





Comparative effectiveness research in COVID-19 using real-world data: methodological considerations

Journal of **Comparative Effectiveness Research**

Stephanie H Read¹ , Artak Khachatryan¹, Aastha Chandak² , Roman Casciano² , Paul Hodgkins³, Richard Haubrich³  & Essy Mozaffari^{*,3}

¹Certara, London, UK

²Certara, New York, NY 10017, USA

³Gilead Sciences, Foster City, CA 94404, USA

*Author for correspondence: Tel.: +1 650 577 6901; Essy.Mozaffari@Gilead.com

“RWD studies continue to emerge as an important source of evidence (i.e., effectiveness and safety) on a treatment’s use in clinical practice that can complement data from RCTs.”

First draft submitted: 23 July 2021; Accepted for publication: 13 August 2021; Published online: 31 August 2021

Keywords: bias • comparative effectiveness studies • confounding • COVID-19 • real-world data

COVID-19, caused by SARS-CoV-2, has caused considerable morbidity and mortality across the world and caused incalculable damage to global economies [1,2]. Randomized controlled trials (RCTs) are the prerequisite for identifying efficacious and safe treatments and have already contributed valuable evidence relating to potential treatments for COVID-19. Real-world data (RWD) studies can complement RCTs by offering opportunities to identify priorities for subsequent RCTs, including treatments that address the host response and to confirm prior RCT findings in groups that are more heterogeneous and with longer follow-up periods.

However, due to their observational nature, RWD studies are susceptible to confounding by indication and immortal time biases [3]. While these methodological considerations are not unique to COVID-19 studies, these limitations have been apparent in a number of recently published COVID-19 comparative effectiveness research (CER) and have also been highlighted in a recent methodological review [4]. Rigorous and transparent methods are required in the conduct of CER in general, but especially during a global public health crises.

Unique features of the pandemic have further complicated the conduct of CER to examine COVID-19 treatments. CER studies are usually undertaken to understand the effectiveness of treatments for conditions that we have a detailed understanding about, in terms of key confounding factors and practice patterns for its management and treatment. During the pandemic, knowledge of natural history, epidemiology and treatment options related to COVID-19 has evolved rapidly. This has led to considerable temporal and regional variations in COVID-19 treatment protocols. Patient case mix and risk of adverse outcomes have also changed over time and vary by region.

This article aims to describe key considerations and potential solutions to challenges faced in conducting CER on newly emerged diseases, specifically focusing on studies among ambulatory or hospitalized COVID-19 patients. Understanding these considerations can facilitate the design, analysis and interpretation of robust CER studies that contribute meaningful evidence for COVID-19 treatments.

Considerations in the presence of temporal variations in the management of COVID-19

The aforementioned temporal variations in our understanding of COVID-19 (which have changed over a time horizon of a few weeks to a few months) have important implications for the design of CER by influencing the relevance of study findings, introducing calendar time biases and complicating the methods for reducing confounding. How these features influence the definition of the study observation window, choice of comparators and analysis of CER are outlined in the following sections.

Defining the study observation window

The definition of the study observation window of CER in the COVID-19 era can affect the relevance, robustness and generalizability of the study findings. Pharmacological and non-pharmacological treatment practices have changed rapidly during the pandemic, with trends suggesting more selective use of treatments over time as new evidence on treatment effectiveness, or lack thereof has emerged. For example, data from the United States (US) have shown that hydroxychloroquine and azithromycin prescription increased dramatically in March 2020 following a single group non-randomized study [5] and the US FDA's subsequent emergency use authorization for hydroxychloroquine [6,7]. Thereafter, prescriptions declined following growing evidence of limited effectiveness and increased risk of harm associated with these treatments and the withdrawal of the emergency use authorization in June 2020. Utilization of remdesivir and dexamethasone increased after May 2020 with the publication of the results from the ACTT-1 and RECOVERY trials [8,9]. Non-pharmacological treatment practices such as proning and use of oxygen therapy have also changed during the pandemic and have contributed to improvements in case fatality rates over time [10].

