



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



ELSEVIER

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Selecting appropriate endpoints for assessing treatment effects in comparative clinical studies for COVID-19



Zachary R. McCaw^{a,1}, Lu Tian^{b,1}, Kevin N. Sheth^c, Wan-Ting Hsu^d, W. Taylor Kimberly^e, Lee-Jen Wei^{f,*}

^a Google, Mountain View, CA, United States of America

^b Department of Biomedical Data Science, Stanford University, Stanford, CA, United States of America

^c Division of Neurocritical Care and Emergency Neurology, Yale School of Medicine, New Haven, CT, United States of America

^d Medical Wisdom, LLC, Brookline, MA, United States of America

^e Division of Neurocritical Care, Department of Neurology, Massachusetts General Hospital, Boston, MA, United States of America

^f Harvard T.H. Chan School of Public Health, Boston, MA, United States of America

ABSTRACT

To evaluate the efficacy and safety of a new treatment for COVID-19 vs. standard care, certain key endpoints are related to the duration of a specific event, such as hospitalization, ICU stay, or receipt of supplemental oxygen. However, since patients may die in the hospital during study follow-up, using, for example, the duration of hospitalization to assess treatment efficacy can be misleading. If the treatment tends to prolong patients' survival compared with standard care, patients in the new treatment group may spend more time in hospital. This can lead to a "survival bias" issue, where a treatment that is effective for preventing death appears to prolong an undesirable outcome. On the other hand, by using hospital-free survival time as the endpoint, we can circumvent the survival bias issue. In this article, we use reconstructed data from a recent, large clinical trial for COVID-19 to illustrate the advantages of this approach. For the analysis of ICU stay or oxygen usage, where the initiating event is potentially an outcome of treatment, standard survival analysis techniques may not be appropriate. We also discuss issues with analyzing the durations of such events.

1. Introduction

When assessing a new treatment vs. standard care in comparative clinical trials for COVID-19, certain key secondary endpoints are related to the duration of specific, undesirable events over the study period. Examples include the durations of supplemental oxygen usage, invasive mechanical ventilation, hospitalization, and ICU stay [1–4]. Heuristically, the longer the duration, the worse the study therapy. However, since death is a possible study outcome, if a patient dies early in the study, their length of hospital stay is rather short. On the other hand, if a patient survives longer, they may spend more time in the hospital. Therefore, if the new treatment tends to provide a survival benefit over standard care, treatment efficacy as evaluated by the duration of, for example, hospitalization, can be misleading. This "survival bias" can make a treatment that prolongs survival appear to be less effective or worse than standard care. Common strategies for mitigating this bias may not be appropriate. For example, one may consider the durations of hospitalization among survivors only, or one may impute the durations of hospitalization for patients who die to an arbitrarily large value. However, neither of these strategies will provide a fair

comparison if the mortality rates of the two arms differ appreciably. In this paper, we present a valid and clinically meaningful alternative based on the mean event-free survival time.

To illustrate this bias and present alternative endpoints, Fig. 1 shows 4 possible observation patterns for a typical COVID-19 study monitoring the duration of hospitalization. For case 1, the patient was discharged alive from the hospital on day 7. Across the 28 days of follow-up, their duration of hospital-free survival was 21 days. For case 2, the patient died after 14 days in hospital. Although the duration of hospitalization was 14 days, their duration of hospital-free survival was 0 days. For case 3, the patient was censored after 21 days in hospital, due for example to late entry into the study. Their duration of hospital-free survival is less than 7 days. Lastly, for case 4, the patient remained in hospital across the 28 days of follow-up. Their duration of hospital-free survival was 0 days. It is important to note that although the length of hospitalization may be misleading as an endpoint in the presence of death, the hospital-free survival time is a clinically interpretable endpoint that is not subject to the survival bias issue.

* Corresponding author at: Department of Biostatistics, Harvard University, 655 Huntington Avenue, Boston, MA 02115, USA.

E-mail address: wei@hsph.harvard.edu (L.-J. Wei).

¹ These authors contributed equally.

<https://doi.org/10.1016/j.cct.2020.106145>

Received 9 July 2020; Received in revised form 18 August 2020; Accepted 7 September 2020

Available online 12 September 2020

1551-7144/ © 2020 Elsevier Inc. All rights reserved.

