Open Forum Infectious Diseases

REVIEW ARTICLE

Overlooked Shortcomings of Observational Studies of Interventions in Coronavirus Disease 2019: An Illustrated Review for the Clinician

Imad M. Tleyjeh,[1,2,3,4] Tarek Kashour,[5] Jay Mandrekar,[6] and Diana B. Petitti[7]

[1]Infectious Diseases Section, Department of Medical Specialties, King Fahad Medical City, Riyadh, Saudi Arabia, [2]Division of Infectious Diseases, Mayo Clinic College of Medicine and Science, Rochester, Minnesota, USA, [3]Division of Epidemiology, Mayo Clinic College of Medicine and Science, Rochester, Minnesota, USA, [4]College of Medicine, Alfaisal University, Riyadh, Saudi Arabia, [5]Department of Cardiac Sciences, King Fahad Cardiac Center, King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia, [6]Department of Biostatistics, Mayo Clinic College of Medicine and Science, Rochester, Minnesota, USA, and [7]Department of Biomedical Informatics, University of Arizona College of Medicine, Phoenix, Arizona, USA

The rapid spread of severe acute respiratory syndrome coronavirus 2 infection across the globe triggered an unprecedented increase in research activities that resulted in an astronomical publication output of observational studies. However, most studies failed to apply fully the necessary methodological techniques that systematically deal with different biases and confounding, which not only limits their scientific merit but may result in harm through misleading information. In this article, we address a few important biases that can seriously threaten the validity of observational studies of coronavirus disease 2019 (COVID-19). We focus on treatment selection bias due to patients' preference on goals of care, medical futility and disability bias, survivor bias, competing risks, and the misuse of propensity score analysis. We attempt to raise awareness and to help readers assess shortcomings of observational studies of interventions in COVID-19.

Keywords. bias; confounding; observational studies.

The 2 fundamental variables of an analytical study are the exposure and the outcome. In observational studies, the frequency of an outcome or exposure is measured and compared between groups. This comparison yields relative frequency measures and describes the association between the exposure and the outcome [1]. Although observational studies complement randomized controlled trials (RCTs), they are prone to certain flaws that can seriously threaten their validity: confounding and bias.

Confounding occurs when the observed relationship between the exposure and the outcome is altered or is accounted for by a third variable (the confounder). Confounders fulfill 3 criteria: (1) they are related to both the exposure and the outcome; (2) they are distributed unequally between the studied groups; and (3) they serve as intermediate step in the causal pathway between the exposure and the outcome [2]. Identifying and taking into account possible confounders is critical when drawing causal inferences from observational studies, especially those assessing treatment effects.

Bias is a systematic error in any process at any stage of inference that produces results or conclusions that differ systematically from the truth [3]. Numerous types of biases that can creep into different stages of clinical research have been recognized and catalogued [3]. Some of these are well known and easily recognizable. Others are frequently missed.

The rapid spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection across the globe and its associated huge burden of morbidity and mortality triggered an unprecedented increase in research activities. Early observations of high mortality in patients hospitalized with coronavirus disease 2019 (COVID-19) led to the use of a variety of pharmacologic treatments based on in vitro studies and/or extrapolation from the effect of treatments on what was considered to be the underlying pathophysiology of COVID-19. A result was an astronomical output of publications of observational studies examining outcomes of treatment in patients with COVID-19. Most of these studies failed to use methodological approaches that systematically deal with biases that arise when examining mortality in severely ill patients and with confounding due to treatment selection. This has not only compromised the validity of these studies and limited their scientific merit but may have resulted in harm through misleading information. For example, hydroxychloroquine (HCQ), convalescent plasma, vitamin C and D, zinc, azithromycin, and recently ivermectin have all

Received 1 June 2021; editorial decision 5 June 2021; accepted 9 June 2021.

Correspondence: Imad M. Tleyjeh, MD, MSc, FACP, FIDSA, Section of Infectious Diseases, King Fahd Medical City, PO Box 59046, Riyadh 11525, Saudi Arabia (tleyjeh.imad@mayo.edu).

Open Forum Infectious Diseases®2021

https://doi.org/10.1093/ofid/ofab317

been promoted and used based on observational studies, the observed effects of which were not subsequently found in RCTs.

In this article, we address a few important biases that can seriously threaten the validity of observational studies of the effect on mortality of treatments for COVID-19 and discuss issues related to the propensity score methods, which have been widely used in observational research about treatments for COVID-19 [1, 4]. We attempt to raise awareness of typical and avoidable biases and to help readers assess the direction and magnitude of bias in published observational studies of COVID-19 treatments to date.

