


**Research Article**

## Making Sense of Non-Randomized Comparative Treatment Studies in Times of Covid-19: A Case Study of Tocilizumab

Ruth Owen<sup>1,2</sup>, Nawab Qizilbash<sup>1,2\*</sup>, Sara Velázquez Díaz<sup>3,4</sup>, José María Castellano Vázquez<sup>3,4,5</sup>, Stuart Pocock<sup>1</sup>

### Abstract

**Background:** Tocilizumab (TCZ) is an interleukin-6 inhibitor and the second established effective drug for the treatment of hospitalized patients with Covid-19. In this study, we sought to validate the recent positive findings from the randomized clinical trial RECOVERY and to evaluate the challenges in the analysis and interpretation of non-randomized comparative effectiveness studies in Covid-19.

**Methods:** We performed a retrospective cohort study using an openly available database of hospitalized Covid-19 patients in Spain. The primary outcome was all-cause in-hospital mortality at 28 days. We used multivariable Fine and Gray competing risk models which adjusted for both fixed and time-variant confounders to investigate the effect of TCZ on the primary outcome.

**Results:** We analyzed 2547 patients hospitalized with Covid-19 between 1<sup>st</sup> January and 28<sup>th</sup> June 2020. Patients in the TCZ group tended to have more severe Covid-19 at admission, as measured by biomarkers of disease severity including CRP, D-dimer and LDH. At 28 days, 91 out of 440 TCZ patients had died compared to 267 out of 2107 patients in the control group. In multivariable analysis, there was no evidence of an association between TCZ and the primary outcome (adjusted hazard ratio 1.20, 95% CI 0.86 to 1.64, P=0.26).

**Conclusions:** Our observational study failed to find a benefit of TCZ on all-cause in-hospital mortality in Covid-19 patients compared with randomized trials, highlighting the impact that unmeasured confounding and other sources of bias can have in a retrospective observational setting. For future observational studies, we recommend prospective data collection to ensure all variables have the necessary quality, completeness and timing for reliable treatment evaluation.

**Keywords:** Suboptimal quality of real-world data can yield unreliable results in comparative treatment studies.

- Statistical methods alone are often inadequate to handle such issues.
- Our observational study failed to confirm the benefit of tocilizumab on all-cause in-hospital mortality in Covid-19 patients found in randomized trials.
- These findings underline the importance of study design and tailored, ideally prospective, data collection in observational comparative treatment studies.

### Affiliation:

<sup>1</sup>London School of Hygiene and Tropical Medicine, London, UK

<sup>2</sup>OXON Epidemiology, Madrid, Spain

<sup>3</sup>HM Hospitales Research Foundation, HM Hospitales Group, Madrid, Spain

<sup>4</sup>Faculty of Medicine, CEU San Pablo University, Madrid Spain

<sup>5</sup>National Center for Cardiovascular Research (CNIC), Instituto de Salud Carlos III, Madrid, Spain

### \*Corresponding author:

Nawab Qizilbash, London School of Hygiene and Tropical Medicine, London, UK.

**Email:** n.qizilbash@oxonepi.com

**Citation:** Ruth Owen, Nawab Qizilbash, Sara Velázquez Díaz, José María Castellano Vázquez, Stuart Pocock. Making Sense of Non-Randomized Comparative Treatment Studies in Times of Covid-19: A Case Study of Tocilizumab. *Journal of Biotechnology and Biomedicine* 7 (2024): 60-66.

