



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.



## Systematic review

## Methodological evaluation of bias in observational coronavirus disease 2019 studies on drug effectiveness

Oksana Martinuka, Maja von Cube, Martin Wolkewitz\*

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Centre, University of Freiburg, Freiburg, Germany

## ARTICLE INFO

## Article history:

Received 21 December 2020

Received in revised form  
12 March 2021

Accepted 21 March 2021

Available online 1 April 2021

Editor: L. Leibovici

## Keywords:

Competing risk bias

Confounding bias

Coronavirus disease 2019

Drug effectiveness

Immortal time bias

## ABSTRACT

**Background and objective:** Observational studies may provide valuable evidence on real-world causal effects of drug effectiveness in patients with coronavirus disease 2019 (COVID-19). As patients are usually observed from hospital admission to discharge and drug initiation starts during hospitalization, advanced statistical methods are needed to account for time-dependent drug exposure, confounding and competing events. Our objective is to evaluate the observational studies on the three common methodological pitfalls in time-to-event analyses: immortal time bias, confounding bias and competing risk bias.

**Methods:** We performed a systematic literature search on 23 October 2020, in the PubMed database to identify observational cohort studies that evaluated drug effectiveness in hospitalized patients with COVID-19. We included articles published in four journals: *British Medical Journal*, *New England Journal of Medicine*, *Journal of the American Medical Association* and *The Lancet* as well as their sub-journals.

**Results:** Overall, out of 255 articles screened, 11 observational cohort studies on treatment effectiveness with drug exposure–outcome associations were evaluated. All studies were susceptible to one or more types of bias in the primary study analysis. Eight studies had a time-dependent treatment. However, the hazard ratios were not adjusted for immortal time in the primary analysis. Even though confounders presented at baseline have been addressed in nine studies, time-varying confounding caused by time-varying treatment exposure and clinical variables was less recognized. Only one out of 11 studies addressed competing event bias by extending follow-up beyond patient discharge.

**Conclusions:** In the observational cohort studies on drug effectiveness for treatment of COVID-19 published in four high-impact journals, the methodological biases were concerningly common. Appropriate statistical tools are essential to avoid misleading conclusions and to obtain a better understanding of potential treatment effects. **Oksana Martinuka, Clin Microbiol Infect 2021;27:949**

© 2021 European Society of Clinical Microbiology and Infectious Diseases. Published by Elsevier Ltd. All rights reserved.

## Introduction

With a growing number of publications on potential therapeutic candidates for coronavirus disease 2019 (COVID-19) treatment, high-quality observational studies have added value to the assessment of drug benefit in the real-world health-care setting [1,2]. However, the observational study design has important limitations and poses several challenges in the data analysis, particularly regarding the time-dependent nature of the data [3].

Ignorance of methodological biases in observational studies with time-to-event analysis may lead to distorted results and false conclusions on the exposure–outcome associations [4,5]. The aim of this article is to review the observational studies on evaluation of drug effectiveness in individuals with COVID-19 with regard to the presence of three methodological biases referred to as immortal time bias, confounding and competing risk bias. This work also aims to give recommendations on avoiding these biases.

In contrast to randomized clinical trials, in observational cohort studies a drug of interest is often prescribed after initiation of a study, e.g. later during follow-up. Immortal time typically occurs when there is a delay or waiting period between cohort entry and the time of the first prescription, which is falsely accounted for as drug-exposed time. Hence, exposed participants must survive the

\* Corresponding author: Martin Wolkewitz, Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Centre, University of Freiburg, Freiburg, Germany.

E-mail address: [wolke@imbi.uni-freiburg.de](mailto:wolke@imbi.uni-freiburg.de) (M. Wolkewitz).

initial time period to receive treatment if they are not assigned to the unexposed cohort [5,6]. Exclusion or misclassification of observation time often leads to immortal time bias and consequently to artificial overestimation or underestimation of drug effectiveness [6,7].

Control of both time-fixed and time-varying confounding is crucial because of the lack of randomization in observational real-world data [8]. In contrast to time-fixed confounding bias, time-varying confounding is commonly encountered in longitudinal observational studies [9–11]. In clinical epidemiology, treatment exposures are often time-varying and the values of potential confounders may change during the observational period, leading to time-varying confounding [12,13]. A lack of control of confounding may lead to biased estimates of treatment effects and causal misinterpretation [11,14].

Another issue that is often observed in observational studies with survival or time-to-event analysis is the competing risk bias. By definition, a competing risk is an event that modifies the chance of occurrence of the primary event of interest and can occur when a patient is at risk of more than one type of event [15].

Competing risk events are frequently observed in hospital epidemiology when the follow-up ends with hospital discharge. In turn, hospital discharge is a competing risk for the hospital death that is often the outcome of interest [7,16]. In survival analyses, the survivor function and the hazard function are the two most common methods for representation of survival data. In the presence of competing risks, the naïve Kaplan–Meier estimator takes the competing risks as censored observations. As a result the Kaplan–Meier analysis overestimates cumulative risks and produces upwards-biased estimates [16–18]. Furthermore, the causal treatment effects cannot be completely assessed if the measures of association are reported only for the event of interest [7,17,18]. The key recommendations to address all the three methodological biases are listed as summary points in [Box 1](#).