Such rapid changes in treatment practices have implications for real-world CER since lags in administrative data availability could potentially render some findings meaningless by the time they are published. For example, a study may identify a new treatment regimen that may have been better than the management practices in place during the study period; however, currently available treatments may have evolved to supersede those that were originally studied. This was exemplified by the identification of remdesivir and dexamethasone as effective COVID-19 treatments and the likely subsequent improved management practices of COVID-19 thereafter. As a result, many CER studies using data from the initial phase of the pandemic will already lack external validity. Understanding trends in treatments and case fatality rates over time is therefore important to understand whether analyses need to account for temporal variations [11]. For example, patients administered the treatment of interest may be matched to comparison group patients hospitalized during the same calendar month. The definition of the allowable matching time-period should be carefully explored given the speed at which the pandemic has progressed. Reporting study outcomes overall and stratified by calendar month will also allow for a determination of whether the observed treatment effect is homogeneous over time, and whether reporting a single overall estimate of the treatment effect is valid. Similarly, restriction of study periods to phases of the pandemic when treatment practices are uniform and most likely to reflect current practices may also help to ensure findings are relevant at time of publication.

Identifying appropriate comparators

The rapid evolution of treatment protocols creates additional challenges in identifying appropriate comparison groups. In ideal settings, patients will be identical across comparator groups in all ways apart from the treatment of interest to reduce confounding by indication [3]. Examples of confounding by indication are abundant in observational research generally and the body of COVID-19 observational research has not been immune to this shortcoming. For example, an increased risk of COVID-19-related death associated with the use of inhaled corticosteroids identified among people with chronic obstructive pulmonary disorder or asthma in the UK could, as described by the authors, be plausibly explained by differences in asthma or chronic obstructive pulmonary disorder severity between users and nonusers of inhaled corticosteroids following the conduct of sensitivity analyses [12]. This highlights the need for careful selection of comparison groups to mitigate possible confounding by indication. This example also highlights good practice in the publication of such findings, with the authors openly describing that the findings were likely due to bias; therefore, ensuring that decision makers are not misled.

For COVID-19 studies, there are two main options for choice of comparison groups: active comparators and best available care designs. Active comparator groups comprise of patients who are administered a different treatment for the same indication and; therefore, should have similar sociodemographic and clinical characteristics to patients receiving the treatment of interest. Options for active comparators in COVID-19 CER have been limited due to the rapidly evolving patterns of use of many treatment candidates over short periods of time [6,7]. For example, in May 2020, remdesivir received emergency use authorization from the FDA for patients with severe COVID-19 and this authorization was subsequently expanded to all hospitalized patients in August 2020. As a result, the indication for remdesivir use changed over time according to disease severity, highlighting the difficulty in selecting appropriate comparator groups.

An alternative to using an active comparator is to use a best available care comparator group, defined as a group of patients who do not receive the treatment of interest; thereby, receiving best available care only. Due to the

absence of active comparator candidates and despite a higher risk of confounding by indication, best available care designs have been widely adopted in COVID-19 CER [13,14]. Patients receiving treatments are likely to be different to those who do not receive treatments and these differences are likely to have changed throughout the pandemic, as exemplified by remdesivir use during the pandemic.

Likewise, the definition of best available care has changed significantly during the pandemic and as such, efforts to account for this should be considered in the study design, such as time-stratified analyses as described earlier. A detailed definition of best available care during the study period should also be provided so that study findings can be placed into context.