**Received:** September 07, 2023

**Accepted:** September 19, 2023

**Published:** January 29, 2024

## Introduction

Tocilizumab is the second established effective drug, after systematic corticosteroids, for the treatment of hospitalized patients with Covid-19. Tocilizumab is a monoclonal antibody that inhibits interleukin -6 (IL-6) and being repurposed from its use as an anti-inflammatory agent for rheumatoid arthritis and cytokine release syndrome to Covid-19 has not been straightforward. IL-6 is cytokine released by macrophages in response to infection and stimulates inflammatory pathways. Levels of IL-6 are correlated with Covid-19 severity [1,2]. In severe Covid-19 there is vascular inflammation and dysfunction [3,4] and IL-6 promotes endothelial dysfunction and impairs vascular permeability. Tocilizumab inhibits this inflammatory process. Given the lack of effective anti-viral agents to treat severe Covid-19, clinicians tried tocilizumab, among others, with early case reports indicating clinical and biochemical improvement [5-7]. This was followed by reports of non-randomized comparative studies using retrospective data largely supporting the clinician-based impressions of benefit. This led to wider use of tocilizumab despite failure to show a survival benefit from early underpowered trials in severe Covid-19. The large RECOVERY trial has shown clear overall benefit (including reduced mortality) in hospitalized Covid-19 patients of all severities, in addition to systematic corticosteroids [8]. With knowledge of this background, we conducted an analysis on a substantial individual patient database from multiple hospitals to evaluate the issues in planning analysis and interpretation of non-randomized comparative effectiveness studies in Covid-19. Besides the results themselves our goal is to provide guidance on appropriate methodology for other non-randomized comparative analyses of therapies in hospitalized patients with Covid-19 and help to understand the discrepancies between the randomized and non-randomized comparative studies of tocilizumab.

## Methods

### Study setting and data source

We analyzed data from the Covid Data Save Lives database, an anonymized dataset [9], which included electronic health records from 2547 patients hospitalized with Covid-19 in 17 hospitals within the Grupo HM Hospitales in Spain. This clinical dataset collects the different components of the Covid-19 treatment process, including detailed information on diagnoses, treatments, admissions, ICU admissions, vital signs, laboratory results, discharge and death. Data on medical history and co-morbidities were extracted from over 5,000 diagnostic and procedural records coded according to the international ICD-10 classification in its latest distributed version [10]. Records of vital sign and laboratory analysis measurements were collected throughout admission, along with the date and time of collection. Information on

tocilizumab (TCZ) administration was obtained from the medication table, including dose, number of doses and date of administration, and was identified by brand name and ATC5/ATC7 classification. Patients were included if they had been admitted to any of the participating hospitals with a diagnosis of COVID POSITIVE or COVID PENDING between 1st January and 28th June 2020. Follow-up was from hospital admission until hospital discharge (either home or to another hospital/center) or death. This study was approved by the ethics committee of HM Hospitals (approval number 20.05.1627-GHM).

### Study endpoints

The primary endpoint was all-cause in-hospital mortality at 28 days. Patients alive were censored at the earliest of end of study date (28<sup>th</sup> June 2020), 28 days follow-up or transfer to another hospital.

### Statistical analysis

Patient characteristics at admission were summarized in the total cohort and compared in patients treated with TCZ versus those never treated with TCZ using Mann-Whitney U tests or independent t-tests for continuous variables and Fisher's exact tests or chi-squared tests for categorical variables where appropriate. Because patients discharged prior to 28 days were likely to have a substantially lower probability of death post discharge, we assumed that those discharged home were alive at 28 days by considering hospital discharge as a competing risk. The cumulative incidence function (CIF) of in-hospital mortality was plotted a) from time of admission in patients not treated with TCZ, that is patients in the control group and patients in the TCZ group censored at time of TCZ administration, and b) from time of TCZ administration in the TCZ group. Fine & Gray models [11] were used in univariable and multivariable analyses of the primary outcome to estimate subdistribution hazard ratios (sHRs). sHRs have a similar interpretation to hazard ratios, except the risk set includes all patients discharged home earlier than 28 days from hospital admission. To account for immortal time bias, TCZ was modelled as a time-updated covariate. That is, patients would transition from 'Not treated with TCZ' to 'Treated with TCZ' at the time of TCZ administration and remain in the TCZ group until the end of follow-up. Covariates considered in multivariable analysis were selected based on existing clinical knowledge and on risk factors of Covid-19 related mortality identified in previous studies [12-14] and are listed in **Supplementary Table 1** along with a summary of missing data. The distribution of each continuous variable was inspected and any variable that was highly positively skewed was log-transformed for analyses. Univariable analyses were first performed to explore the associations of patient characteristics at admission with all-cause mortality at 28 days in the dataset. A multivariable model was then fitted to investigate the effect of TCZ on 28-day mortality after

adjustment for confounding. All potential confounders were first included in the model, then removed sequentially using a stepwise method with  $p > 0.20$  as the criterion for exclusion. Both age and sex were forced into the model.

