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.

Clinical Microbiology and Infection 28 (2022) 371–374

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Narrative review

Signals were broadly positive for months, but never definitive: the tocilizumab story

Alessandro Cozzi-Lepri [1,*], Colette Smith [1], Cristina Mussini [2]

[1)] *Centre for Clinical Research, Epidemiology, Modelling and Evaluation (CREME), Institute for Global Health, UCL, London, UK*
[2)] *Department of Infectious Diseases, Azienda Ospedaliero-Universitaria Policlinico of Modena, Modena, Italy*

ARTICLE INFO

Article history:
Received 19 August 2021
Received in revised form
25 October 2021
Accepted 25 October 2021
Available online 9 November 2021

Editor: L Leibovici

Keywords:
Covid-19
Evidence-based medicine
Meta-analysis
Propensity score methods
Tocilizumab

ABSTRACT

Background: Most treatment guidelines for coronavirus disease 2019 (COVID-19) currently recommend tocilizumab in combination with dexamethasone in critically ill patients who are exhibiting rapid respiratory decompensation.
Aims: To produce a critical review and summary of the pathway which led to the repurposing of tocilizumab for COVID-19 treatment, from *in vitro* observations to guidelines recommendations.
Sources: All studies evaluating the effectiveness of tocilizumab to treat COVID-19 disease published between July 2020 and July 2021.
Content: Two large and methodologically well conducted observational studies, the TESEO and the STOP COVID cohorts, showed a reduction in the risk of invasive mechanical ventilation or death in patients treated with tocilizumab as compared to standard of care in 2020. Concomitantly, and up to February 2021, a number of randomized trials (RCTs) with small sample sizes were showing discrepant results. These RCTs had a number of issues: small sample size, various designs and inclusion criteria, and different dosages of tocilizumab used. The confidence interval of the meta-analytic estimate for the RCT results was consistent with the hypothesis of no efficacy of tocilizumab. In our opinion, this was mainly because the meta-analysis included small and heterogeneous studies. These results led to a delay in the inclusion of tocilizumab in guidelines which occurred only in the summer of 2021.
Implications: Although observational studies are unable to control for unmeasured confounding, they can be put together quickly during a pandemic and promptly provide important information. The large sample size allows us to investigate effect measure modifiers and to better target interventions. It is key that the effect size is somewhat large (RR > 2), all sources of bias are properly accounted for, and the direct evidence is weighted against these factors. It appears to us that for tocilizumab, not having dismissed the results of *carefully designed and analysed* observational studies in 2020 could have prevented many deaths over those months. **Alessandro Cozzi-Lepri, Clin Microbiol Infect 2022;28:371**

The clinical picture

The standard of care of patients with severe coronavirus disease 2019 (COVID-19) pneumonia changed dramatically during 2020, whilst treating physicians balanced the urgent need to lower the mortality risk with quickly accumulating evidence on the efficacy of candidate clinical interventions.

Study designs to assess the efficacy of interventions

Randomized clinical trials (RCTs) are considered the reference standard for evaluating the efficacy and safety of medical interventions. Although the application of this rigorous strategy, which is valid in general, becomes even more relevant during a pandemic, it can be a challenge to rapidly organize and conduct RCTs. Conversely, collection of real-world data from routine clinical practice is relatively simpler and, by using modern techniques of data capture, can be done almost in real time during a pandemic. In this scenario, the key question still holds: to what extent can we

* Corresponding author. Alessandro Cozzi-Lepri, Medical Statistics and Epidemiology, Centre for Clinical Research, Epidemiology, Modelling and Evaluation (CREME), Institute for Global Health, UCL, Rowland Hill St, London, NW3 2PF, UK.
E-mail address: a.cozzi-lepri@ucl.ac.uk (A. Cozzi-Lepri).

https://doi.org/10.1016/j.cmi.2021.10.018
1198-743X/

trust evidence from non-randomized studies, in particular observational studies using real-world data?