TREATMENT SELECTION BIAS

Overview

Bias in an epidemiologic study refers to a systematic error that leads to an incorrect estimate of the true effect of an exposure on an outcome [3]. A major concern in observational studies of COVID-19 treatment and mortality outcome is bias that arises from selection of treatment of patients based on factors that predict outcome. For example, patients who have advanced cancer or end-stage comorbid diseases or those with disability may not be offered a given treatment (medical futility and disability bias).

Patients' Preferences, Medical Futility, and Disability as Factors in Treatment Selection

In the early months of the SARS-CoV-2 pandemic and in the absence of a standard protocol in a real-world setting outside the context of a randomized trial, treatment might be affected by whether a patient had a do not resuscitate (DNR) or do not ventilate (DNV) advance directive at admission or the patient and/or family decided after admission that the patient should not be resuscitated and/or ventilated. Moreover, during pandemic waves when health care systems are overwhelmed, patients with disabilities or those whose care has been judged as futile are less likely to receive advanced critical care [5–7]. These patients are more likely to die because the reason for the DNR and/or DNV directive is usually poor health status. If these patients are less likely to be treated and are included in the comparator group in the analysis, the estimated effect of treatment on mortality will be biased in favor of the treatment. For example, in a non–COVID-19 observational study of patients admitted to an intensive care unit with sepsis, adding DNR status to a multivariable model assessing the association between activated protein C and mortality led to an important shift in the estimated effect of activated protein C on mortality [8].

DNR and DNV orders are a specific category of a larger set of statements that are patients' preferences for life-sustaining treatments. Walkey et al [9] have discussed the importance of accounting for patient preferences for life-sustaining treatment both in observational studies of treatments and in clinical trials.

There is suggestive evidence that patients hospitalized with COVID-19 who had preferences to forgo or were not offered mechanical ventilation were less likely to be treated with HCQ, for example, at least in some settings (Table 1). They are also much more likely to die. In these observational studies, the bias in the estimated effect of HCQ on mortality because of failure

Table 1. Comparison of 3 Large Coronavirus Disease 2019 Cohorts With a Focus on "Goal of Care"

Study	Data Set		
Arshad et al [10]	Henry Ford Health System (6 hospitals) in southeast Michigan: 10 March–2 May 2020		
Variable	HCQ Group (n = 1985)	Non-HCQ (n = 556)	Comment
ICU admission	26.9%	14.6%	166/615 (27%) of those who died did not receive MV
Ventilatory support	20.2%	8.6%	
Mortality	16.1%	25.4%	
Richardson et al [11]	Northwell Health System (12 hospitals) in New York City, Long Island, and Westchester County, New York: 1 March–4 April 2020		
Variable	HCQ Group	Non-HCQ	271/553 (49%) of those who died did not receive MV
ICU admission	NA	NA	
Ventilatory support	NA	NA	
Mortality	NA	NA	
Catteau et al [12]	Belgian National COVID-19 Hospital Surveillance Data: 14 March–24 May 2020		
Variable	HCQ Group (n = 4542)	Non-HCQ (n = 3533)	1512/1881 (80.4%) of patients who died did not receive MV 861/975 (88%) of non-HCQ patients who died did not receive MV 300/800 (37.5%) of HCQ patients who died did not receive MV
ICU admission within 24 hours	6.9%	2.7%	
Ventilatory support	11.4%	3.3%	
Mortality	17.7%	27.1%	

Abbreviations: COVID-19, coronavirus disease 2019; HCQ, hydroxychloroquine; ICU, intensive care unit; MV, mechanical ventilation; NA, not applicable (did not examine HCQ); NR, not reported.

to account for a preference not to be ventilated would be large (Table 1).

A 2020 systematic review and meta-analysis [13] of observational studies of HCQ and mortality in patients hospitalized for COVID-19 identified 15 studies that were not considered "critically biased." None of the 15 studies presented information on whether patients with DNR or DNV advance directives were admitted to the hospital, whether these patients were included in the study, and, if included, how they were handled in the analysis.

SURVIVOR BIAS

Treatments including medications (eg, antivirals, immunomodulators, anticoagulation) or other interventions (eg, respiratory support, prone positioning) are being widely tested using cohort studies of patients hospitalized for COVID-19. In observational studies, "immortal time" refers to the period between the time point when patients enter the study cohort (in most COVID-19 studies this is the admission time) and the point when they receive the examined treatment. During the period between admission and treatment initiation, death cannot occur in the treatment group because those patients must, by design, survive long enough to receive treatment. In other words, the patients who survive to receive treatment are considered "immortal" between admission and treatment (Figure 1). Outcomes, such as death, that occur among patients during this immortal time can be attributed to only the "untreated/comparator" group. Not accounting for this immortal time in the design or analysis of observational studies leads to what is known as immortal time bias, survivor bias, or time-dependent bias [14].