In multivariable analyses, the following variables were fitted as time-updated covariates: saturated oxygen, diastolic blood pressure (DBP), systolic blood pressure (SBP), heart rate, alanine transaminase (ALT), aspartate transaminase (AST), C-reactive protein (CRP), creatinine, d-dimer, eosinophils, glucose, lactate dehydrogenase (LDH), lymphocytes, monocytes, neutrophils, platelet count, potassium, sodium, urea, white blood cell count (WBC) and treatment with steroids. To avoid adjusting for variables on the causal pathway of the association between TCZ and mortality, these covariates were only updated up to the point of TCZ administration for patients in the TCZ group. Additionally, the multivariable model was refitted using only the baseline values of the time-updated covariates, in order to explore the impact of ignoring time-variant confounding. A subgroup analysis of CRP at admission was performed ( $CRP \geq 130$ mg/L vs  $CRP < 130$ mg/L) by including an interaction term in the multivariable model. Medical history and comorbidities were assumed to be absent if they had not been reported in the diagnostics table. The first vital sign or laboratory analysis measurement taken within 3 days of admission was considered the baseline value. For patients with no measurements reported during admission, baseline values were imputed using multiple imputation with chained equations and assumed constant throughout follow-up. The imputation models included the primary outcome, the Nelson-Aalen estimator of the cumulative hazard function and baseline variables with no missing data. Twenty imputation sets were generated and combined using Rubin's rule [15]. Statistical analyses were performed using Stata v16.1 (StataCorp LP College Station, TX, USA). A two-sided  $p$ -value of 0.05 was considered statistically significant in all analyses.

## Results

Of 2547 patients admitted to hospital with a diagnosis of Covid-19, 440 received at least one dose of TCZ. **Table 1** summarizes patient characteristics at hospital admission in the total cohort and according to treatment with TCZ. Compared to the control group, patients receiving TCZ were younger (mean age 66.5 vs 68.2 years), had lower saturated oxygen (median 94.0% vs 95.3%), were more likely to be male (71.8% vs 56.1%) and obese (11.8% vs 6.9%). They also had higher AST, CRP, D-dimer and LDH levels, all of which are markers of disease severity in Covid-19 patients. In the control group, there was higher prevalence of dementia (4.7% vs 0.9%), renal disease (7.5% vs 4.1%) and diabetes (17.8% vs 12.3%) and controls tended to have a higher Charlson comorbidity index.

Patients in the control group were followed up for a median time of 6 days (IQR 4-9) from hospital admission. Patients in the TCZ group were followed up for a median time of 14 days (IQR 9-23) from hospital admission and a median time of 11 days (IQR 6-18) from TCZ administration. The median time to TCZ administration from hospital admission was 3 days (IQR 1-4) and the majority of TCZ patients received one dose (79.8%), with the remaining patients receiving two. At 28 days follow-up, a total of 358 patients (14.1%) had died, 1893 (74.3%) had been discharged home, 141 (5.5%) had been transferred to another hospital or center and 155 (6.1%) patients remained in hospital. The proportion of deaths at 28 days follow-up was higher in the TCZ group, 91 patients of whom died (20.7%) compared to the control group, 267 patients of whom died (12.7%).

**Figure 1A** shows the cumulative mortality from time of hospital admission in patients not treated with TCZ or up to the point of treatment, which is relatively constant for the first 9 days then appears to decline for the remainder of follow-up. However, this should be interpreted with caution as the censoring of patients at time of TCZ administration may lead to an over-optimistic estimate of survival due to the selection of more severe patients for treatment with TCZ. In contrast, there is a fairly steady increase in mortality from time of tocilizumab administration to 28 days follow-up in patients treated with tocilizumab (**Figure 1B**).