Statistical analysis considerations: minimizing confounding with propensity scores

Propensity score (PS) methods are a key analytic tool for addressing systematic differences in measured baseline characteristics between exposed and unexposed groups [15]. PS can be used to balance groups using four approaches: matching, inverse probability weighting, stratification and adjustment. To ensure balance between groups, it is important that the logistic regression model used to estimate the PS is correctly specified and includes all relevant baseline variables, particularly variables that are associated with treatment allocation or are associated with both the outcome and treatment allocation [16]. Typically, established understanding and published literature can inform which variables should be included within the model, but this knowledge has not yet been fully established for COVID-19. In the early phase of the pandemic, case series were the dominant source of information but we now know that the risk of adverse COVID-19 outcomes such as hospital admissions, mechanical ventilation and death varies by demographic characteristics and specific comorbidities including obesity, hypertension and asthma [17]. This lack of understanding of key variables influencing outcomes created a substantial challenge for developing well-specified PS models. Despite many unknowns, RWD are often rich sources of information, providing robust details on patients relating to comorbidities, treatments received and demographic characteristics enabling the consideration of an exhaustive list variables for exploration and ultimate inclusion in PS models.

Another important consideration in PS model specification is to account for the temporal variations in treatment allocation, case mix and case fatality, each of which can lead to bias [19,20]. For example, according to England's Hospital Episode Statistics administrative dataset, proportions of older, frail, female and white patients admitted to hospital increased steadily between March and May 2020 as the outbreak moved out of larger cities with younger populations [18]. Potential options for minimizing the influence of these temporal variations during statistical analyses is the inclusion of calendar time of hospitalization within PS models or generation of calendar time-specific PS. In the latter approach, separate PS models are derived for each calendar time period to estimate time specific propensity of treatment receipt. The generated PS can then be used to match patient pairs within each time period to create a full study cohort.

As with any observational study, consideration must be given to the conduct of extensive sensitivity analyses to explore the robustness of the findings when different approaches to specifying and using the PS are applied [21]. For example, in a study of hydroxychloroquine in hospitalized patients with COVID-19, inverse probability weighting was used in the primary analysis but sensitivity analyses using PS matching and PS adjusted models were undertaken to confirm the robustness of the findings [22]. In stark contrast, in a study investigating a multi-drug therapy for the treatment of COVID-19 in Mexico, no methods were employed to balance the characteristics of treatment and comparison groups, despite the comparison group being older and having more comorbidities than the treatment group [23]. In this example, the multi-drug therapy was found to be more effective at reducing hospitalization and death than standard care. Given the high likelihood of confounding by indication in this study, the potential for these findings to mislead decision makers is considerable.

Statistical analysis considerations: avoiding immortal time biases

Immortal time bias is a common methodological challenge in CER and a noted challenge in studies of COVID-19 treatments [4,24]. Immortal time relates to the follow-up time during which, the outcome of interest could not occur and it typically arises through improper or imbalanced definition of the index date for each comparison group. If time between index date and treatment initiation is incorrectly classified as exposed person-time, this will result in biased estimates. Furthermore, if an event had occurred in this window (such as a death), this would preclude treatment and the event would not be counted toward the treatment group. As a result, by definition, time from index date to treatment initiation becomes immortal time for the treatment group as an event may not occur during this time. However, when treatment initiation is used as the index date for the exposed group (to eliminate

immortal time), while some other time point, such as date of hospitalization is used for the unexposed group, bias is also introduced. Several COVID-19 CER have been found to be at high risk of immortal time biases, including a study assessing the effectiveness of tocilizumab in the treatment of patients hospitalized with COVID-19 [25]. In this study, the index date for both groups was set to date of hospitalization and; therefore, time between date of hospitalization and tocilizumab initiation was immortal time in the exposed group.

To prevent immortal time biases, exposed and unexposed groups can be matched on index date. For example, in a study assessing the effectiveness of remdesivir for COVID-19 treatment, patients hospitalized with COVID-19 and who subsequently initiated remdesivir were matched to other patients hospitalized with COVID-19 who did not initiate remdesivir at the same time point [26]. Time-dependent PS were also used to derive matched sets that were similar in terms of disease severity and other potential factors influencing treatment allocation.

