluated only patients with moderate COVID-19 and other assessed only patients with severe disease. The percentage of patients with comorbidities in the RCTs included in this study ranged from 71% to 87.1%. Hypertension was the most frequently recorded comorbidity in all RCTs (24.6% to 52%). Baseline score on ordinal scales about oxygen use of patients presented worse results in Beigel *et al* and Kalil *et al* [24,25], where percentage of patients with high-flow oxygen devices or non-invasive ventilation was 17.6% to 21.8% and patients with invasive mechanical ventilation or extracorporeal membrane oxygenation ranged from 10.5% to 29.6%. RCTs used different ordinal scales: one study applied a six-category ordinal scale, one evaluated patients with a seven-category scale, two RCTs considered an eight-category scale and one applied a ten-category scale. Data about prognosis-related factors of patients were detailed in Table 3.

Wang *et al* assessed more than 200 patients in subgroup analysis regarding time of antiviral drug for time to clinical improvement and mortality at day 28; and less than 130 patients for viral RNA load on upper respiratory tract [8]. More than 1,000 patients were included in subgroup analysis of Beigel *et al* and Kalil *et al* for time to clinical improvement and better clinical status score at day 15 [24,25]. The rest of endpoints in RCTs presented insufficient data about sample sizes of subsets. Table 1 provides the number of patients involved in subgroup analysis.

DISCUSSION

Results in patients with severe COVID-19 suggested that early use of remdesivir provides benefit compared to late use only in time to clinical improvement and better clinical status score at day 15 [24]. The effect of remdesivir in these endpoints has been statistically superior in patients with <7 days of symptoms. However, Barratt-Due *et al* found no differences between early and late use of remdesivir at 7 days cut-off for oropharyngeal viral clearance [26]. Oropharyngeal viral clearance is a surrogate endpoint with little clinical effect. It is a local measure and severe COVID-19 is a systemic disease with mainly pulmonary involvement. This could be an explanation for inconsistency of early and late subgroup results at 7 days cut-off presented in Barratt-Due *et al* and Beigel *et al* [24]. On the other hand, no greater effect of early use of remdesivir was observed for other cut-offs and outcomes. Thus, previous experience of greater benefit from early neuraminidase treatment in patients with influenza infection was partially confirmed in early use of remdesivir against COVID-19 [27], since benefit was observed in outcomes with limited clinical relevance.

The results of our study are complementary to WHO Solidarity trial data [28], which is the most important study on the use of COVID-19 treatments. Authors of WHO Solidarity trial concluded that remdesivir showed little or no effect on hospitalized patients with COVID-19, according to overall mortality, time of hospitalization and initiation of ventilation. This study

was not included in our review due to lack of subset analysis of remdesivir effect regarding the time of symptom onset.

For cut-off of 9 days of symptoms, our review found apparent heterogeneity between subgroups of 10-day course of remdesivir for clinical status on day 11 in moderate COVID-19 trial of Spinner *et al* [23]. These differences among subsets regarding symptom onset were calculated from graphical representations –without exact data provided by authors–. Furthermore, *post hoc* nature of this subset analysis should not be forgotten. Likewise, Spinner *et al* and other similar studies reported no differences between subgroups of early and late use of 5-day remdesivir regimen for cut-off of 9 days of symptoms [23,24]. The apparent differences between subgroups in 10-day course of remdesivir can be attributed to a multiplicity of determinations and limitations of subgroups [10].

Subgroup analyses should be considered with caution due to their limitations [9,10,29]. Wang *et al* committed one of the most frequent methodological errors. They considered intragroup differences between remdesivir and placebo in early treatment subset for time to clinical improvement and viral load in upper respiratory tract at 10-day cut-off [8,15,30,31]. No interaction test was calculated with complementary subset. Numerical differences in time to clinical improvement favorable to early use of remdesivir was highlighted, facilitating an inadequate interpretation of subgroup analysis that may influence clinical decision-making. In addition, differences of effect observed in a subset analysis should not always apply. First, clinical relevance of endpoints needs to be evaluated. Both viral load and time to clinical improvement can be considered as measures of little clinical relevance if they are not related with a reduction in mortality.

Our review found patients with different characteristics in RCTs about the use of remdesivir in COVID-19 [8,23–26]. There were studies with only moderate or severe COVID-19 and others recruited both populations. Variability was also observed in oxygenation and ventilation for COVID-19 patients in RCTs. Beigel *et al* and Kalil *et al* presented the worst results of baseline score on ordinal scales about oxygen use [24,25]. Further, patients were evaluated using different scales. On the other hand, Beigel *et al* and Kalil *et al* developed subgroup analysis regarding the timing of remdesivir with the largest sample sizes. Generally, studies did not provide all the essential information for the subset data.