The univariable associations between each potential confounder and the primary endpoint are shown in Supplementary Table 2. In the unadjusted analysis, TCZ was associated with a higher risk of mortality (sHR 2.35, 95% CI 1.86 to 2.98,  $P < 0.001$ , Table 2). Time-invariant confounders adjusted for in the multivariable model were age, sex, oxygen at admission, treatment with steroids, CHF, ischemic heart disease, pulmonary disorder, cancer and PVD. Time-updated covariates were saturated oxygen, heart rate, ALT, AST, CRP, glucose, LDH, lymphocytes, monocytes, platelet count, sodium, urea and WBC. After adjustment for time-invariant confounding and baseline values of time-variant confounders, the harmful effect of TCZ decreased in magnitude but remained statistically significant (sHR 1.92, 95% CI (1.42 to 2.60),  $P < 0.001$ ). After adjustment for time-variant and -invariant confounding the effect of TCZ was no longer statistically significant (sHR 1.20, 95% CI 0.86 to 1.64,  $P = 0.26$ ). Subgroup analysis showed some evidence of a difference in the effect of TCZ according to CRP at admission ( $P$  for interaction = 0.03), suggesting a detrimental effect of TCZ in 1906 patients with a  $CRP < 130$  at hospital admission (sHR 1.66, 95% CI (1.11 to 2.48)), whereas there was no significant effect of TCZ in 641 patients with  $CRP \geq 130$  (sHR 0.93, 95% CI (0.63 to 1.37), Supplementary Figure 1).