Considerations in the presence of regional variations in treatments

The presence of regional variations in treatment protocols is another factor that must be taken into consideration in the design and analysis of COVID-19 CER studies. These regional variations were likely caused by a lack of or differences in guideline recommendations, at least in the early phase of the pandemic, which meant that hospitals often employed their own treatment protocols. Other contributors to such regional variations include drug shortages, which meant lottery systems and first-come first-serve systems may have been put in place. Similarly, some experimental drugs may have only been available at select institutions that were participating in clinical trials. Evidence also supports the presence of regional variations in COVID-19 patient outcomes. COVID-19 mortality rates were found to vary considerably across hospitals in the US and were particularly high in hospitals with higher community incidence [27]. The latter finding may relate to a stressed healthcare system operating under the condition of a pandemic. Differences in patient outcomes across regions may also reflect differences in the age distribution and prevalence of comorbidities within the regions [28]. Despite the potential biases stemming from these regional variations, few multi-center COVID-19 studies have utilized methods to address this issue.

Design & statistical analysis considerations

These regional variations have important implications for the design of CER. With regard to the choice of study population, consideration should be given to the exclusion of hospitals in which clinical trials are underway since trial participants may be difficult to identify using administrative healthcare databases and these participants are unlikely to be comparable with patients who were not trial participants. Where sample size allows, consideration should be given to conducting studies within specific regions or hospitals to ensure homogeneity in treatment protocols and patient outcomes. Alternatively, adjustment for region in regression models or various matching approaches may be utilized to ensure regional variations in treatments and outcomes do not bias estimates of associations between treatments and outcomes [29]. Most simply, in addition to matching of treatment and comparison groups being based on the PS with a pre-specified difference, comparison groups may be force matched within the same hospital or region to mitigate confounding by region. However, this within-cluster matching strategy may lead to large numbers of patients being excluded where matches cannot be identified. To overcome this limitation, a preferential within-cluster matching may be utilized, which is a two-step matching approach. The first step identifies matches within the same 'cluster', which can be specified to be the same hospital or geographic region. For those unmatched at this first step, the next step goes beyond the 'cluster' definition to find matches. Thereby, this approach combines the advantages of a pure within- and between-cluster matching in terms of bias reduction and reducing the number of unmatched patients. Regional disparities may also be addressed in PS estimations through multilevel models with the inclusion of hospital-level or region-level covariates as fixed effects or random effects [30,31]. To account for heterogeneity in outcomes across the US, a study which used a national claims database to assess the effectiveness of metformin in reducing COVID-19-related death, geographic location of hospital was a covariate in the PS logistic regression model. Mixed effects logistic regression with state-level random effects were additionally used to account for regional differences [13].

Conclusion

During public health crises, it is essential that evidence from CER is of sufficient quality to enable decision-making that facilitates the best policy and treatment models to improve both individual and public health. Hastily conducted studies have the potential to lead researchers to flawed conclusions, with significant consequences for patients and health care delivery. However, RWD studies continue to emerge as an important source of evidence (i.e., effectiveness

and safety) on a treatment's use in clinical practice that can complement data from RCTs. When used correctly and with consideration of the potential pitfalls, CER utilizing RWD can help answer questions quickly to guide and/or reinforce treatment decisions contemporaneously. The limitations of CER are well documented and there are many nuances of conducting CER to evaluate treatments for a new condition. Best practices for conducting CER during a pandemic include a thorough exploration of the likely temporal and regional effects that may influence research findings by, for example, reporting analyses overall and stratified by calendar time and region to check the homogeneity of the findings. Rigorous methods, such as matching or inverse probability weighting should be used to address the most common source of bias in CER. Given the many unknowns of a pandemic, extensive sensitivity analyses to explore the robustness of findings are an essential component of the CER toolkit.