## BOX 1

### Summary points

- Time zero, time of treatment initiation and end of follow-up should be clearly described in the research methods. To avoid immortal time bias, follow-up time and treatment status should be properly accounted by the design or statistical methods.
- To avoid competing risk bias, a competing risk analysis has to be performed to describe results on all cause-specific hazards and visualized using cumulative hazard functions. If there is no follow-up beyond hospital discharge, the discharge has to be handled as a ‘competing event’ in the statistical analysis.
- Immortal time bias, time-fixed confounding and competing risk bias can be addressed simultaneously by applying a cause-specific Cox regression for an event of interest and a competing event with the inclusion of treatment as a time-dependent covariate.
- If robust time-varying clinical data are available and applicable, time-varying confounding should be addressed using a marginal structural Cox model or other g-methods for causal inference.
- Data analysis should be performed in a stepwise manner, starting from simple straightforward methods to increased model complexity.

## Methods

### Study selection, inclusion and exclusion criteria

A literature search of observational studies on drug effectiveness in hospitalized patients with COVID-19 was performed in four journals: *the British Medical Journal (The BMJ)*, *the New England Journal of Medicine (NEJM)*, *the Journal of the American Medical Association (JAMA)*, and *The Lancet*. Also, their sub-journals were selected. These journals were chosen because they are the leading and most cited medical journals with high impact factors. This review required original retrospective observational cohort studies with primary data including time-to-event data analysis. As a result, comments, correspondence, opinions, researcher letters and audio interviews were excluded. The inclusion criteria were hospitalized patients diagnosed with COVID-19. Both single-centre and multicentre studies were included. The literature search included studies that investigated drug effectiveness alone or in combination with standard therapy. Studies that focused on drug preventive rather than therapeutic effects were excluded. Articles that investigated the adverse effects of drug therapy only were not included either. The selection included studies with a comparative control group in which patients either received standard therapy or no treatment. Studies with different time-dependent clinical outcomes, such as in-hospital mortality, overall survival and survival without transfer to the intensive care unit (ICU), were selected for our review. Finally, the study selection was limited to articles published between the start of the pandemic in December 2019 and 23 October 2020. The selection procedure of our literature search is shown in the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow diagram ([Fig. 1](#)).

### Data sources and search strategy

The systematic literature review was performed in the PubMed database on 23 October 2020. The review was conducted according to the PRISMA guidelines for reporting systematic reviews [19]. The list of keywords and detailed search strategy are described in the Supplementary material ([Tables S1 and S2](#)).

### Methodological bias assessment

The assessment of biases was performed independently by all three authors. Each observational study was reviewed in-depth to determine the possible presence of immortal time bias, confounding bias and competing risk bias by screening the Methods as well as the Results section. The design, methods and statistical techniques were evaluated. Studies were considered to be susceptible to the immortal time bias if the time period before treatment allocation was not addressed in the analysis and, as a result, time-dependent drug exposure was statistically handled as a time-fixed exposure. The susceptibility to both time-fixed and time-varying confounding biases was evaluated. The occurrence of time-fixed confounding biases was determined if baseline covariates were not adjusted. However, the evaluation of correctness, adequacy and effectiveness of the applied adjustment methods and the degree of residual confounding were outside the scope of this review. The time-varying confounders were evaluated by examining the presence of time-dependent treatment and time-varying covariates. Studies were considered vulnerable to time-varying confounding bias if sophisticated statistical methods for controlling of time-dependent confounders were not applied. Studies were considered as being susceptible to competing risk bias if follow-up of patients was ensured only until discharge and