**Table 1:** Patient characteristics at hospital admission.

| Variable                                | All patients         | TCZ                  | No TCZ               | P       |
|-----------------------------------------|----------------------|----------------------|----------------------|---------|
|                                         | N=2547               | N=440                | N=2107               |         |
| Age (years), mean (SD)                  | 67.9 (16.7)          | 66.5 (12.9)          | 68.2 (17.4)          | 0.02    |
| Female, N (%)                           | 1050 (41.2)          | 124 (28.2)           | 926 (43.9)           | <0.0001 |
| Confirmed Covid-19 diagnosis, N (%)     | 2317 (91.0)          | 438 (99.5)           | 1879 (89.2)          | <0.0001 |
| Supplemental oxygen, N (%)              | 2067 (81.2)          | 321 (73.0)           | 1746 (82.9)          | <0.0001 |
| Treatment with steroids, N (%)          | 1140 (44.8)          | 404 (91.8)           | 736 (34.9)           | <0.0001 |
| Temperature (c), mean (SD)              | 36.6 (0.8)           | 36.9 (0.9)           | 36.6 (0.7)           | <0.0001 |
| Heart rate (bpm), mean (SD)             | 85.0 (16.1)          | 87.7 (15.5)          | 84.5 (16.2)          | 0.0002  |
| Saturated oxygen (%), median (IQR)      | 95.0 (93.0-97.0)     | 94.0 (90.0-95.7)     | 95.3 (94.0-97.0)     | <0.0001 |
| SBP (mmHg), mean (SD)                   | 128.2 (20.7)         | 128.0 (19.2)         | 128.3 (21.0)         | 0.78    |
| DBP (mmHg), mean (SD)                   | 73.4 (12.4)          | 73.7 (11.9)          | 73.3 (12.4)          | 0.52    |
| Charlson comorbidity index, N (%)       |                      |                      |                      | 0.01    |
| 0                                       | 1466 (57.6)          | 279 (63.4)           | 1187 (56.3)          |         |
| 1                                       | 572 (22.5)           | 87 (19.8)            | 485 (23.0)           |         |
| 2                                       | 258 (10.1)           | 43 (9.8)             | 215 (10.2)           |         |
| 3                                       | 126 (4.9)            | 12 (2.7)             | 114 (5.4)            |         |
| 4+                                      | 125 (4.9)            | 19 (4.3)             | 106 (5.0)            |         |
| MI, N (%)                               | 59 (2.3)             | 10 (2.3)             | 49 (2.3)             | 0.95    |
| CHF, N (%)                              | 145 (5.7)            | 19 (4.3)             | 126 (6.0)            | 0.17    |
| PVD, N (%)                              | 76 (3.0)             | 12 (2.7)             | 64 (3.0)             | 0.73    |
| Cerebrovascular disease, N (%)          | 100 (3.9)            | 13 (3.0)             | 87 (4.1)             | 0.25    |
| Dementia, N (%)                         | 104 (4.1)            | 4 (0.9)              | 100 (4.7)            | 0.0002  |
| Chronic Pulmonary Disease, N (%)        | 352 (13.8)           | 59 (13.4)            | 293 (13.9)           | 0.78    |
| Rheumatic disease, N (%)                | 42 (1.6)             | 4 (0.9)              | 38 (1.8)             | 0.22    |
| Renal disease, N (%)                    | 175 (6.9)            | 18 (4.1)             | 157 (7.5)            | 0.01    |
| Hypertension, N (%)                     | 1140 (44.8)          | 190 (43.2)           | 950 (45.1)           | 0.46    |
| Diabetes, N (%)                         | 428 (16.8)           | 54 (12.3)            | 374 (17.8)           | 0.005   |
| Cancer, N (%)                           | 194 (7.6)            | 39 (8.9)             | 155 (7.4)            | 0.28    |
| Liver disease, N (%)                    | 61 (2.4)             | 13 (3.0)             | 48 (2.3)             | 0.4     |
| Prior stroke, N (%)                     | 45 (1.8)             | 7 (1.6)              | 38 (1.8)             | 0.76    |
| Ischaemic heart disease, N (%)          | 163 (6.4)            | 28 (6.4)             | 135 (6.4)            | 0.97    |
| Asthma, N (%)                           | 122 (4.8)            | 21 (4.8)             | 101 (4.8)            | 0.99    |
| Obesity, N (%)                          | 197 (7.7)            | 52 (11.8)            | 145 (6.9)            | 0.0004  |
| ALT (U/L), median (IQR)                 | 25.2 (16.0-42.0)     | 31.6 (20.0-48.7)     | 23.8 (15.0-40.0)     | <0.0001 |
| AST (U/L), median (IQR)                 | 31.3 (22.0-49.0)     | 42.0 (29.6-59.8)     | 29.2 (21.0-45.0)     | <0.0001 |
| Creatinine (mg/dL), median (IQR)        | 0.90 (0.71-1.10)     | 0.94 (0.77-1.11)     | 0.89 (0.71-1.10)     | 0.005   |
| CRP (mg/L), median (IQR)                | 65.0 (23.8-130.7)    | 116.8 (64.8-209.6)   | 53.5 (18.3-113.9)    | <0.0001 |
| D-dimer (ng/mL), median (IQR)           | 735.5 (445.0-1419.0) | 823.0 (525.0-1346.0) | 722.0 (431.0-1439.0) | 0.01    |
| Eosinophils (x10e3/uL), median (IQR)    | 0.01 (0.00-0.05)     | 0.00 (0.00-0.02)     | 0.01 (0.00-0.07)     | <0.0001 |
| Glucose (mg/dl), median (IQR)           | 113.1 (99.7-136.5)   | 119.0 (105.0-142.7)  | 112.0 (98.3-134.0)   | <0.0001 |
| LDH (U/L), median (IQR)                 | 522.0 (396.0-679.1)  | 662.2 (536.8-851.0)  | 493.4 (380.5-636.6)  | <0.0001 |
| Lymphocytes (x10e3/uL), median (IQR)    | 1.1 (0.8-1.5)        | 0.8 (0.6-1.2)        | 1.1 (0.8-1.5)        | <0.0001 |
| Monocytes (x10e3/uL), median (IQR)      | 0.5 (0.3-0.7)        | 0.4 (0.3-0.6)        | 0.5 (0.3-0.7)        | <0.0001 |
| Neutrophils (x10e3/uL), median (IQR)    | 4.8 (3.3-6.9)        | 5.6 (3.8-8.4)        | 4.6 (3.2-6.6)        | <0.0001 |
| Platelet count (x10e3/uL), median (IQR) | 206 (160-269)        | 199 (155-256)        | 208 (161-272)        | 0.14    |
| Potassium (mmol/L), mean (SD)           | 4.2 (0.6)            | 4.2 (0.5)            | 4.3 (0.6)            | 0.006   |
| Sodium (mmol/L), mean (SD)              | 137.2 (4.8)          | 135.9 (4.0)          | 137.6 (4.9)          | <0.0001 |

|                                           |                  |                  |                  |       |
|-------------------------------------------|------------------|------------------|------------------|-------|
| <b>Urea (mg/dL), median (IQR)</b>         | 35.0 (26.0-50.0) | 35.8 (26.9-48.6) | 34.8 (26.0-50.3) | 0.95  |
| <b>WBC count (x10e3/uL), median (IQR)</b> | 6.6 (5.0-8.9)    | 7.0 (5.3-10.1)   | 6.6 (5.0-8.7)    | 0.003 |