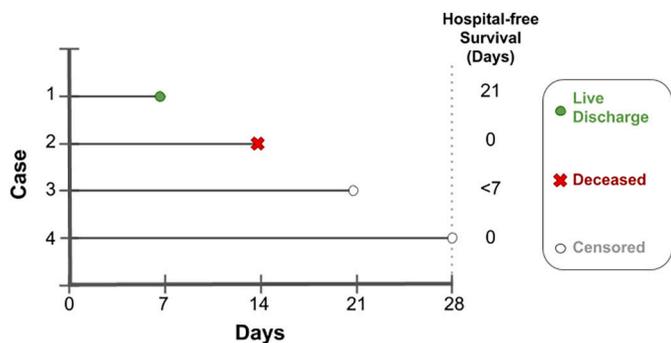


Fig. 1. Possible patterns for the duration of hospitalization in trials for COVID-19.

2. Example

We now use a recent large clinical trial, ACTT-1 [1], to illustrate the advantages of using the event-free survival time for assessing treatment efficacy. Adaptive COVID-19 Treatment Trial (ACTT-1) is an ongoing, double-blinded, randomized, placebo-controlled trial of remdesivir versus placebo among patients hospitalized with COVID-19. Across the 28 days of follow-up, patient health was classified on an 8-point ordinal scale, spanning from category 1, discharge from hospital with no limitation of activities, to category 8, death. The primary endpoint was time-to recovery, defined as the first time during follow-up on which the patient reached categories 1, 2, or 3. Patients in these categories were either no longer hospitalized, or hospitalized but no longer requiring supplemental oxygen or attentive medical care (e.g., remained

hospitalized for infection-control reasons). There were 538 and 521 patients assigned to remdesivir and placebo, respectively. Although not yet reported, key secondary endpoints discussed in the trial protocol included the durations of hospitalization, supplemental oxygen, and mechanical ventilation [5].

Since the paper presented no results for the length of hospitalization endpoint, we considered a similar endpoint, the time patients spent in hospital receiving attentive medical care, to demonstrate the appropriate analysis of event-free survival data. This undesirable endpoint is complementary to the study's primary endpoint of time to recovery. To this end, we scanned the cumulative recovery rate curves in Fig. 2A and also the survival curves in Supplementary Fig. S3 of the original paper [1,6]. The reconstructed data consist of the patient's duration of the above undesirable event (hospitalized and receiving supplemental oxygen or ongoing medical care). As illustrated in our Fig. 1, the patient's observed duration of the undesirable event ended upon recovery, death, or censoring. There were 44 and 60 deaths and 132 and 169 censorings during follow-up for remdesivir and placebo. To analyze the reconstructed data, Fig. 2A presents the cumulative incidence curves for the proportion of patients not requiring attentive hospital care (including deceased patients). The areas above the cumulative incidence curves (Figs. 2B & C) represent the mean times, up to day 28, that patients spent in hospital receiving attentive care [7–10]. These were 13.4 days for placebo, and 12.0 days for remdesivir. The difference was 1.4 days (95% CI, 0.19 to 2.6 days; $P = 0.024$) in favor of remdesivir. That is, on average, patients receiving placebo spent an additional 1.4 days hospitalized and receiving oxygen or medical care across the 28 days of follow-up. Note that, in this case, the length of hospital stay was truncated by death. Since there were more deaths in the standard