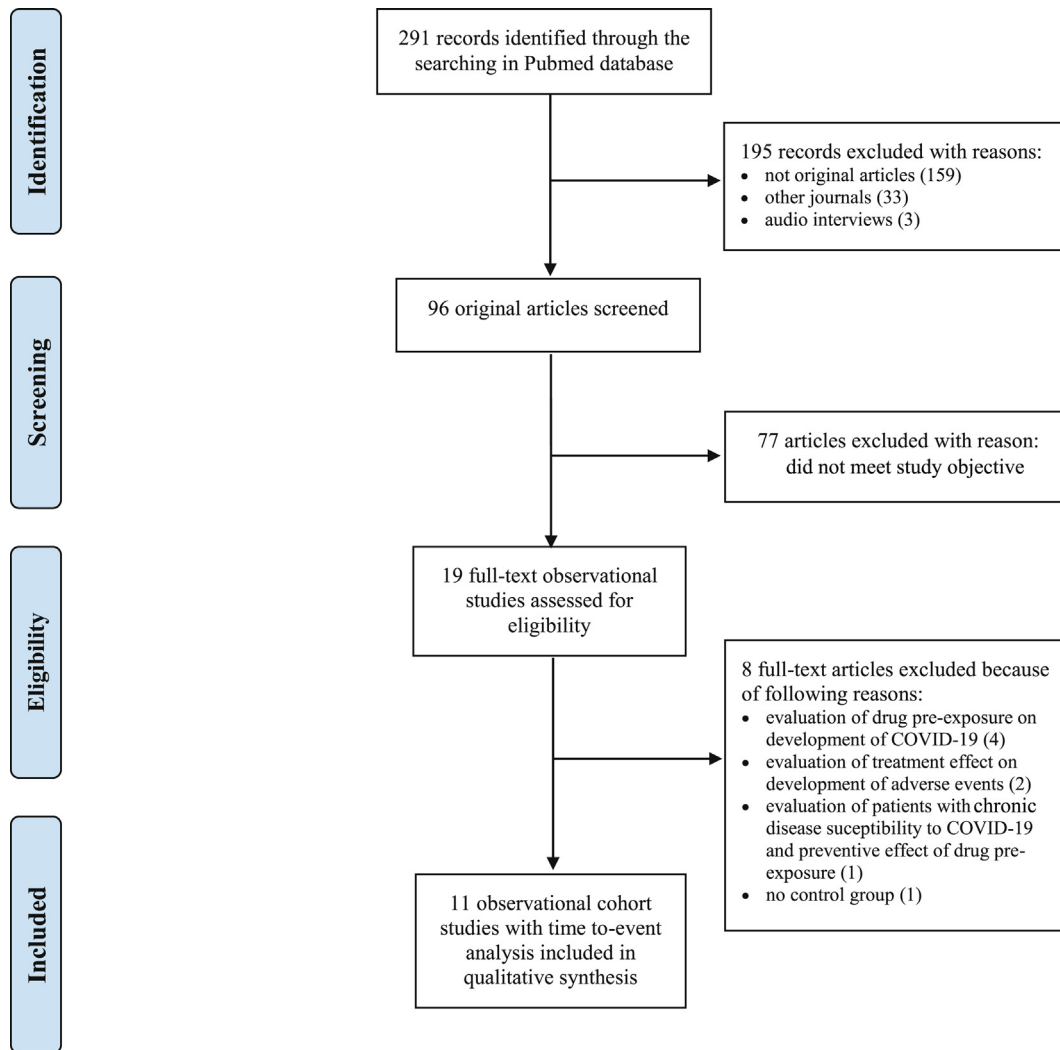


Fig. 1. Flow diagram of the identified studies in the systematic literature review.

individuals who were discharged alive were censored at the time of discharge for quantifying the probability of experiencing the event of interest. To identify whether the biases were addressed, the study methodology (i.e. study design and applied analytical methods) and results were evaluated. In addition, the supplementary materials of the included articles were checked.

## Results

Out of 255 articles screened, 11 observational cohort studies on the drug effectiveness for COVID-19 treatment in hospitalized patients were included; six articles from *The Lancet* and its sub-journals, three articles from *the Journal of the American Medical Association* and its sub-journals as well as one from *the British Medical Journal* and one from *the New England Journal of Medicine* (Table 1) [20–30]. These observational studies investigated the effectiveness of drugs such as anakinra, azithromycin, chloroquine or hydroxychloroquine, methylprednisolone and tocilizumab. These drugs were administered alone or in combination with standard therapy. All of these studies were susceptible to at least one of the three discussed types of bias (Fig. 2). The results and examples of the identified biases are given in the following sections.

### Occurrence of immortal time bias

Overall, eight studies were susceptible to immortal time bias [20–25,27,29]; in the three remaining studies, the start of the follow-up and/or the start of treatment exposure administration were not clearly defined [26,28,30]. In two of these studies, treatment was artificially converted into a time-fixed exposure by using different time scales for the treated and untreated cohorts, i.e. for the untreated group the start of follow-up was at time zero whereas for the exposed cohort the start of follow-up was set at the time of treatment [21,25]. As a result, artificial conversion of the time scale made the evaluation of the treatment impossible because of starting events before follow-up. It is likely that the three studies with unclear start of treatment administration are also susceptible to immortal time bias (Table 1) [26,28,30].

To account for immortal time bias, four studies performed sensitivity analyses and included the results in the secondary analyses or in the supplementary materials [22,24,27,29]. To address the immortal time bias, methods such as the landmark analysis [22], the target trial emulation analysis [24,27] and the time-dependent Cox regression analysis [29] were performed (Table 1). For instance, Rosenberg et al. accounted for delayed