**Abbreviations:** ALT – Alanine Transaminase, AST – Aspartate Transaminase, CHF – Congestive Heart Failure, CRP – C-reactive Protein, DBP – Diastolic Blood Pressure, IQR – Interquartile Range, LDH – Lactate Dehydrogenase, MI – Myocardial Infarction, PVD – Peripheral Vascular Disease, SBP – Systolic Blood Pressure, SD – Standard Deviation, TCZ – tocilizumab, WBC – White Blood Cell Count

**Table 2:** Unadjusted and adjusted effect of tocilizumab on 28-day all-cause mortality

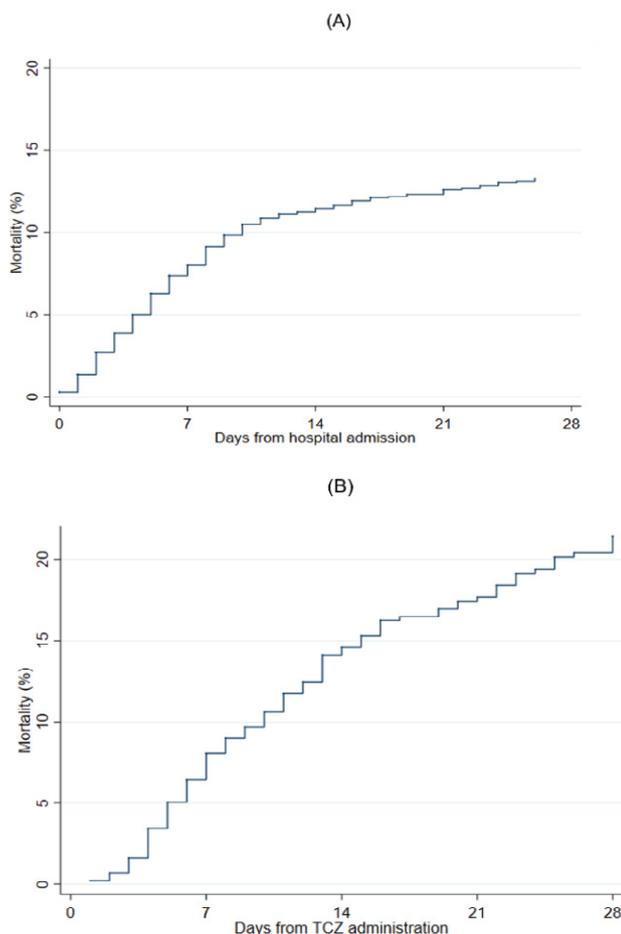
|           | TCZ          |            | No TCZ       |            | Unadjusted                |        | Adjusted for time invariant confounding <sup>2</sup> |        | Adjusted for time invariant and variant confounding <sup>3</sup> |      |
|-----------|--------------|------------|--------------|------------|---------------------------|--------|------------------------------------------------------|--------|------------------------------------------------------------------|------|
|           | No. Patients | No. deaths | No. patients | No. deaths | sHR <sup>1</sup> (95% CI) | P      | sHR <sup>1</sup> (95% CI)                            | P      | sHR <sup>1</sup> (95% CI)                                        | P    |
| Mortality | 440          | 91         | 2107         | 267        | 2.35 (1.86, 2.98)         | <0.001 | 1.92 (1.42, 2.60)                                    | <0.001 | 1.20 (0.86, 1.64)                                                | 0.26 |

**Abbreviations:** CI – Confidence interval, sHR – Sub distribution Hazard Ratio, TCZ – Tocilizumab

1 Sub distribution Hazard ratios were estimated using Fine & Gray models which considered discharge from hospital prior to 28 days as a competing risk

2 Adjusted for age, sex, oxygen at admission, treatment with steroids, CHF, ischaemic heart disease, pulmonary disorder, cancer and PVD and for baseline values at hospital admission of saturated oxygen, heart rate, ALT, AST, CRP, glucose, LDH, lymphocytes, monocytes, platelet count, sodium, urea and WBC.