This work is an illustrative example of how a methodological assessment of subset analysis can avoid making premature statements. Two previous publications were used. Sun *et al* established an adequate basis for the assessment of subgroup analyzes, but inexperienced evaluators may doubt the importance or order about interpretation criteria [9]. This possible limitation can be minimized by validated tool of Gil-Sierra *et al* [20], that also values additional considerations with respect to the first methodology.

Although many drugs –such as lopinavir/ritonavir or hydroxychloroquine– have been widely used, only glucocorticoids and vaccines showed a clear reduction of mortality in COV-

ID-19 [3, 4, 32,33]. Uncertainty about the effect of therapeutic alternatives is still high [5]. The development of systematic methodologies to evaluate new scientific evidence is necessary to reduce superfluous or negative effects of drugs -and unnecessary expenses- in patients with COVID-19. Minimum costs of production for remdesivir have been estimated at US \$0.93/day [34]. However, remdesivir acquisition price is much higher than costs of production. Therefore, optimization of the use of this antiviral is very important for efficiency of health systems due to its important economic impact.

CONCLUSION

We conducted a study with a systematic search and application of an established methodology for interpretation of subgroup analysis about early use of remdesivir in COVID-19. Moreover, our review detailed essential estimates for interpretation of subset analyzes not previously described. This work found a statistically significant superior benefit of early use of remdesivir for patients with severe COVID-19 and <7 days of symptoms for time to clinical improvement and better clinical status score at day 15. No greater benefit was associated with early use of remdesivir in other outcomes or time cuts. Finally, it seems reasonable to apply the 7-day cut-off from symptoms onset to evaluate the early use of remdesivir for COVID-19 in future studies.

FUNDING

None to declare.

CONFLICTS OF INTEREST

Manuel David Gil-Sierra: membership of an advisory board (consultation fees), lectures for Janssen Pharmaceutica and Pfizer (reimbursement for attending symposia) of another cancer drugs. The rest of the authors have no conflict of interest to declare.

REFERENCES

1. Johns Hopkins Coronavirus Resource Center. COVID-19 map. Available in: <https://coronavirus.jhu.edu/map.html>. Accessed August 20, 2021.
2. Potere N, Di Nisio M, Rizzo G, et al. Low-Dose subcutaneous tocilizumab to prevent disease progression in patients with moderate COVID-19 pneumonia and hyperinflammation. *Int J Infect Dis*. 2020; S1201-9712(20)30616-0. Doi: 10.1016/j.ijid.2020.07.078
3. Cao B, Wang Y, Wen D, et al. A trial of lopinavir-ritonavir in adults hospitalized with severe COVID-19. *N Engl J Med*. 2020;382(19):1787-99. Doi: 10.1056/NEJMoa2001282
4. Horby P, Lim WS, Emberson JR, et al. Dexamethasone in hospitalized patients with COVID-19 – preliminary report. *N Engl J Med*. 2021;384(8):693-704. Doi: 10.1056/NEJMoa2021436
5. Siemieniuk RA, Bartoszko JJ, Ge L, et al. Drug treatments for COVID-19: living systematic review and network meta-analysis. *BMJ*. 2020;370:m2980. Doi: 10.1136/bmj.m2980
6. Baden LR, El Sahly HM, Essink B, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med*. 2021;384(5):403-416. Doi: 10.1056/NEJMoa2035389
7. Polack FP, Thomas SJ, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med*. 2020;383(27):2603-2615. Doi: 10.1056/NEJMoa2034577.
8. Wang Y, Zhang D, Du G, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet*. 2020;395(10236):1569-78. Doi: 10.1016/S0140-6736(20)31022-9
9. Sun X, Ioannidis JPA, Agoritsas T, et al. How to use a subgroup analysis: users' guide to the medical literature. *JAMA*. 2014;311(4):405-11. Doi: 10.1001/jama.2013.285063
10. Brookes ST, Whitley E, Peters TJ, et al. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess*. 2001;5(33):1-56. Doi: 10.3310/hta5330
11. Lagakos SW. The challenge of subgroup analyses – reporting without distorting. *N Engl J Med*. 2006;354(16):1667-9. Doi: 10.1056/NEJMp068070
12. Gillenwater S, Rahaghi F, Hadeh A. Remdesivir for the treatment of COVID-19 - preliminary report. *N Engl J Med*. 2020;383. Doi: 10.056/NEJMc2022236
13. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ*. 2010;340:c117. Doi: 10.1136/bmj.c117
14. Sociedad Española de Farmacia Hospitalaria (2016). Grupo GENESIS. Utilidades: calculadoras y programas. Available in: https://gruposedetrabajo.sefh.es/genesis/index.php?option=com_content&view=article&id=46&Itemid=42. Accessed August 15, 2021.
15. Altman DG, Matthews JN. Statistics notes. Interaction 1: Heterogeneity of effects. *BMJ*. 1996;313(7055):486. Doi: 10.1136/bmj.313.7055.486
16. Hancock MJ, Kjaer P, Korsholm L, Kent P. Interpretation of subgroup effects in published trials. *Phys Ther*. 2013;93(6):852-9. Doi: 10.2522/ptj.20120296
17. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med*. 1998;17(24):2815-34. Doi: 10.1002/(sici)1097-0258(19981230)17:24<2815::aid-sim110>3.0.co;2-8
18. Tierney JF, Stewart LA, Ghersi D, et al. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials*. 2007;8:16. Doi: 10.1186/1745-6215-8-16
19. Vale CL, Tierney JF, Stewart LA. Effects of adjusting for censoring on meta-analyses of time-to-event outcomes. *Int J Epidemiol*. 2002;31(1):107-11. Doi: 10.1093/ije/31.1.107
20. Gil-Sierra MD, Fénix-Caballero S, Abdel Kader-Martin L, et al. Checklist for clinical applicability of subgroup analysis. *J Clin Pharm Ther*. 2020;45(3):530-8. Doi: 10.1111/jcpt.13102