Survivor bias is common in the medical literature including in studies published in high-impact-factor journals. In 2004, van Walraven et al [15] examined all observational studies that used a survival analysis in top medical journals between 1998 and 2004. Of 682 eligible studies, 127 (18.6%) contained a time-dependent factor, of which 52 (40.9% of studies with a time-dependent factor) were susceptible to survivor bias. In approximately two-thirds (67.3%) of these susceptible studies, the bias affected a variable mentioned in the study abstract. Correction of the bias could have qualitatively changed the study's conclusions in more than one-half of the studies [15].

Beyersmann et al [16] showed, using a simple mathematical tool, that survivor bias inevitably leads to biased effect estimation, because the number of individuals at risk of exposure is distorted over the course of time. Beyersmann and colleagues' [16] model showed 3 possible effects of survivor bias on time to study end point (outcome). First, if the time-dependent exposure has no real effect on the outcome, survivor bias will result into erroneous positive association (better) with the outcome. Second, if the time-dependent exposure has a real negative (worse) effect on the outcome, survivor bias will exaggerate this negative association. Third, if the time-dependent exposure has a real positive (better) effect on the outcome, time-dependent bias will show less pronounced effect.

The Cox proportional hazards regression model is commonly used to analyze data from studies that seek to estimate the effect of treatment on the time to an event (eg, death). A Cox regression analysis estimates the hazard rate—the probability of having the event (eg, death) given that the patient has survived to a specific time. The pure effect of survivor bias on the outcome hazard is as follows: The hazard for the untreated patients is always overestimated, and the hazard for the treated patients is always underestimated. Thus, the hazard ratio (HR) (comparing treated with untreated) is always underestimated. As a rule of thumb, the magnitude of the bias depends on 2 components: first, the time to the exposure (the longer the time to exposure, the larger the bias), and second, the time to the outcome (the shorter the time, the stronger the bias) (Figure 1).

The following analytical approaches are used to account for survivor bias [17, 18]: (1) model treatment as a time-dependent variable in the Cox regression analysis; (2) landmark analysis; (3) structural nested accelerated failure time model; (4) and marginal structure models. The advantages and disadvantages of these approaches are discussed in statistical papers and are beyond the scope of this review.

In 2 recent systematic reviews [19, 20], we observed that only 4 of 18 cohort studies of tocilizumab in COVID-19 and 3 of 19 studies of corticosteroids in COVID-19 adjusted for survivor bias. In one study by Wu et al [21] of 1514 COVID-19 patients, the authors reported on the adjusted HR for mortality for corticosteroid treatment. With corticosteroid treatment considered as a time-fixed variable in the Cox regression analysis, the adjusted HR for mortality in severe COVID-19 was 1.77 (95% confidence interval [CI], 1.08–2.89) while it increased by 60% to 2.83 (95% CI, 1.72–4.64) when corticosteroid treatment was used as a time-dependent variable.

COMPETING RISKS

In survival data, the outcome of interest is time to the occurrence of a certain event. An important feature of survival data is censoring, which occurs when the exact survival time is unknown [22]. Censoring occurs, for example, when a subject is lost to follow-up, withdraws from the study, or does not experience the event of interest before the end of the study. Conventional methods used in the analysis of survival data include the Kaplan-Meier method and Cox proportional hazards regression. A Cox regression analysis estimates the hazard rate—the probability of having the event (eg, death) given that the patient had survived to a specific time [22]. Both methods assume that censoring is independent or noninformative. Noninformative censoring means that individuals who are censored have the