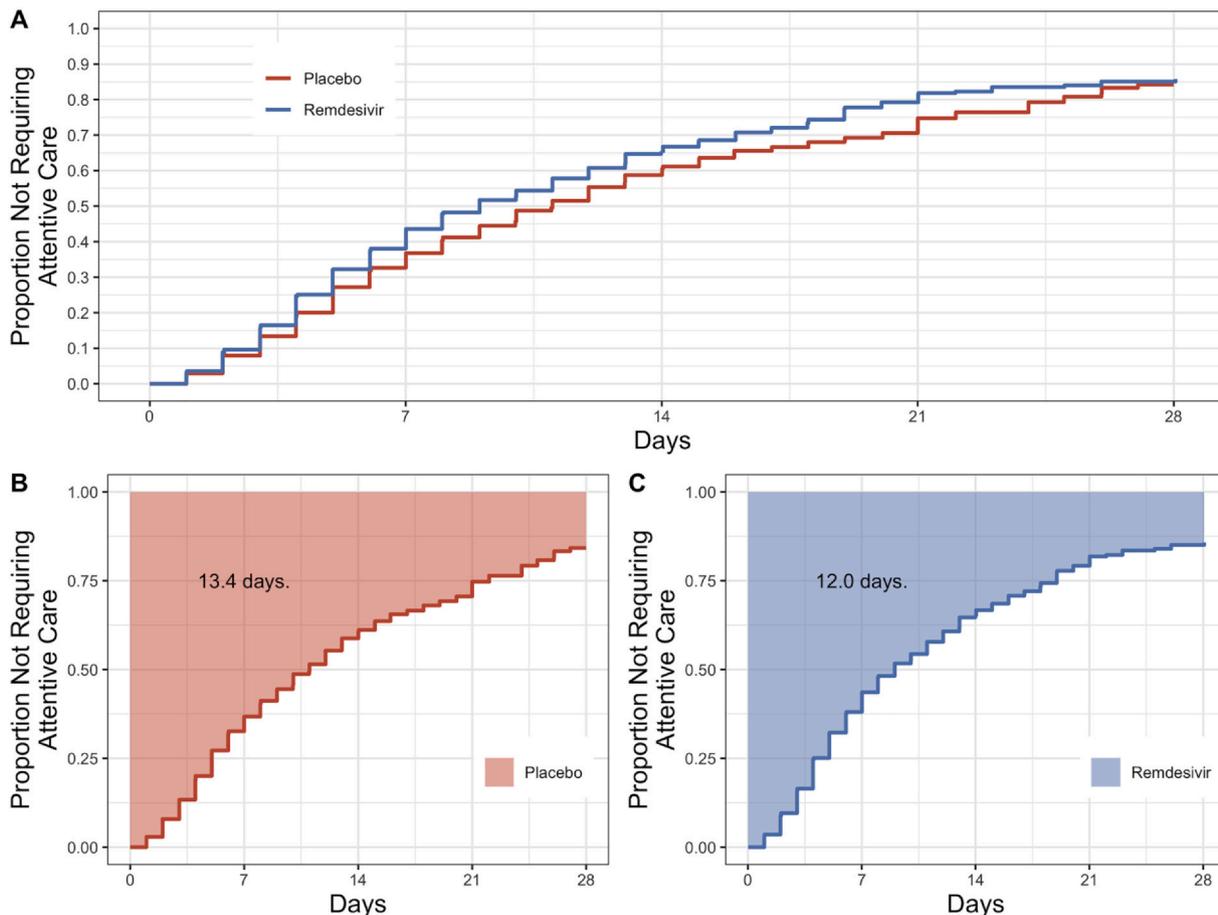


Fig. 2. (A) Cumulative incidence curves for the proportion of patients not requiring attentive care. (B & C) Mean time spent in hospital receiving attentive care via the area above the cumulative incidence curve.