**Table 1**  
Characteristics of included studies and their features of immortal time bias

First author [reference]	Journal title abbreviation	Study baseline (time zero)	Start of treatment exposure	Susceptibility to immortal time bias	Accounted for bias		Methods addressed immortal time bias
					Primary analysis	Secondary analysis	
Biran [20]	<i>Lancet Rheumatol</i>	admission to ICU	early in ICU	yes	no	no	not reported
Cavalli [21]	<i>Lancet Rheumatol</i>	admission to hospital	day zero (different time scales for treated and untreated patients)	yes <sup>a</sup>	no	no	not reported
Geleris [22]	<i>N Engl J Med</i>	24 hours after admission to hospital	within 48 hours or during follow-up period	yes	no	yes	landmark analysis
Guaraldi [23]	<i>Lancet Rheumatol</i>	admission to tertiary hospital	at time of hospital admission up to 24 hours	yes	no	no	not reported
Gupta [24]	<i>JAMA Intern Med</i>	admission to hospital	within 2 days of ICU admission	yes	no	yes	nested target trial emulation analysis
Huet [25]	<i>Lancet Rheumatol</i>	admission to hospital	day zero (different time scales for treated and untreated patients)	yes <sup>a</sup>	no	no	not reported
Kuderer [26]	<i>Lancet</i>	start of follow-up not clearly defined	start of treatment not clearly defined	start of follow-up, start of treatment is not clearly defined	n/a	n/a	not reported
Mahévas [27]	<i>BMJ</i>	admission to hospital	within 48 hours, more than 48 hours or during follow-up	yes	no	yes	target trial emulation analysis (mimicking an ITT and an as- treated)
Mehra [28]	<i>Lancet</i>	start of follow-up not clearly defined	within 48 hours after diagnosis established	start of follow-up is not clearly defined	n/a	n/a	not reported
Rosenberg [29]	<i>JAMA</i>	admission to hospital	at any time during hospitalization	yes	no	yes	sensitivity analysis with time- dependent treatment status
Wu [30]	<i>JAMA Intern Med</i>	admission to hospital	start of treatment not clearly defined	start of treatment is not clearly defined	n/a	n/a	not reported

Abbreviations: ICU, intensive care unit; ITT, intention-to-treat analysis; n/a, not available.

Journal title abbreviations: BMJ, *the British Medical Journal*; JAMA, *the Journal of the American Medical Association*; JAMA Intern Med, *the Journal of the American Medical Association Internal Medicine*; N Engl J Med, *the New England Journal of Medicine*; Lancet, *The Lancet*; Lancet Rheumatol, *The Lancet Rheumatology*.

<sup>a</sup> Treatment was artificially converted to a time-fixed exposure.

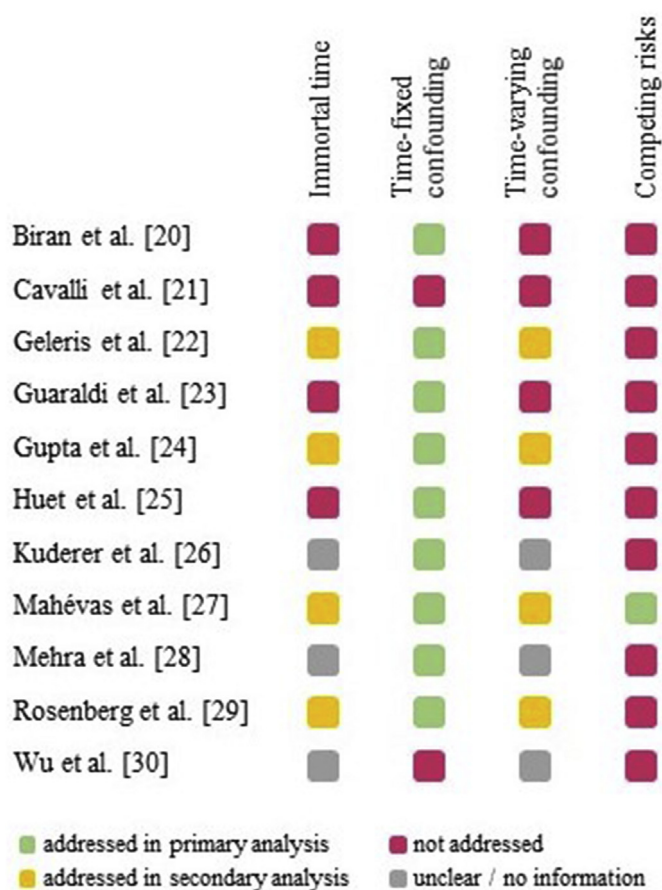


Fig. 2. The summary of biases for the included studies.

treatment start by using the time-dependent Cox proportional hazards model in their sensitivity analysis. The resulting hazard ratio for in-hospital mortality corrected for the immortal time bias increased in all treatment groups. In particular, for the group treated with hydroxychloroquine in combination with azithromycin it increased from 1.35 (95% CI 0.76–2.40) to 1.83 (95% CI 1.02–3.28) and for hydroxychloroquine alone from 1.08 (95% CI 0.63–1.85) to 1.46 (95% CI 0.84–2.55), respectively [29].

#### Occurrence of confounding bias

Out of the 11 studies susceptible to confounding, nine studies used methods to account for the time-fixed confounding bias [20,22–29]; the two remaining studies showed results from the crude (unadjusted) analysis only [21,30]. The most frequently used methods to control for time-fixed confounding were adjustment in a regression model or stratification. Seven studies applied the propensity score weighting methods for confounding adjustment (Table 2).

In contrast to the time-fixed confounding bias, the time-varying confounding bias was less recognized. In eight studies where the exposure was time-varying (see *Immortal time bias* section), we presume that treatment initiation was caused by evolving clinical characteristics of the patient, which led to time-varying confounding (Table 2) [20–25,27,29]. In these studies, treatment exposure was analysed as a baseline covariate and, as a result, time-

varying confounding was not addressed. For instance, covariates such as blood cell count as well as biochemical, coagulation and inflammatory parameters were more likely routinely collected and influenced the subsequent decisions on drug administration and on the outcome.