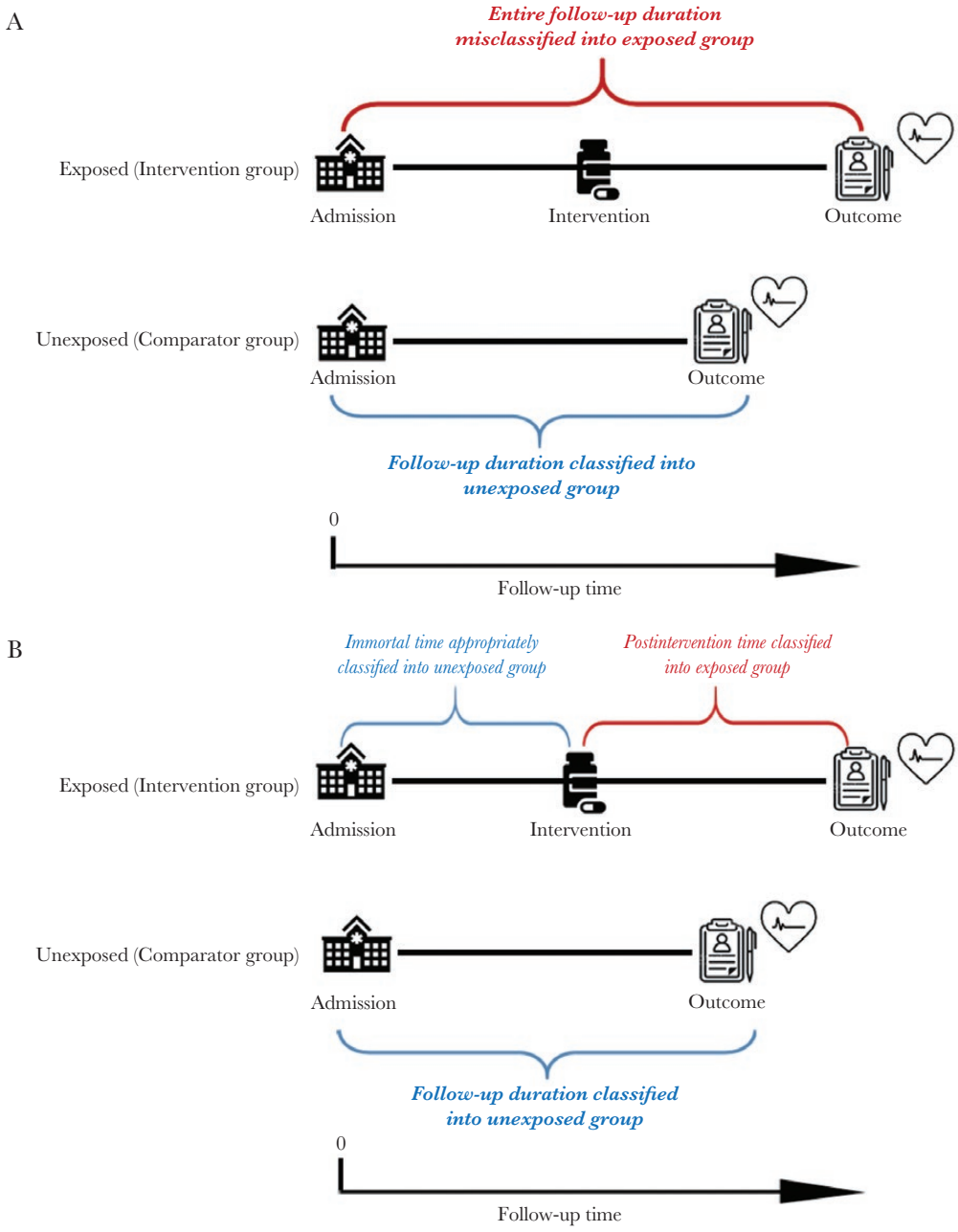

Figure 1. *A*, Illustration of survivor bias due to misclassification of preintervention immortal time (time between admission to intervention). *B*, Illustration of appropriate classification of preintervention immortal time: absence of survivor bias.

same future risk of the event of interest as subjects under observation. However, this assumption is not always true.

A common approach used by investigators is to examine COVID-19 patients' hospital mortality data without any follow-up beyond hospital discharge. In this scenario, discharged patients are treated as censored observations when using survival analysis. The fundamental assumption that the death hazard remains the same after censoring is violated here; discharged patients have usually recovered and thus have lower death hazards than patients who remain hospitalized. If no follow-up beyond hospital discharge is available, discharge from hospital and in-hospital death are therefore considered "competing events." A competing event is an event whose occurrence precludes the occurrence of the event of interest and the incidence of these events is called "competing risks" [23]. The same principle applies when, for example, recovery from COVID-19 is being considered as the primary outcome. Patients who die are censored and are therefore wrongly assumed to have similar risk of recovery compared to those who remain alive and hospitalized.

Competing risk bias is common in studies, even those published in top medical journals. In 2016, Schumacher et al [24] assessed 219 original articles published in the *New England Journal of Medicine* in 2015. They identified 192 (88%) publications with a time-related primary end point, of which 136 studies (62%) used statistical methodology for time-to-event data. In 51 of the 136 studies, competing risks were present. The competing risks were adequately dealt with in only 26 of these 51 studies (51%). The remaining 25 studies (49%) were susceptible to competing risk bias.

Competing events affect our ability to appropriately compare risks for a given treatment in an observational study. For example, we might be interested to know whether convalescent plasma transfusion decreases the risk of in-hospital mortality in COVID-19 patients. If convalescent plasma is associated with the competing event (discharge), then this can have a large effect on the estimated risk of in-hospital mortality in patients treated with plasma transfusion. Table 2 summarizes a few examples from high-impact journals of studies of the effect of treatment on mortality that failed to consider competing events in their analysis.