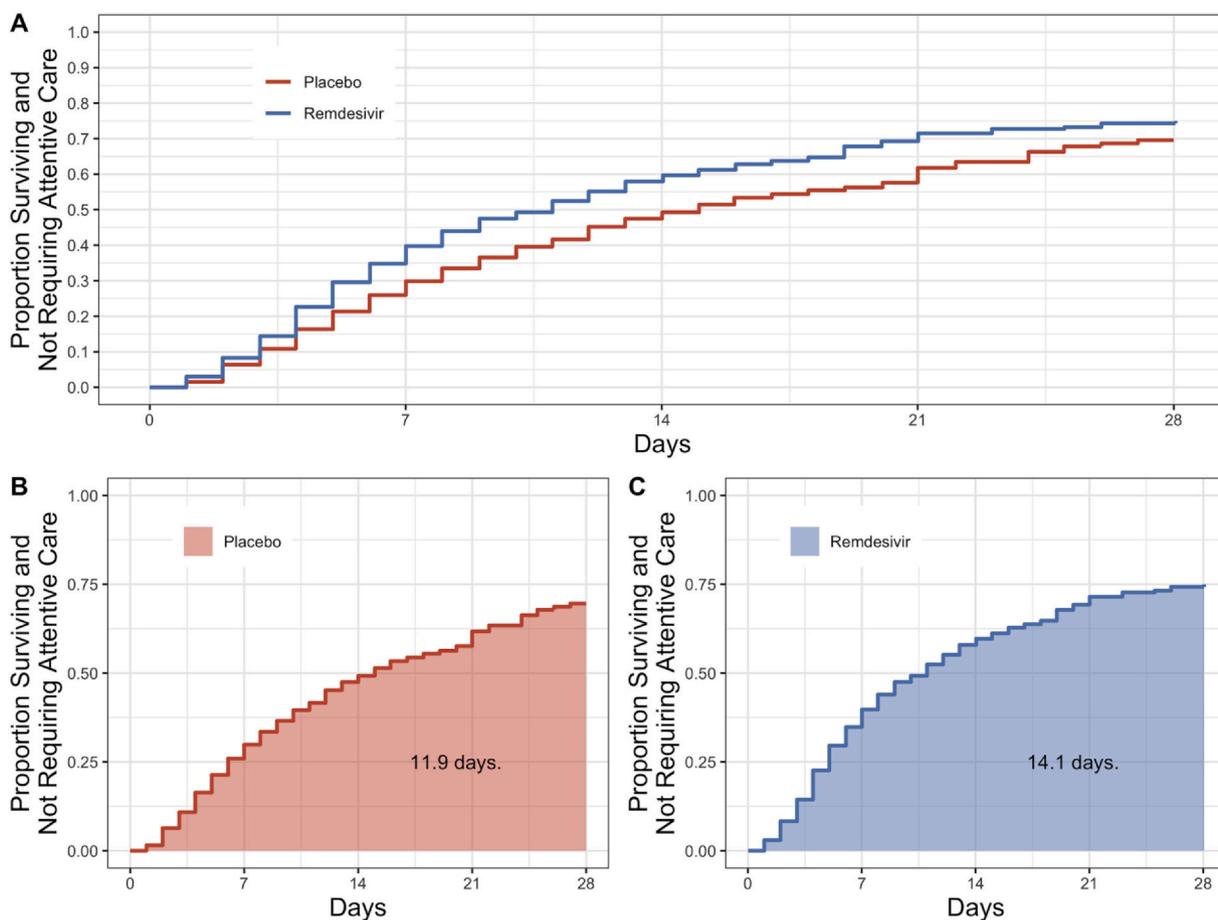


Fig. 3. (A) Cumulative incidence curves for the proportion of patients surviving and not requiring attentive care. (B & C) Mean event-free survival as the area under the cumulative incidence curve.

care arm, this analysis penalizes remdesivir.

Now, to analyze the event-free survival times across the 28 days of follow-up, Fig. 3A presents the cumulative incidence curves for the proportion of patients surviving and not receiving attentive care. In contrast to Fig. 2A, these curves are constructed while treating death as a competing risk [11–13]. That is, patients who die were removed from the risk set and no longer eligible to experience recovery. The areas below these cumulative incidence curves (Figs. 3B & C) represent the mean event-free survival times up to day 28 [14]. These were 11.9 days for placebo, and 14.2 days for remdesivir. The difference was 2.2 days (95% CI, 0.89 to 3.52 days; $P < 0.001$) in favor of remdesivir. Thus, on average, patients receiving remdesivir survived event-free for an additional 2.2 days across the 28 days of follow-up. Note that the difference in mean event-free survival is highly statistically significant in favor of remdesivir in contrast to the moderately significant difference obtained using the duration of the need for medical attention. This may reflect the fact that remdesivir numerically prolonged patients' survival [14].

3. Discussions

Shortening the duration of hospitalization is not equivalent to extending the duration of hospital-free survival. A harmful treatment may shorten the duration of hospitalization by increasing the mortality rate without increasing the recovery rate. Such a treatment would reduce the duration of hospital-free survival. Conversely, a beneficial treatment may extend the duration of hospital-free either by reducing the mortality rate or by increasing the recovery rate. Only in the absence of death is reducing the hospital stay equivalent to extending the duration of hospital-free survival. However, if the mortality rates of the two arms

are similar, then comparing the durations of hospitalization may be appropriate.

When dealing with, for example, the duration of ICU stay or oxygen utilization, the data are more complex than those from the endpoint discussed above since patients may enter the ICU during the study. This complexity extends to other cases where initiation of the event of interest does not necessarily start when the patient enters the study [15]. Fig. 4 provides various possible patterns with respect to the ICU endpoint. For example, in case 4, the patient was in ICU from day 7 to 14. For case 5, the patient died during the ICU stay. The ICU-free survival days are given on the right-hand side of Fig. 4. The complexity of such data arises because entry into the ICU is likely dependent on the

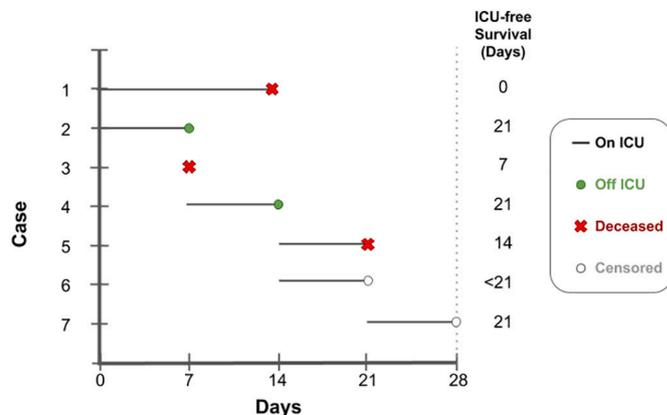


Fig. 4. Possible patterns for the duration of ICU stay.

patient's eventual outcome, such as hospital discharge or death. Therefore, like case 6, the censoring was not independent of the ICU-free survival time and one cannot analyze the duration of ICU stay or ICU-free survival time via standard survival analysis methods. Further statistical research is needed to delineate the analytic procedures needed to obtain valid inference on the average ICU-free survival time.