The time-varying confounding was controlled and secondary results were presented by four studies. For example, in the study conducted by Geleris et al. the landmarking analysis was based on the value of time-varying exposure at the landmark point (24 and 48 hours), after which the time-varying exposure may change value [22]. In the studies conducted by Gupta et al. and Mahévas et al. an observational target trial emulation methodology was used and appropriate adjustment methods, like inverse probability weighting, were applied [24,27]. In the study conducted by Rosenberg et al. a time-dependent Cox model that accounted for time-dependent treatment was used [29].

#### Occurrence of competing risk bias

Several time-to-event primary outcomes were investigated in the studies, such as development of acute respiratory distress syndrome, admission to ICU, administration of invasive mechanical ventilation, in-hospital death or 30-day in-hospital mortality, survival without transfer to ICU and overall survival. These end points were studied as a single event, or as a composite end point of several events (Table 3).

All in all, ten out of the eleven studies seemed to be susceptible to competing risk bias because a follow-up ended with hospital discharge and discharged patients were censored in probability analyses [20–26,28–30]. In nine out of the ten studies, the classical Kaplan–Meier method was applied to display the treatment effect and to compare survival probabilities between the treatment and control groups [20–25,27,29,30]. In one study, performed by Mahévas et al., the competing event of discharge alive for the primary outcome (survival without transfer to ICU) was addressed by the extension of follow-up for discharged patients, and the corrected Kaplan–Meier curves were presented in the supplementary material [27].

All of the studies used a regression model (logistic regression and/or the Cox proportional hazard regression model) to quantify the treatment effect for the primary event of interest. The Cox proportional hazards model provides a correct estimate of the treatment effect on the event of interest. Nonetheless, the regression analysis is incomplete if the hazard ratio for the competing event is not reported. Only in one sensitivity analysis, conducted by Gupta et al., discharge was addressed by keeping discharged patients in the risk set and not treating them as randomly censored. As a result, the adjusted hazard ratio of the event of interest shifted from 0.71 (95% CI 0.56–0.92) to 0.72 (95% CI 0.56–0.93). Keeping discharged patients in the risk set is targeted at estimating a sub-distribution hazard [31]. We remark that this approach is not appropriate if the treatment is time-dependent [24].

#### Discussion

Since the number of studies on COVID-19 treatment effectiveness and the speed of publishing new data in journals has drastically increased [32], unbiased results from observational studies are extremely important as a complement to randomized controlled trials. The methodological challenges in studying observational COVID-19 data and performing statistical analysis on drug effectiveness has been described in detail elsewhere [33,34]. However, our review has demonstrated that methodological issues such as

immortal time bias, confounding bias and competing risk bias are commonly found in articles published in high-impact medical journals.

To our knowledge, for treatment effectiveness assessment, the best practice is to perform analysis in a stepwise manner, starting from simple straightforward methods to increased model complexity. To ensure a complex approach and enhance confidence, the results from simple and extended models should be presented side by side [35]. This approach can be used to address time-to-treatment, confounders, and competing risks issues. For details on the proper statistical techniques and analysis of time-to-event data in observational studies, we highly recommend the guidance of Kragh Andersen et al. [36].

As the primary step, a complete competing risk analysis should be presented that includes a cause-specific Cox regression analysis for the event of interest and for all competing events [7,33,36,37]. An initial model should include baseline covariates, further, time-fixed confounders need to be adjusted. This analysis provides insights into the effect of treatment and covariates on each event type separately [38]. To account for non-randomized treatment administration and reduce the effects of confounding, the model can be

extended using more complex procedures, such as propensity score-based methods. Further, to estimate the causal treatment effect, more sophisticated analysis for time-varying confounders, such as g-methods (i.e. the inverse probability of treatment weighted marginal structural models, g-computation formula and g-estimation of structural nested models) must be applied [8,10]. For instance, in the excellent study by Ursino et al., the g-computation approach was applied to evaluate the potential treatment effect of corticosteroids and interleukin antagonists like tocilizumab and anakinra [39]. In addition, we refer to the textbook of Hernán et al. for the detailed tutorial on causal inference and g-methods [40].

In the presence of competing risks, we discourage the use of Kaplan–Meier plots for effect visualization because of the high risk of potentially misleading conclusions. Instead, cumulative cause-specific hazards can be calculated and should be plotted for the events of interest and for the competing events [41]. This method allows us to account for competing events and to display treatment exposure differences [16]. More generally, the multistate methodology offers several advantages for survival data analysis with multiple outcomes. A multistate model allows us to evaluate the