3 Adjusted for age, sex, oxygen at admission, treatment with steroids, CHF, ischaemic heart disease, pulmonary disorder, cancer and PVD as time-invariant covariates and for saturated oxygen, heart rate, ALT, AST, CRP, glucose, LDH, lymphocytes, monocytes, platelet count, sodium, urea and WBC as time-varying covariates.



**Abbreviations:** TCZ-Tocilizumab

**Figure 1:** Cumulative incidence of mortality (A) from time of hospital admission in patients not treated with tocilizumab or up to the point of treatment and (B) from time of tocilizumab administration in 440 patients treated with tocilizumab.

## Discussion

Our non-randomized comparative analysis did not find the beneficial effect of TCZ established by the very recent large RECOVERY randomized trial. While the 95% confidence intervals of our analysis overlapped with that of RECOVERY, the point estimate did not fall within the 95% CI of the RECOVERY trial’s estimate. We found an apparent significant excess risk of death with TCZ in an unadjusted analysis but this attenuated to a much lesser non-significant excess after adjustment for many variables at baseline and further adjustment for time-varying covariates. The differences in the results from the fully adjusted multivariable model and the model with only adjustment for baseline values highlights the danger of ignoring time-variant confounding. However, the way to account for this type of confounding is not always clear and can require sophisticated statistical methods, resulting in effect estimates that are difficult to interpret. We took care to avoid adjustment of variables on the causal pathway of the association of interest. Allowance for missing data using multiple imputation made little difference. We did not find convincing evidence that the effect of TCZ was beneficial in patients with a high CRP as a marker of more severe disease. Allowing for the beneficial effect of corticosteroids made negligible difference. Our findings also differ from three other non-randomized studies of TCZ at moderate risk of bias [16-18] which found more extreme beneficial effects than RECOVERY. Our findings are similar to the unadjusted analysis on this same but smaller and earlier dataset [19]. We did not confirm the subgroup finding of benefit in patients with high CRP [19] which has also not been confirmed by other observational studies and the RECOVERY trial used CRP ≥ 75 mg/L as an inclusion criterion, with a median value of 143 (IQR: 107-207) in the

randomized population. TCZ was administered to patients with severe Covid-19 based on the mortality rate of the group (and comparable to other observational studies and the severely ill subgroup in RECOVERY). The reasons for these discrepancies are difficult to identify. Our population was reasonably large and included all, unselected hospitalized patients with moderate and severe Covid-19. RECOVERY found benefits in both these groups, dispelling earlier claims of the beneficial effects of TCZ only in severe or critically ill Covid-19 [20]. The timing of administration of TCZ was soon after admission, without huge variation and all patients received mostly one or two doses on consecutive days, so unlikely to be factors affecting the results. Many covariates were included in our model comparable to other non-randomized studies and most established risk factors were confirmed in this dataset. Our study has some limitations. Only death was analyzed and other endpoints such as use of ICU and mechanical ventilation are important for hospital resource utilization in times of a pandemic. However, all-cause death is the outcome with least scope for misclassification and no missing data. We adjusted for competing risk. With sufficient data, if this endpoint cannot provide a valid finding other endpoints used in observational studies [16-18] are open to greater difficulties in analysis.

Other repurposed drugs which were used in the hospitals of the dataset were not included with the exception of corticosteroids but there is no good evidence for their beneficial effect (SOLIDARITY). We did not use propensity scores for adjustment but our statistical models using multivariable adjustment were appropriate and there is little evidence that propensity score adjustments would give different results [21]. The quality of the data did not allow accurate analyses of the timing of administration of TCZ in relation to mechanical ventilation or date of ICU admission. Clinical severity was not always recorded or assessable at baseline or the time of administration of TCZ, however, the analysis used laboratory markers of severity such as CRP and D-dimer and LDH. During the worst weeks of the pandemic from which this dataset arises, comorbidities at admission were not comprehensively collected, with some chronic conditions being recorded after admission. While many of the chronic conditions recorded after admission were ascribed to being present at baseline for analysis, the multivariable analysis of baseline variables may have been suboptimal, allowing for considerable residual confounding. Some key variables were absent such as smoking and BMI, and obesity defined clinically was present in only a minority of patients, again leading to potential residual confounding. Important time-varying covariates were collected or recorded in a non-uniform, real-world fashion and this may have led to suboptimal post-baseline adjustments. Time since symptom onset was also not recorded and there were a lot of missing data. While negative controls for both exposure and endpoints have been proposed [22], we had a gold standard reference

of a large reliable randomized trial with precise estimates containing large clinical subgroups for severity of Covid-19 and use of corticosteroids. For the same reason, we did not calculate an E-value, a quantitative bias analysis to assess the strength of association between an unmeasured confounder and the exposure or outcome, conditional on measured covariates that would be necessary to fully explain observed effects [23].