Failing to account for competing risks generally leads to an overestimation of the cumulative incidence of the event of interest. The extent to which the cumulative incidence is overestimated is related to the proportion of subjects experiencing the event of interest and the competing event. In a recent study [29], investigators simulated a fictive clinical trial on COVID-19 mimicking studies investigating interventions such as hydroxychloroquine, remdesivir, or convalescent plasma. The outcome was time from randomization until in-hospital death. Six scenarios for the effect of treatment on death and recovery were considered. The HR and the 28-day absolute risk reduction of in-hospital death were estimated using the Cox proportion hazards and the Fine and Gray models [30]. Estimates were then compared with the true values, and the magnitude of misestimation was quantified. The simulation showed that the shorter the median time to recovery (as a competing event with death), the more overestimated the association between treatment and in-hospital mortality (ie, the more perceived benefit of treatment on survival).

Analysis of survival data with competing events requires special considerations [31, 32]. Methods available to correctly analyze these data include estimating the risks of events over time and determining how exposures of interest affect risk. Depending on the research question, in the presence of competing events, survival data should be analyzed using either a cause-specific hazard model or a subdistribution hazard model (Figure 2) [30]. The cumulative incidence function (CIF) should be used to estimate the cumulative incidence instead of the Kaplan-Meier method. Cumulative incidence is defined as the probability that a particular event has occurred before a given time. The CIF quantifies the cumulative probability of cause-specific failure in the presence of competing events without assumptions about the dependence among the events. It denotes the probability of experiencing a specified event before time (t) and before the occurrence of a different type of event.

CAUSAL INFERENCE

Overview

Over the last decade, there has been increasing interest in the use of real-world data in the evaluation of treatments of all kinds. This interest has arisen in parallel with growing use of electronic health records (EHRs) in health care settings and concomitant development of large databases of information derived from EHRs, often linked with data from other data sources.

The challenges in using real-world data to draw conclusions about the causal effects of treatments on outcomes are numerous. Statistical methods that seek to enhance the ability to draw causal inferences from observational studies of treatment have evolved rapidly. Propensity methods have come to be used with increasing frequency in studies that seek to draw causal inferences about the effect of treatment on outcomes using observational data, although many other methods have been proposed to deal with confounding. These include, among others, inverse probability weighting, covariate balancing techniques, and machine learning.

Table 2. Selected Examples of Coronavirus Disease 2019 Observational Studies From High-Impact Journals With Competing Events

Study	Intervention	Outcome	Competing Event	Survival Analysis	Discharged Patients Censored	Competing Risk Analysis
Biran et al [25]	Tocilizumab	In-hospital mortality	Discharge alive	Yes	Yes	No
Geleris et al [26]	HCQ	Composite of intubation or death	Discharge alive before intubation	Yes	Yes	No
Huet et al [27]	Anakinra	Composite of intubation or death	Discharge alive before intubation	Yes	Yes	No
Rosenberg et al [28]	HCQ with or without azithromycin	In-hospital mortality	Discharge alive	Yes	Yes	No

Abbreviation: HCQ, hydroxychloroquine.