By strictly applying the evidence-based medicine (EBM) hierarchy, only by conducting high-quality RCTs (using concealment of allocation, blinding, adequate power, no loss to follow-up, etc.) we can be certain that people receiving the intervention are as similar as possible to those who did not; thus, any difference in outcome can be ascribed to the intervention.

The role of observational studies

In contrast, if evidence of high effectiveness of a certain intervention has been obtained from observational studies alone, these treatments are typically not licensed for routine use, since the interventions are believed to be supported by relatively poor evidence. Indeed, results of observational studies are unlikely to trigger Food and Drug Administration (FDA) emergency use approvals (EUA). Observational studies are vulnerable to a number of factors, including the presence of confounding (both measured and unmeasured) and differential levels of clinical monitoring by intervention arm, which could bias the estimate of the effect of the intervention. Specifically, randomized studies are placed higher in the hierarchy of EBM because bias due to unmeasured confounding is minimized. Therefore, over the last two decades an increasing number of epidemiologists and statisticians have advocated the use of propensity score adjustments and marginal structural models to at least properly control for all sources of measured confounding. If the objective is to provide an estimate of the causal effect of an intervention which is as close as possible to the estimate that could be obtained in such hypothetical randomized clinical trials, all other sources of bias being equal, marginal structural models represent the best available tool to date to make valid inference from observational data [1–3].

The tocilizumab story

Tocilizumab is an interleukin-6 (IL-6) receptor antagonist. As elevated levels of IL-6 consistently predicted both severe prognosis and mortality, there was a strong rationale for using this compound in COVID-19 disease [4]. The drug was originally indicated for the treatment of rheumatoid arthritis, and more recently for the treatment of cytokine-release syndrome secondary to chimeric antigen receptor T-cell therapy (CAR-T) [5]. The fact that the cytokine storm described during severe COVID-19 could be considered similar to that occurring after CAR-T, and the availability of data on the tolerability of the drug in humans with rheumatoid arthritis and CAR-T, led clinicians—initially in China, then by Europe and the USA—to administer tocilizumab to COVID-19 patients [6]. Quickly, real-world data were put together in order to gather evidence on the effectiveness of the drug in reducing the risk of invasive mechanical ventilation and death [7,8]. These studies were conducted with very variable inclusion criteria, ranging from patients with very mild disease to patient populations who were admitted to the ICU. Among these, we want to highlight two studies: one in Italy (the TESEO cohort, our group) and one in the USA (the STOP-COVID cohort based in 68 tertiary hospitals) [7,8]. Both studies included critically ill patients and used a marginal structural model with inverse probability weights to control for measured imbalanced characteristics of patients who were treated with tocilizumab compared to those who were not.

The TESEO cohort is a large multicentre observational study of 544 patients with COVID-19 admitted to three tertiary hospitals in the Emilia Romagna region during the first wave of the pandemic [7]. The study showed high effectiveness of tocilizumab compared to standard of care for reducing the risk of both invasive mechanical ventilation and death. At day 14 after hospital admission, the proportion of patients with this composite outcome was 23% (95%CI 16–29%) for the tocilizumab group versus 37% (31–42%) for the standard-of-care group (log rank p 0.0023). From fitting a marginal structural Cox regression model, the adjusted hazard ratio (aHR) was 0.53 (95%CI 0.31–0.89), thus showing high effectiveness of tocilizumab in reducing the risk of these events. The effect of tocilizumab was even greater in people with a baseline PaO_2/FiO_2 value of <150 mmHg (aHR 0.19, 95%CI 0.08–0.44), suggesting that the drug might be particularly indicated for critically ill patients. The article was published in August 2020 [7]. These results were replicated a few months later on new data in a completely different geographical setting in the STOP COVID cohort, published in November 2020 [8]. Here, 28-day mortality risks were 28% in the tocilizumab group versus 37% in the non-tocilizumab group, with an aHR for death of 0.71 (95%CI 0.56–0.92). Again, the effect was stronger in the subset of patients who were admitted to the ICU within 3 days of symptoms onset (aHR 0.41, 95%CI 0.23– 0.74) and largely attenuated in those admitted to the ICU >3 days after symptom onset. TESEO and STOP COVID also shared the same type of imbalance observed between the two treatment strategies, with patients treated with tocilizumab typically having more comorbidities, higher prevalence of hypoxaemia and higher levels of inflammatory markers.