In conclusion, our study failed to find a benefit of TCZ on all-cause in-hospital mortality in Covid-19 compared with randomized trials, which we believe is due to a combination of residual confounding and suboptimal adjustment with timevarying covariates, largely due to deficiencies in the quality of data. Other non-randomized comparative studies at moderate risk of bias found exaggerated benefits compared with randomized trials. Thus, many non-randomized comparative studies of treatments using retrospective data appear unreliable in hospitalized Covid-19. For future observational studies, we recommend tailored, ideally, prospective data collection to ensure all variables have the necessary quality, completeness and timing for reliable treatment evaluation.

## References

1. Aziz M, Fatima R, Assaly R. Elevated interleukin-6 and severe COVID-19: A meta-analysis. *J Med Virol* 92 (2020): 2283-2285.
2. Zhu J, Pang J, Ji P, et al. Elevated interleukin-6 is associated with severity of COVID-19: A meta-analysis. *J Med Virol* 93 (2021): 35-37.
3. Ackermann M, Verleden SE, Kuehnel M, et al. Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *N Engl J Med* 383 (2020): 120-128.
4. Varga Z, Flammer AJ, Steiger P, et al. Endothelial cell infection and endotheliitis in COVID-19. *Lancet* 395 (2020): 1417-1418.
5. Antwi-Amoabeng D, Kanji Z, Ford B, et al. Clinical outcomes in COVID-19 patients treated with tocilizumab: An individual patient data systematic review. *J Med Virol* 92 (2020): 2516-2522.
6. Xu X, Han M, Li T, et al. Effective treatment of severe COVID-19 patients with tocilizumab. *Proc Natl Acad Sci USA* 117 (2020): 10970-10975.
7. Zhang X, Song K, Tong F, et al. First case of COVID-19 in a patient with multiple myeloma successfully treated with tocilizumab. *Blood Adv* 4 (2020): 1307-1310.
8. Horby PW, Pessoa-Amorim G, Peto L, et al. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): preliminary results of a randomized, controlled, open-label, platform trial (2021).

9. HM Hospitales. Covid Data Saves Lives [Available from: <https://www.hmhospitales.com/coronavirus/covid-data-save-lives/english-version>].
10. World Health Organization. The ICD-10 classification of mental and behavioral disorders: clinical descriptions and diagnostic guidelines. World Health Organization (1992).
11. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* 94 (1999): 496-509.
12. Gupta S, Hayek SS, Wang W, et al. Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US. *JAMA Intern Med* (2020).
13. Malik P, Patel U, Mehta D, et al. Biomarkers and outcomes of COVID-19 hospitalisations: systematic review and meta-analysis. *BMJ Evid Based Med* (2020).
14. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 584 (2020): 430-436.
15. Rubin D. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons (1987).
16. 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* 2 (2020): e603-e612.
17. Gupta S, Wang W, Hayek SS, et al. Association Between Early Treatment With Tocilizumab and Mortality Among Critically Ill Patients With COVID-19. *JAMA Intern Med* 181 (2021): 41-51.
18. Ruiz-Antorán B, Sancho-López A, Torres F, et al. Combination of Tocilizumab and Steroids to Improve Mortality in Patients with Severe COVID-19 Infection: A Spanish, Multicenter, Cohort Study. *Infect Dis Ther* 10 (2021): 347-362.
19. Martínez-Sanz J, Muriel A, Ron R, et al. Effects of tocilizumab on mortality in hospitalized patients with COVID-19: a multicentre cohort study. *Clin Microbiol Infect* 27 (2021): 238-243.
20. Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19. *New England Journal of Medicine* (2021).
21. 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* 69 (2017): 345-357.
22. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 21 (2010): 383-388.
23. Vander Weele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med* 167 (2017): 268-274.