**Table 2**  
Characteristics of included studies and their features of time-fixed and time-varying confounding bias

First author [reference]	Type of exposure	Susceptibility to time-fixed/ time-varying confounders	Accounted for time-fixed confounding bias	Accounted for time-varying confounding bias		Methods controlling for confounding bias	
				Primary analysis	Secondary analysis	Time-fixed	Time-varying
Biran [20]	time-dependent exposure	yes/yes	yes	no	no	semiparametric multivariate methods: regression models, PS matched analysis and PS covariate adjustment	not reported
Cavalli [21]	time-fixed exposure at day zero	yes/no	no	no	no	n/a	not reported
Geleris [22]	time-dependent exposure	yes/yes	yes	no	yes	stratification; semiparametric multivariate methods: regression models, PS matched analysis and PS covariate adjustment, IPW	landmark approach
Guaraldi [23]	time-dependent exposure	yes/yes	yes	no	no	stratification; semiparametric multivariate methods: regression models and IPW	not reported
Gupta [24]	time-dependent exposure	yes/yes	yes	no	yes	semiparametric multivariate methods: regression models and IPW	by conducting the nested target trial emulation approach and using IPW
Huet [25]	time-fixed exposure at day zero	yes/no	yes	no	no	semiparametric multivariate methods: regression model	not reported
Kuderer [26]	unclear defined	yes/unclear	yes	unclear	n/a	semiparametric multivariate methods: regression models standardization;	not reported
Mahévas [27]	time-dependent exposure	yes/yes	yes	no	yes	semiparametric multivariate methods: regression models, PS and IPTW	by conducting the target trial emulation approach and using IPTW
Mehra [28]	unclear defined	yes/unclear	yes	unclear	n/a	semiparametric multivariate methods: regression models, PS matched analysis	not reported
Rosenberg [29]	time-dependent exposure	yes/yes	yes	no	yes	by design: hospital-stratified random sampling; at analysis stage: semiparametric multivariate methods: regression models	time-dependent Cox regression
Wu [30]	unclear defined	yes/unclear	no	unclear	n/a	n/a	not reported

Abbreviations: IPW, inverse probability weighting; IPTW, inverse probability of treatment weighting; PS, propensity score.

**Table 3**  
Characteristics of included studies and their features of competing risk events

First author [reference]	Primary end point/outcome	Competing event	Competing risk analysis		Cause-specific regression analysis for competing event	Graphical representation of survival curves
			In primary analysis	In secondary analysis		
Biran [20]	in-hospital mortality	discharge alive	no	no	no	event-free survival probabilities; i.e. KM plot <sup>a</sup> for overall survival
Cavalli [21]	overall survival (at day 21), MV-free survival	discharge alive, discharge without need for MV	no	no	no	event-free survival probability; i.e. KM plots for overall survival and MV free survival
Geleris [22]	intubation or death without intubation as a composite endpoint	discharge alive without need for intubation	no	no	no	event-free survival probability; i.e. KM plot
Guaraldi [23]	composite of IMV or death, in-hospital death	discharge alive without need for IMV	no	no	no	cumulative incidence probabilities for MV or death, and death alone; i.e. 1–KM
Gupta [24]	in-hospital death (30-day mortality)	discharge alive	no	yes <sup>b</sup>	no <sup>b</sup>	cumulative incidence probabilities for mortality; i.e. 1–KM
Huet [25]	admission to ICU for IMV or death as a composite endpoint	discharge alive before ICU or without need for IMV	no	no	no	event-free survival probabilities; i.e. KM plots for event-free of IMV, death, and IMV or death
Kuderer [26]	30-day all-cause mortality	discharge alive	no	no	no	forest plot for 30-day all-cause mortality
Mahévas [27]	survival without transfer to ICU (at day 21)	no competing event	not necessary, CR is addressed by extended follow-up		no competing risk bias	event-free survival probability; i.e. KM plot for survival without ICU admission
Mehra [28]	in-hospital mortality	discharge alive	no	no	no	forest plot for in-hospital mortality
Rosenberg [29]	in-hospital mortality	discharge alive	no	no	no	cumulative incidence probabilities for in-hospital mortality; i.e. 1–KM
Wu [30]	development of ARDS and death among those with ARDS	discharge alive among those with ARDS	no	no	no	event-free survival probability; i.e. KM plots for overall survival

Abbreviations: ARDS, acute respiratory distress syndrome; CR, competing risk; ICU, intensive care unit; IMV, invasive mechanical ventilation; MV, mechanical ventilation; KM, Kaplan–Meier.

<sup>a</sup> The KM curves provided for the secondary end point (overall survival).

<sup>b</sup> Discharge was addressed by keeping discharged patients in the risk set and not treating them as randomly censored. Keeping discharged patients in the risk set targets at estimating a sub-distribution hazard. This approach is not appropriate if the treatment is time-dependent.



time dynamics of disease progression by describing the occurrence of the events as transitions between different states [42]. This also makes it possible to visualize primary outcomes of interest and competing events simultaneously over a time period in a single plot [43].

The article by Lapadula et al. presents an appropriate analysis on antiviral treatment assessment in patients with COVID-19 [44]. In their study, the landmark analysis was applied to estimate survival rates conditional on a landmark time (7 days after intubation) to adjust for immortal time bias. The exposure was modelled to each treatment (remdesivir, hydroxychloroquine and lopinavir/ritonavir) as time-dependent covariates. The competing risk bias was addressed by estimation of two cause-specific hazard functions: for the event of interest such as in-patient hospital death and for the competing event such as discharge, respectively. As a result, the competing risk analysis showed that remdesivir was significantly associated with hospital discharge (hazard ratio 2.25; 95% CI 1.27–3.97) and with a non-significantly lower mortality (hazard ratio 0.73; 95% CI 0.26–2.1). Moreover, the cumulative hazard plots were used to visualize potential treatment effects [44]. This example addresses all three methodological biases and clearly demonstrates the need to study competing events.