Author contributions

This manuscript was devised by E Mozaffari, A Chandak and R Casciano while S Read, A Khachatryan, P Hodgkins and R Haubrich made substantial contributions to the design and content. S Read drafted the manuscript, all authors contributed to the critical revision of the manuscript. All authors have read and approved the final version of the manuscript and all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Financial & competing interests disclosure

S Read, A Khachatryan, A Chandak and R Casciano are employees of Certara, which was contracted by Gilead Sciences for writing this manuscript (provided funding for this research). P Hodgkins, R Haubrich and E Mozaffari are employees and shareholders at Gilead Sciences. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Gilead Sciences funded Certara to write this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

Papers of special note have been highlighted as: ● of interest

1. Gorbalenya A, Baker S, Baric R *et al.* The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5(4), 536–544 (2020).
2. World Health Organization. WHO COVID-19 dashboard (2020). <https://covid19.who.int/>
3. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report – Part II. *Value Health* 12(8), 1053–1061 (2009).
4. Renoux C, Azoulay L, Suissa S. Biases in evaluating the safety and effectiveness of drugs for covid-19: designing real-world evidence studies. *Am. J. Epidemiol.* 190(8), 1452–1456 (2021).
- **This helpful publication provides a thorough description of additional biases, including selection bias in COVID-19 studies.**
5. Gautret P, Lagier J-C, Parola P *et al.* Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int. J. Antimicrob. Agents* 56(1), 105949 (2020).
6. Fan X, Johnson BH, Johnston SS, Elangovanraaj N, Coplan P, Khanna R. Evolving treatment patterns for hospitalized COVID-19 patients in the United States in April 2020–July 2020. *Int. J. Gen. Med.* 14, 267–271 (2021).
- **Describes the extensive changes in treatment patterns over a short period of time; therefore, clearly highlighting the need for consideration of practice changes.**
7. Vaduganathan M, Van Meijgaard J, Mehra MR, Joseph J, O'Donnell CJ, Warraich HJ. Prescription fill patterns for commonly used drugs during the COVID-19 pandemic in the United States. *JAMA* 323(24), 2524–2526 (2020).
8. Beigel JH, Tomashek KM, Dodd LE *et al.* Remdesivir for the treatment of Covid-19 – final report. *N. Engl. J. Med.* 383(19), 1813–1826 (2020).
9. RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with Covid-19. *N. Engl. J. Med.* 384(8), 693–704 (2021).
10. Thompson AE, Ranard BL, Wei Y, Jelic S. Prone positioning in awake, nonintubated patients with COVID-19 hypoxemic respiratory failure. *JAMA Intern. Med.* 180(11), 1537–1539 (2020).