After August 2020, randomized studies started to be published. The trials were highly heterogeneous in inclusion criteria and design. For example, the CORIMUNO TOCI study, RCT-TCZ-COVID-19 study and Brazilian COVID-19 trials were not double-blind and did not have placebo controls [9–11], and only REMAP CAP included mainly critically ill patients [12]. Exact dosing of tocilizumab also varied by trial. In the RECOVERY, EMPACTA and REMAP-CAP [12–14] patients received one dose of tocilizumab that could be repeated after 12 or 24 hours on the basis of clinical evaluation, while in the COVACTA just one dose was used, and in the RCT-TCZ-COVID-19 always two doses were used [10,15].

By the summer of 2020, because of the conflicting nature of the results of these initital trials, tocilizumab was excluded in many countries from the list of recommended regimens for the treatment of COVID-19 disease, and its use was recommended against outside of clinical trials.

In early January 2021 the results of the REMAP CAP trials were published, and there was a need to resynthesize all the evidence coming from RCTs [12]. The best way to pool all the information together is to conduct a meta-analysis including all trials regardless of design features and inclusion criteria. This was indeed the approach used by the MRC Population Health Research Unit (https://twitter.com/rupert_pearse/status/1349424862876594179/photo/1). Such a meta-analysis indicated an effect of tocilizumab, albeit of small magnitude, in reducing the risk of 28-day mortality (meta-analytic OR 0.83, 95%CI 0.66–1.04, $p = 0.11$). As a consequence, the drug was seen more positively, although more definitive evidence was needed to recommend its routine use, as reported in a commentary published at the time in the BMJ [16].

However, looking at some of the trials included in the meta-analysis individually, the study conducted by Salvarani et al. included only patients with mild disease severity with a median PaO_2/FiO_2 ratio of 265 mmHg [10]. This is clearly a population with no ongoing cytokine storm for whom tocilizumab should not be indicated. Second, the sample size of the BAAC Bay trial [17] was small, and because of the specific wording used in the conclusion paragraph of the abstract—i.e. *"tocilizumab was not effective from preventing intubation or death"*—its inconclusive evidence has been widely interpreted as 'no effect of the drug'.

Despite recent efforts to better teach statistical hypothesis testing and avoid p-value misconceptions, unfortunately it is not uncommon to confuse the concept of 'no statistical significance' with that of 'no effect' [18].

Importantly, when restricting to trials with >150 participants per arm and predominantly including patients with more severely compromised respiratory function at entry (EMPACTA, CORIMUNO and REMAP CAP) there was strong evidence that tocilizumab reduced the risk of invasive mechanical ventilation or death by a remarkable 40% [15,9,12]. In March 2021 the results of the largest trial conducted to date, the RECOVERY trial, were finally publicly reported and confirmed a reduction in mortality (RR 0.86, 95%CI 0.77−0.96, p = 0.007) in patients treated with dexamethasone plus tocilizumab compared to those treated with dexamethasone alone [13]. On the basis of these results, several guidelines—including the Italian Society of Infectious diseases, UK, NIH and IDSA—were eventually modified to recommend the use of tocilizumab in critically ill patients [19−22]. Interestingly, along the lines of our arguments, Lawrence et al. recently proposed that ideally meta-analyses should be conducted using individual patient's (IP) data rather than an assemblage of summary statistics. Indeed, to cite a recent example involving another repurposed drug, most of the flaws identified in trials of ivermectin would have been immediately detected in an IP meta-analysis. Such an approach would also have led to a better assessment of the effect measure modifiers in these studies [23].