With our study, we assessed the presence of methodological biases and provided recommendations. Several study limitations are noteworthy, such as the inclusion of only four journals and their sub-journals. As a result, we cannot extend the results obtained to other journals. Another limitation is that an incomplete or unclear description of the study methodology or/and results in the investigated publications could have resulted in incorrect interpretation of the analyses.

## Conclusions

Immortal time bias, confounding bias and competing risk bias alone or in combination were present in all of the 11 reviewed observational studies on treatment effectiveness evaluation for hospitalized patients with COVID-19. These biases may have led to a severe overestimation or underestimation of COVID-19 treatment effectiveness estimates. As a result, the drugs may appear to be either more effective or conversely to have little to no effect. In fact, making valid causal inferences from real-world observational data is a demanding task that requires high-quality data and adequate statistical methods as well as clinical knowledge and statistical expertise.

Even though every study has its unique features, which should be addressed in a tailor-made analysis, the summary points (Box 1) and the following recommendations could be taken into account to prevent or decrease the occurrence and the severity of the methodological biases. To overcome immortal time bias, time-dependent treatment exposures should be included in the Cox regression model as a time-dependent variable [45]. In the presence of time-varying confounders, a marginal structural Cox model can be used and estimated using inverse probability of treatment weights [8]. In the competing risks setting, the cause-specific hazard ratios could also be obtained from the time-dependent Cox regression model. Furthermore, treatment effects should be visualized with cumulative hazard functions for the event of interest and all competing events [37,46]. Finally, corrected estimates should be reported as primary findings in the primary results section.

## Transparency declaration

The authors declare no conflicts of interest. This study has been funded by the German Research Foundation (DFG) with project grant WO 1746/5-1.

## Acknowledgements

We acknowledge Dr. Klaus-Dieter Wolkewitz and Prof. Martin Schumacher for reviewing the manuscript. We thank Layla Nieden for editing a previous version of the review.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cmi.2021.03.003>.

## References

- [1] Radford MJ, Foody JM. How do observational studies expand the evidence base for therapy? *JAMA* 2001;286:1228–30.
- [2] Gilmartin-Thomas JF, Liew D, Hopper I. Observational studies and their utility for practice. *Aust Prescr* 2018;41:82–5.
- [3] Prada-Ramallal G, Takkouche B, Figueiras A. Bias in pharmacoepidemiologic studies using secondary health care databases: a scoping review. *BMC Med Res Methodol* 2019;19:53.
- [4] Sharma M, Nazareth I, Petersen I. Observational studies of treatment effectiveness: worthwhile or worthless? *Clin Epidemiol* 2019;11:35–42.
- [5] Suissa S, Dell'Aniello S. Time-related biases in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2020;29:1101–10.
- [6] Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf* 2007;16:241–9.
- [7] Wolkewitz M, Schumacher M. Survival biases lead to flawed conclusions in observational treatment studies of influenza patients. *J Clin Epidemiol* 2017;84:121–9.
- [8] Fewell Z, Hernán MA, Wolfe F, Tilling K, Choi H, Sterne JAC. Controlling for time-dependent confounding using marginal structural models. *Stata J* 2004;4:402–20.
- [9] Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. *BMJ* 2017;359:j4587.
- [10] Williamson T, Ravani P. Marginal structural models in clinical research: when and how to use them? *Nephrol Dial Transplant* 2017;32:ii84–90.
- [11] Pierce M, Dunn G, Millar T. Confounding in longitudinal studies in addiction treatment research. *Addict Res Theory* 2017;25:236–42.
- [12] Pazzagli L, Linder M, Zhang M, Vago E, Stang P, Myers D, et al. Methods for time-varying exposure related problems in pharmacoepidemiology: an overview. *Pharmacoepidemiol Drug Saf* 2018;27:148–60.
- [13] Becker SO, Ichino A. Estimation of average treatment effects based on propensity scores. *Stata J* 2002;2:358–77.
- [14] 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.
- [15] Dutz A, Löck S. Competing risks in survival data analysis. *Radiother Oncol* 2019;130:185–9.
- [16] Wolkewitz M, Schumacher M. Neuraminidase inhibitors and hospital mortality in British patients with H1N1 influenza A: a re-analysis of observational data. *PLoS One* 2016;11:e0160430.
- [17] Wolkewitz M, Cooper BS, Bonten MJM, Barnett AG, Schumacher M. Interpreting and comparing risks in the presence of competing events. *BMJ* 2014;349:g5060.
- [18] Schuster NA, Hoogendijk EO, Kok AAL, Twisk JWR, Heymans MW. Ignoring competing events in the analysis of survival data may lead to biased results: a nonmathematical illustration of competing risk analysis. *J Clin Epidemiol* 2020;122:42–8.
- [19] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
- [20] Biran N, Ip A, Ahn J, Go RC, Wang S, Mathura S, et al. Tocilizumab among patients with COVID-19 in the intensive care unit: a multicentre observational study. *Lancet Rheumatol* 2020;2:e603–12.
- [21] Cavalli G, de Luca G, Campochiaro C, Della-Torre E, Ripa M, Canetti D, et al. Interleukin-1 blockade with high-dose anakinra in patients with COVID-19, acute respiratory distress syndrome, and hyperinflammation: a retrospective cohort study. *Lancet Rheumatol* 2020;2:e325–31.
- [22] Geleris J, Sun Y, Platt J, Zucker J, Baldwin M, Hripcsak G, et al. Observational study of hydroxychloroquine in hospitalized patients with Covid-19. *N Engl J Med* 2020;382:2411–8.
- [23] Guaraldi G, Meschiari M, Cozzi-Lepri A, Milic J, Tonelli R, Menozzi M, et al. Tocilizumab in patients with severe COVID-19: a retrospective cohort study. *Lancet Rheumatol* 2020;2:e474–84.
- [24] Gupta S, Wang W, Hayek SS, Chan L, Mathews KS, Melamed ML, et al. Association between early treatment with tocilizumab and mortality among critically ill patients with COVID-19. *JAMA Intern Med* 2020. <https://doi.org/10.1001/jamainternmed.2020.6252>.