21. Goldman JD, Lye DCB, Hui DS, et al. Remdesivir for 5 or 10 days in patients with severe COVID-19. *N Engl J Med.* 2020; 383(19):1827-1837. Doi: 10.1056/NEJMoa2015301
22. European Medicines Agency. Veklury. Available in: <https://www.ema.europa.eu/en/medicines/human/EPAR/veklury>. Accessed August 18, 2021.
23. Spinner CD, Gottlieb RL, Criner GJ, 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(11):1048-1057. Doi: 10.1001/jama.2020.16349.
24. Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the treatment of Covid-19 - final report. *N Engl J Med.* 2020; 383(19):1813-1826. Doi: 10.1056/NEJMoa2007764.
25. Kalil AC, Patterson TF, Mehta AK, et al. Baricitinib plus remdesivir for hospitalized adults with Covid-19. *N Engl J Med.* 2021;384(9):795-807. Doi: 10.1056/NEJMoa2031994.
26. Barratt-Due A, Christoffer Olsen I, Nezvalova-Henriksen K, et al. Evaluation of the effects of remdesivir and hydroxychloroquine on viral clearance in COVID-19: A randomized trial. *Ann Intern Med.* 2021;M21-0653. Doi: 10.7326/M21-0653.
27. Muthuri SG, Venkatesan S, Myles PR, et al. Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A H1N1pdm09 virus infection: a meta-analysis of individual participant data. *Lancet Respir Med.* 2014;2(5):395-404. Doi: 10.1016/S2213-2600(14)70041-4
28. Pan H, Peto R, Henao-Restrepo AM, et al. Repurposed antiviral drugs for Covid-19 - interim WHO Solidarity trial results. *N Engl J Med.* 2021;384(6):497-511. Doi: 10.1056/NEJMoa2023184.
29. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet.* 2005;365(9454):176-86. Doi: 10.1016/S0140-6736(05)17709-5
30. Matthews JN, Altman DG. Interaction 3: How to examine heterogeneity. *BMJ.* 1996;313(7061):862. Doi: 10.1136/bmj.313.7061.862
31. Matthews JNS, Altman DG. Statistics notes. Interaction 2: compare effect sizes not P values. *BMJ.* 1996;313(7060):808. Doi: 10.1136/bmj.313.7060.808
32. Alia E, Grant-Kels JM. Does hydroxychloroquine combat COVID-19? A timeline of evidence. *J Am Acad Dermatol.* 2020; 83(1):e33-e34. Doi: 10.1016/j.jaad.2020.04.03
33. Polack FP, Thomas SJ, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med.* 2020; 383(27):2603-2615. Doi: 10.1056/NEJMoa2034577
34. Hill A, Wang J, Levi J, et al. Minimum costs to manufacture new treatments for COVID-19. *J Virus Erad.* 2020;6(2):61-69. Doi: 10.1016/S2055-6640(20)30018-2