Note that if there are no censored observations before Day 28, then the analysis of ICU-stay is straightforward. In future trials for COVID-19, the trial size may be larger and the length of study follow-up may be longer. Therefore, at the interim analysis, there are likely to be a sizable number of censored observations, and one is likely to encounter the same challenges faced in the ACTT-1 trial.

In conclusion, the conventional approach to analyzing the duration of an undesirable clinical event may not be appropriate. On the other hand, a simple conversion of this endpoint to the event-free survival counterpart is clinically interpretable and leads to a fair assessment of treatment efficacy.

References

- [1] J.H. Beigel, K.M. Tomashek, L.E. Dodd, et al., Remdesivir for the treatment of COVID-19 – a preliminary report, *NEJM* (2020 May 22), <https://doi.org/10.1056/NEJMoa2007764>.
- [2] J.D. Goldman, D.C.B. Lye, D.S. Hui, et al., Remdesivir for 5 or 10 days in patients with severe COVID-19, *NEJM* (2020 May 27), <https://doi.org/10.1056/NEJMoa2015301>.
- [3] B. Cao, Y. Wang, D. Wen, et al., A trial of Lopinavir-ritonavir in adults hospitalized with severe COVID-19, *NEJM* 382 (19) (2020 May 7) 1787–1799, <https://doi.org/10.1056/NEJMoa2001282>.
- [4] Y. Wang, D. Zhang, G. Du, et al., Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial, *Lancet* 395 (10236) (2020 May 16) 1569–1578, [https://doi.org/10.1016/S0140-6736\(20\)31022-9](https://doi.org/10.1016/S0140-6736(20)31022-9).
- [5] https://www.nejm.org/doi/suppl/10.1056/NEJMoa2007764/suppl_file/nejmoea2007764_protocol.pdf.
- [6] P. Guyot, A.E. Ades, M.J. Ouwens, et al., Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves, *BMC Med. Res. Methodol.* 12 (2012) 9, <https://doi.org/10.1186/1471-2288-12-9>.
- [7] H. Uno, B. Claggett, L. Tian, et al., Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis, *J. Clin. Oncol.* 32 (22) (2014) 2380–2385.
- [8] K. Pak, H. Uno, D.H. Kim, et al., Interpretability of cancer clinical trial results using restricted mean survival time as an alternative to the hazard ratio, *JAMA Oncol.* 3 (12) (2017) 1692–1696.
- [9] D.H. Kim, H. Uno, L.J. Wei, Restricted mean survival time as a measure to interpret clinical trial results, *JAMA Cardiol.* 2 (11) (2017) 1179–1180.
- [10] Z.R. McCaw, A.R. Orkaby, L.J. Wei, et al., Applying evidence-based medicine to shared decision making: value of restricted mean survival time, *Am. J. Med.* 132 (1) (2019) 13–15, <https://doi.org/10.1016/j.amjmed.2018.07.026>.
- [11] P.C. Austin, D.S. Lee, J.P. Fine, Introduction to the analysis of survival data in the presence of competing risks, *Circulation* 133 (6) (2016 Feb 9) 601–609, <https://doi.org/10.1161/CIRCULATIONAHA.115.017719>.
- [12] B. Lau, S.R. Cole, S.J. Gange, Competing risk regression models for epidemiologic data, *Am. J. Epidemiol.* 170 (2) (2009) 244–256, <https://doi.org/10.1093/aje/kwp107>.
- [13] L. Zhao, L. Tian, B. Claggett, et al., Estimating treatment effect with clinical interpretation from a comparative clinical trial with an end point subject to competing risks, *JAMA Cardiol.* 3 (4) (2018 Apr 1) 357–358, <https://doi.org/10.1001/jamacardio.2018.0127>.
- [14] Z.R. McCaw, L. Tian, J.L. Vassay, et al., How to quantify and interpret treatment effects in comparative clinical studies of COVID-19, *Ann. Intern. Med.* (2020), <https://doi.org/10.7326/M20-4044>.
- [15] B. Huang, L. Tian, Z.R. McCaw, et al., Analysis of response data for assessing treatment effects in comparative clinical studies, *Ann. Intern. Med.* (2020), <https://doi.org/10.7326/M20-0104>.