- [25] Huet T, Beaussier H, Voisin O, Jouveshomme S, Dauriat G, Lazareth I, et al. Anakinra for severe forms of COVID-19: a cohort study. *Lancet Rheumatol* 2020;2:e393–400.
- [26] Kuderer NM, Choueiri TK, Shah DP, Shyr Y, Rubinstein SM, Rivera DR, et al. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. *Lancet* 2020;395:1907–18.
- [27] Mahévas M, Tran V-T, Roumier M, Chabrol A, Paule R, Guillaud C, et al. Clinical efficacy of hydroxychloroquine in patients with covid-19 pneumonia who require oxygen: observational comparative study using routine care data. *BMJ* 2020;369:m1844.
- [28] Mehra MR, Desai SS, Ruschitzka F, Patel AN. RETRACTED: hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *Lancet* 2020. [https://doi.org/10.1016/S0140-6736\(20\)31180-6](https://doi.org/10.1016/S0140-6736(20)31180-6).
- [29] Rosenberg ES, Dufort EM, Udo T, Wilberschied LA, Kumar J, Tesoriero J, 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.
- [30] Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med* 2020;180:934–43.
- [31] Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999;94:496.
- [32] Horbach SP. Pandemic Publishing: medical journals drastically speed up their publication process for Covid-19. *bioRxiv* 2020;2020. <https://doi.org/10.1101/2020.04.18.045963>. 04.18.045963.
- [33] Wolkewitz M, Lambert J, von Cube M, Bugiera L, Grodd M, Hazard D, et al. Statistical analysis of clinical COVID-19 data: a concise overview of lessons learned, common errors and how to avoid them. *Clin Epidemiol* 2020;12:925–8.
- [34] Wolkewitz M, Puljak L. Methodological challenges of analysing COVID-19 data during the pandemic. *BMC Med Res Methodol* 2020;20:81.
- [35] Zwietering MH. Quantitative risk assessment: is more complex always better? Simple is not stupid and complex is not always more correct. *Int J Food Microbiol* 2009;134:57–62.
- [36] Kragh Andersen P, Pohar Perme M, van Houwelingen HC, Cook RJ, Joly P, Martinussen T, et al. Analysis of time-to-event for observational studies: guidance to the use of intensity models. *Stat Med* 2021;40:185–211.
- [37] Huebner M, Wolkewitz M, Enriquez-Sarano M, Schumacher M. Competing risks need to be considered in survival analysis models for cardiovascular outcomes. *J Thorac Cardiovasc Surg* 2017;153:1427–31.
- [38] Wolbers M, Koller MT, Stel VS, Schaar B, Jager KJ, Leffondré K, et al. Competing risks analyses: objectives and approaches. *Eur Heart J* 2014;35:2936–41.
- [39] Ursino M, Dupuis C, Buetti N, de Montmollin E, Bouadma L, Golgran-Toledano D, et al. Multistate modeling of COVID-19 patients using a large multicentric prospective cohort of critically ill patients. *J Clin Med* 2021;10. <https://doi.org/10.3390/jcm10030544>.
- [40] Hernán MA, Robins JM. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC; 2020.
- [41] Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation* 2016;133:601–9.
- [42] Beyersmann J, Wolkewitz M, Allignol A, Grambauer N, Schumacher M. Application of multistate models in hospital epidemiology: advances and challenges. *Biom J* 2011;53:332–50.
- [43] von Cube M, Grodd M, Wolkewitz M, Hazard D, Wengenmayer T, Canet E, et al. Harmonizing heterogeneous endpoints in coronavirus disease 2019 trials without loss of information. *Crit Care Med* 2021;49:e11–e19.
- [44] Lapadula G, Bernasconi DP, Bellani G, Soria A, Rona R, Bombino M, et al. Remdesivir use in patients requiring mechanical ventilation due to COVID-19. *Open Forum Infect Dis* 2020;7:ofaa481.
- [45] Jones M, Fowler R. Immortal time bias in observational studies of time-to-event outcomes. *J Crit Care* 2016;36:195–9.
- [46] Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J Clin Epidemiol* 2013;66:648–53.