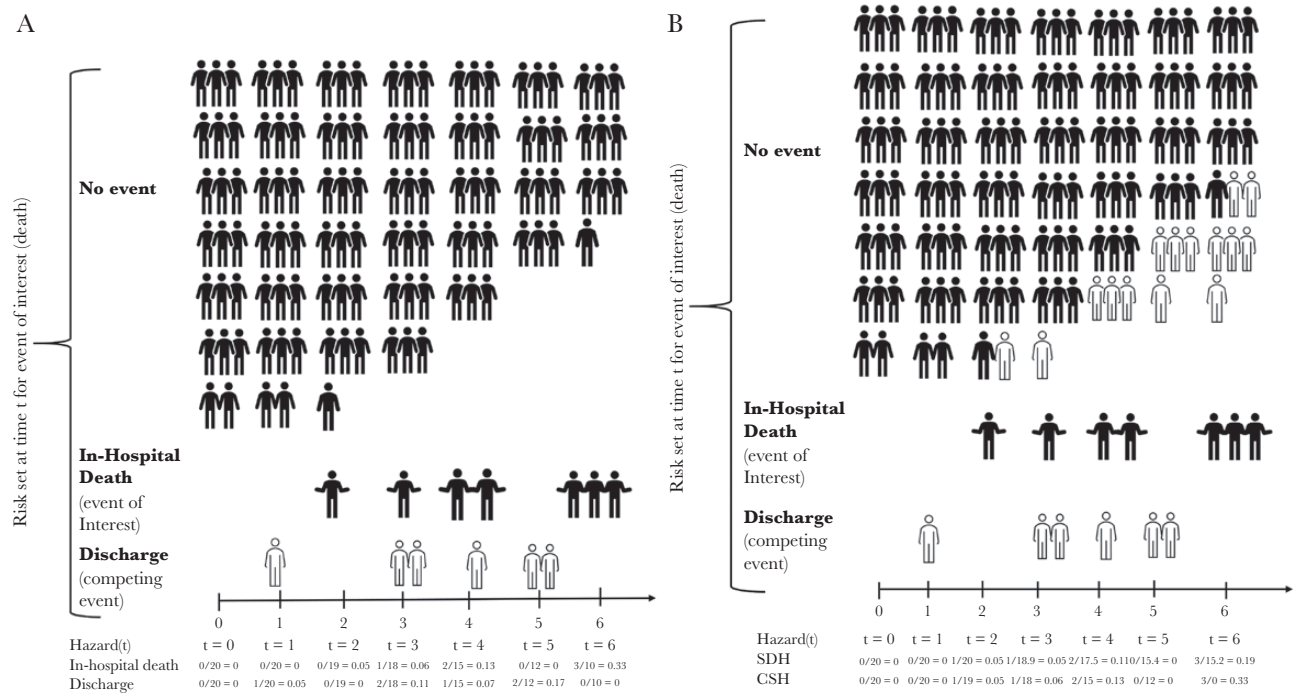

Figure 2. Modified from Sapir-Pichhadze et al [23]. *A*, Cause-specific hazard ratio (HR) for each event: the HR for death (the event of interest) and the HR for discharge (the competing event). These can be obtained from separate Cox proportional hazards regression models. This approach provides an etiological exploration of risk factors and shows how risk factors are associated with each event; direct and indirect effects can be distinguished. *B*, The subdistribution hazard function estimates the hazard rate for death at time t based on the risk set that remains at time t after accounting for all previously occurring event types, which includes competing events (death and discharge). As a time-averaged risk comparison, subdistribution HRs extend overall risk ratios. Abbreviations: CSH, cause-specific hazard; SDH, subdistribution hazard.

Propensity Score Methods

A concern about conventional regression analysis is that the regression model may be overfitted when the number of covariates is large compared with the number of outcome events. In observational studies of COVID-19 treatments and mortality, the number of factors that are potential confounders or are related to treatment selection is large. Thus, model overfitting is a valid concern. Propensity score methods theoretically reduce the problem of model overfitting because they can balance a large number of covariates across treatment groups by weighting on a single score [33, 34].

Propensity score matching has been particularly popular in the medical literature [35]. As described in detail in the next section, observational studies of COVID-19 treatments in hospitalized patients that used propensity methods have most often used propensity score matching [19, 20].

Propensity score matching has the advantage of being simple to present and interpret. However, this simplicity hides the complexity of implementing "matching on the propensity score" [36]. Many decisions must be made, including whether to do pair-matching (1-to-1) or 1-to-many matching and the choice among many matching methods (eg, nearest neighbor, caliper matching, radius matching, kernel matching). A decision must be made about whether to match with or without replacement. These choices can affect conclusions and there is no definitive advice regarding how and when to choose a certain technique [36].

Propensity score matching can exclude patients for whom no matching patient exists. For example, in a study of HCQ and mortality that used 1-to-1 propensity score matching, only 1820 patients (910 pairs of HCQ treated/HCQ untreated) from a pool of 3372 eligible patients with COVID-19 contributed to the final analysis [37]. Exclusion of some patients because they have no matching patients is a disadvantage of propensity score matching because precision and generalizability are reduced [38].

Propensity Scores in Observational Studies of Treatment for COVID-19: Corticosteroids and Tocilizumab as Case Studies

In a systematic review [19] of corticosteroid outcomes in patients with COVID-19, 9 cohort studies examining mortality in relation to corticosteroid treatment were identified. Of these, 3 (33%) used a propensity method to control for confounding. Two studies used propensity score matching. One study did not provide details about the matching ratio, the methods for matching, or the software used. Another study provided information on the matching ratio (pairs), stated that matching was done without replacement, and identified the software package used but did not state the method for matching.

In another systematic review of tocilizumab outcomes in patients with COVID-19 [20], 18 observational studies examining mortality met the authors' criteria for inclusion. Of these, 7 (39%) used propensity scores as an approach to control for confounding. As found in other evaluations of the implementation of propensity score methods, the completeness of the description of the methods was uneven. Five of the 7 studies that used a propensity score method used propensity score matching. Only 2 of the 5 provided information on all of the following: the matching ratio, the method for matching, whether the matching was done with or without replacement, the caliper, and the software used. One study that used propensity score matching provided information only on the matching ratio.