11. Nguyen NT, Chinn J, Nahmias J *et al.* Outcomes and mortality among adults hospitalized with COVID-19 at US medical centers. *JAMA Netw. Open* 4(3), e210417–e210417 (2021).
12. Schultze A, Walker AJ, Mackenna B *et al.* Risk of COVID-19-related death among patients with chronic obstructive pulmonary disease or asthma prescribed inhaled corticosteroids: an observational cohort study using the OpenSAFELY platform. *Lancet Respir. Med.* 8(11), 1106–1120 (2020).
13. Bramante CT, Ingraham NE, Murray TA *et al.* Metformin and risk of mortality in patients hospitalised with COVID-19: a retrospective cohort analysis. *Lancet Healthy Longev.* 2(1), e34–e41 (2021).
14. Arshad S, Kilgore P, Chaudhry ZS *et al.* Treatment with hydroxychloroquine, azithromycin, and combination in patients hospitalized with COVID-19. *Int. J. Infect. Dis.* 97, 396–403 (2020).
15. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate. Behav. Res.* 46(3), 399–424 (2011).
- **A publication providing a comprehensive introduction to propensity scores and their implementation.**
16. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am. J. Epidemiol.* 163(12), 1149–1156 (2006).
17. Williamson EJ, Walker AJ, Bhaskaran K *et al.* Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 584(7821), 430–436 (2020).
18. Navaratnam AV, Gray WK, Day J, Wendon J, Briggs TW. Patient factors and temporal trends associated with COVID-19 in-hospital mortality in England: an observational study using administrative data. *Lancet Respir. Med.* 9(4), 397–406 (2021).
19. Bechman K, Yates M, Mann K *et al.* Inpatient COVID-19 mortality has reduced over time: results from an observational cohort. *SSRN* (2021). <https://ssrn.com/abstract=3786058>
20. Garcia-Vidal C, Cózar-Llistó A, Meira F *et al.* Trends in mortality of hospitalised COVID-19 patients: a single centre observational cohort study from Spain. *Lancet Reg. HealthEur.* 3, 100041 (2021).
21. Olender SA, Perez KK, Go AS *et al.* Remdesivir for severe COVID-19 versus a cohort receiving standard of care. *Clin. Infect. Dis.* doi: 10.1093/cid/ciaa1041 (2020) (Epub ahead of print).
22. Geleris J, Sun Y, Platt J *et al.* Observational study of hydroxychloroquine in hospitalized patients with Covid-19. *N. Engl. J. Med.* 382(25), 2411–2418 (2020).
- **Uses several approaches to balancing treatment and comparison groups and is; therefore, a strong example of good practice in conducting comparative effectiveness research during COVID-19.**
23. Lima-Morales R, Méndez-Hernández P, Flores YN *et al.* Effectiveness of a multidrug therapy consisting of ivermectin, azithromycin, montelukast, and acetylsalicylic acid to prevent hospitalization and death among ambulatory COVID-19 cases in Tlaxcala, Mexico. *Int. J. Infect. Dis.* 105, 598–605 (2021).
24. Franklin JM, Lin KJ, Gatto NM, Rassen JA, Glynn RJ, Schneeweiss S. Real-world evidence for assessing pharmaceutical treatments in the context of COVID-19. *Clin. Pharmacol. Ther.* 109(4), 816–828 (2021).
- **Provides some additional considerations relating to research questions, data availability and challenges in COVID-19 RWD studies.**
25. Guaraldi G, Meschiari M, Cozzi-Lepri A *et al.* Tocilizumab in patients with severe COVID-19: a retrospective cohort study. *Lancet Rheumatol.* 2(8), e474–e484 (2020).
- **A COVID-19 comparative effectiveness research which utilized rigorous methods for controlling for immortal time biases and confounding by indication.**
26. Garibaldi BT, Wang K, Robinson ML *et al.* Comparison of time to clinical improvement with vs without remdesivir treatment in hospitalized patients with COVID-19. *JAMA Netw. Open* 4(3), e213071–e213071 (2021).
27. Asch DA, Sheils NE, Islam MN *et al.* Variation in US hospital mortality rates for patients admitted with COVID-19 during the first 6 months of the pandemic. *JAMA Intern. Med.* 181(4), 471–478 (2021).
28. Bialek S, Bowen V, Chow N *et al.* Geographic differences in COVID-19 cases, deaths, and incidence—United States, February 12–April 7, 2020. *Morb. Mortal. Wkly Rep.* 69(15), 465 (2020).
29. Mozaffari E, Chandak A, Zhang Z *et al.* Remdesivir treatment is associated with improved survival in hospitalized patients with COVID-19. *WMEF21-2507* (2021). <https://www.abstractsonline.com/pp8/#/9286/presentation/11269>
30. Di Castelnuovo A, Costanzo S, Antinori A *et al.* Use of hydroxychloroquine in hospitalised COVID-19 patients is associated with reduced mortality: findings from the observational multicentre Italian CORIST study. *Eur. J. Intern. Med.* 82, 38–47 (2020).
31. Bartoletti M, Marconi L, Scudeller L *et al.* Efficacy of corticosteroid treatment for hospitalized patients with severe COVID-19: a multicentre study. *Clin. Microbiol. Infect.* 27(1), 105–111 (2021).