Limitations of Propensity Scores and Regression to Address Bias and Confounding

Neither propensity score methods nor conventional regression can disentangle the causal effect of one treatment (eg, remdesivir) from the causal effects of other medications (eg, steroids) when the same patient receives both treatments. Neither propensity score methods nor conventional regression can assess and balance all the many factors that come into play in the clinical management of seriously ill patients during the course of hospital treatment [39]. With both conventional regression analysis and propensity score methods, the validity of conclusions about a causal effect of treatment is based on the assumption of absence of unobserved treatment selection factors and residual confounding.

FINAL THOUGHTS

Although an RCT is considered the gold standard to test the efficacy of any intervention, data from well-designed observational studies complement RCTs [40–42]. Observational studies provide important foundation data to plan RCTs such as hypothesis generation and RCT sample size calculation. Moreover, observational studies help examine the generalizability of RCT findings. As compared to observational studies, RCTs include standardized patient care with protocols and exclude certain patient groups. Moreover, observational studies and meta-analyses of these studies may offer higher external validity than a single RCT owing to their potentially large size and the ability to include a patient sample that is representative of the average patient population. A recent example is the concordance between the efficacy and effectiveness of COVID-19 vaccines in RCTs and observational studies, respectively.

Finally, in addition to the rigorous conduct of observational studies, investigators need to comply with the guidelines for reporting of observational studies (Strengthening the Reporting of Observational Studies in Epidemiology [STROBE] statement) [43]. The STROBE statement serves as a common and important construct to report observational research in a standardized and rigorous manner.

CONCLUSIONS

Observational studies can provide valuable information that helps us understand diseases like COVID-19 and explore potential benefits of different therapeutic interventions. However, they are prone to bias and confounding. In this review, we discussed treatment selection bias, survivor bias, competing risks, and the misuse of propensity score analysis. We attempted to raise awareness and to help readers assess shortcomings of observational studies of interventions in COVID-19.

Notes

Potential conflicts of interest. All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

REFERENCES

1. Jepsen P, Johnsen SP, Gillman MW, Sørensen HT. Interpretation of observational studies. Heart **2004**; 90:956–60.
2. Porta M, Porta M, ed. A Dictionary of Epidemiology. 6th ed. New York: Oxford University Press; **2014**.
3. Sackett DL. Bias in analytic research. J Chronic Dis **1979**; 32:51–63.
4. Nørgaard M, Ehrenstein V, Vandenbroucke JP. Confounding in observational studies based on large health care databases: problems and potential solutions—a primer for the clinician. Clin Epidemiol **2017**; 9:185–93.
5. Dyer C. Many "do not resuscitate" decisions in England during pandemic ignored equality and human rights laws, says CQC. BMJ **2021**; 372:n754.
6. Tleyjeh PIM, Tlayjeh H. Perceived efficacy of hydroxychloroquine in observational studies: results of the confounding effect of "goals of care." Int J Antimicrob Agents **2021**; 57:106308.
7. Andrews EE, Ayers KB, Brown KS, et al. No body is expendable: medical rationing and disability justice during the COVID-19 pandemic. Am Psychol **2021**; 76:451–61.
8. Bradford MA, Lindenauer PK, Wiener RS, Walkey AJ. Do-not-resuscitate status and observational comparative effectiveness research in patients with septic shock. Crit Care Med **2014**; 42:2042–7.
9. Walkey AJ, Barnato AE, Wiener RS, Nallamothu BK. Accounting for patient preferences regarding life-sustaining treatment in evaluations of medical effectiveness and quality. Am J Respir Crit Care Med **2017**; 196:958–63.
10. Arshad S, Kilgore P, Chaudhry ZS, et al; Henry Ford COVID-19 Task Force. Treatment with hydroxychloroquine, azithromycin, and combination in patients hospitalized with COVID-19. Int J Infect Dis **2020**; 97:396–403.
11. Richardson S, Hirsch JS, Narasimhan M, et al; Northwell COVID-19 Research Consortium. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA **2020**; 323:2052–9.
12. Catteau L, Dauby N, Montourcy M, et al; Belgian Collaborative Group on COVID-19 Hospital Surveillance. Low-dose hydroxychloroquine therapy and mortality in hospitalised patients with COVID-19: a nationwide observational study of 8075 participants. Int J Antimicrob Agents **2020**; 56:106144.
13. Fiolet T, Guihur A, Rebeaud ME, Mulot M, Peiffer-Smadja N, Mahamat-Saleh Y. Effect of hydroxychloroquine with or without azithromycin on the mortality of coronavirus disease 2019 (COVID-19) patients: a systematic review and meta-analysis. Clin Microbiol Infect **2021**; 27:19–27.
14. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. BMJ **2010**; 340:b5087.
15. van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. J Clin Epidemiol **2004**; 57:672–82.
16. Beyersmann J, Gastmeier P, Wolkewitz M, Schumacher M. An easy mathematical proof showed that time-dependent bias inevitably leads to biased effect estimation. J Clin Epidemiol **2008**; 61:1216–21.
17. Karim ME, Gustafson P, Petkau J, Tremlett H; Long-Term Benefits and Adverse Effects of Beta-Interferon for Multiple Sclerosis (BeAMS) Study Group.

Comparison of statistical approaches for dealing with immortal time bias in drug effectiveness studies. Am J Epidemiol **2016**; 184:325–35.
18. Zhou Z, Rahme E, Abrahamowicz M, Pilote L. Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: a comparison of methods. Am J Epidemiol **2005**; 162:1016–23.
19. Tlayjeh H, Mhish OH, Enani MA, et al. Association of corticosteroids use and outcomes in COVID-19 patients: a systematic review and meta-analysis. J Infect Public Health **2020**; 13:1652–63.
20. Tleyjeh IM, Kashour Z, Damlaj M, et al. Efficacy and safety of tocilizumab in COVID-19 patients: a living systematic review and meta-analysis. Clin Microbiol Infect **2021**; 27:215–27.
21. Wu J, Huang J, Zhu G, et al. Systemic corticosteroids show no benefit in severe and critical COVID-19 patients in Wuhan, China: a retrospective cohort study. medRxiv [Preprint]. Posted online 14 May **2020**. doi:10.1101/2020.05.11.20097709.
22. Bewick V, Cheek L, Ball J. Statistics review 12: survival analysis. Crit Care **2004**; 8:389–94.
23. Sapir-Pichhadze R, Pintilie M, Tinckam KJ, et al. Survival analysis in the presence of competing risks: the example of waitlisted kidney transplant candidates. Am J Transplant **2016**; 16:1958–66.
24. Schumacher M, Ohneberg K, Beyersmann J. Competing risk bias was common in a prominent medical journal. J Clin Epidemiol **2016**; 80:135–6.
25. Biran N, Ip A, Ahn J, et al. Tocilizumab among patients with COVID-19 in the intensive care unit: a multicentre observational study. Lancet Rheumatol **2020**; 2:e603–12.
26. Geleris J, Sun Y, Platt J, et al. Observational study of hydroxychloroquine in hospitalized patients with Covid-19. N Engl J Med **2020**; 382:2411–8.
27. Huet T, Beaussier H, Voisin O, et al. Anakinra for severe forms of COVID-19: a cohort study. Lancet Rheumatol **2020**; 2:e393–400.
28. Rosenberg ES, Dufort EM, Udo T, et al. Association of treatment with hydroxychloroquine or azithromycin with in-hospital mortality in patients with COVID-19 in New York State. JAMA **2020**; 323:2493–502.
29. Oulhaj A, Ahmed LA, Prattes J, et al. The competing risk between in-hospital mortality and recovery: a pitfall in COVID-19 survival analysis research. medRxiv [Preprint]. Posted online 15 July **2020**. doi:10.1101/2020.07.11.20151472.
30. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc **1999**; 94:496–509.
31. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. Circulation **2016**; 133:601–9.
32. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res **2011**; 46:399–424.
33. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika **1983**; 70:41–55.
34. Martens EP, Pestman WR, de Boer A, et al. Systematic differences in treatment effect estimates between propensity score methods and logistic regression. Int J Epidemiol **2008**; 37:1142–7.
35. Ali MS, Prieto-Alhambra D, Lopes LC, et al. Propensity score methods in health technology assessment: principles, extended applications, and recent advances. Front Pharmacol **2019**; 10:973.
36. Streiner DL, Norman GR. The pros and cons of propensity scores. Chest **2012**; 142:1380–2.
37. Singh S, Khan A, Chowdhry M, Chatterjee A. Outcomes of hydroxychloroquine treatment among hospitalized COVID-19 patients in the United States—real-world evidence from a federated electronic medical record network. medRxiv [Preprint]. Posted online 19 May **2020**. doi:10.1101/2020.05.12.20099028.
38. Elze MC, Gregson J, Baber U, et al. Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. J Am Coll Cardiol **2017**; 69:345–57.
39. Colaneri M, Bogliolo L, Valsecchi P, et al. Tocilizumab for treatment of severe COVID-19 patients: preliminary results from SMAtteo Covid19 registry (SMACORE). Microorganisms **2020**; 8:695.
40. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. Br J Cancer **2014**; 110:551–5.
41. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med **2000**; 342:1887–92.
42. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med **2000**; 342:1878–86.
43. Vandenbroucke JP, von Elm E, Altman DG, et al; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Ann Intern Med **2007**; 147:W163–94.