Indeed, we should also note at this point that other drugs (such as hydroxychloroquine, azithromycin and ivermectin) appeared to be promising to treat COVID-19 disease during the early phase of the pandemic. However, we do not think that the pathway from observational studies to RCTs for these drugs is comparable to that we describe here for tocilizumab. Briefly, hydroxychloroquine showed promising reductions of SARS Cov-2 replication in *in vitro* studies, but we are not aware of high-quality observational studies in humans. Indeed, one initial study lacked a control group and others were criticized for being affected by confounding and immortal-time bias [26,27]. Similarly, invermectin's use was advocated following the results of a meta-analysis which was later retracted by the authors due to the inclusion of at least two studies which had been poorly designed, not peer-reviewed, or affected by clear bias [28-30].

Conclusions

Classical principles of EBM are not under scrutiny here. High-quality evidence of the safety and efficacy of new interventions is vital. It is indeed reasonable to always treat results of observational studies cautiously because of the issue, among others, of unmeasured confounding. It is also reasonable not to rely on the results of single randomized trials in isolation, and to ensure that meta-analyses consider differences in study designs and patient populations before pooling the results [23].

Nevertheless, there are anomalies in this 'tocilizumab story' which seem to suggest that perhaps more caution is required when performing some of these steps, especially while we are working under the pressure of a devastating pandemic.

First, the interpretation of some of the early small individual trials has been incorrect or misleading. In addition, the meta-analyses of these trials performed in January 2021 focused on a mortality endpoint only. Although death is certainly a more solid and less subjective endpoint, this choice appears to be debatable because saturation of ICUs was one of the most critical issues, especially during the first peak of the pandemic. Further, over the same time period, other candidate pharmaceutical interventions have passed through the pathway from *in vitro* analysis through observational studies, randomized trials and subsequent trigger of FDA EUA and inclusion in treatment guidelines. These include antivirals, which were included in treatment guidelines without evidence of a difference with respect to mortality [21,22].

In parallel with the development and increasing use in the research community of methods based on propensity scores, Howick et al. have proposed a revised form of the Bradford Hill's viewpoints to establish causality in epidemiology [24]. This suggests that randomized and non-randomized studies should no longer be separated but instead classified together as study designs providing 'direct evidence' for the effectiveness of an intervention. According to Howick et al., the accumulation of 'direct' evidence demonstrating that the effect size is greater than the combined influence of plausible confounders and other potential biases is more important than the actual study design (experimental versus observational) [24]. In this particular case, patients treated with tocilizumab in observational studies were on average more critically ill than those who did not receive the drug. Thus, if anything, the effect of tocilizumab could have even been underestimated in these studies. Importantly, our key point is that *carefully designed and analysed* observational studies can also play a key role in advancing our knowledge of treating COVID-19. In a recent study by Shepshelovich et al., for COVID-19 treatment comparisons, a large discrepancy between results of observational studies and trials was shown, although 30% of the observational studies reported only crude mortality rates (univariable analyses without controlling for confounding) and only 55% used propensity adjustment analysis [25]. In addition, only eight non-randomized studies (8%) contained any type of adjustment for immortal time bias.

Taking all these arguments together, in a scenario of particularly severe outcomes occurring on a rapid timescale, is it sensible to withdraw or recommend against the use of a promising intervention, despite the fact that has been shown in well-conducted non-randomized studies to reduce mortality by as much as 40%?

Indeed, many patients admitted during the second wave in 2020 and early 2021 could not have access to tocilizumab because of guideline recommendations. Crucially, many lives could have been saved by introducing tocilizumab in routine clinical practice as early as August 2020 (the date of publication of the TESEO study) or even in November 2020 (pre-print of the STOP COVID study), and this is certainly something which regulatory agencies, researchers, infectious disease clinicians and the community as a whole should be reflecting upon.

Author contributions

AC-L: data analysis, data interpretation, writing and revising for intellectual content. CS: writing and revising for intellectual content. CM: writing and revising for intellectual content.

Transparency declaration

None of the authors declares any conflicts of interest. No external funding was received for this work.

Acknowledgements

The authors wish to show their appreciation to Professor Miguel Hernan for inspiration and sharing his results from the STOP COVID collaboration cohort (https://www.lshtm.ac.uk/media/44276). ACL work has been co-funded by the EuCARE project funded by the EU under the HORIZON Europe programme, GA n.6944397.

References

[1] Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550–60.
[2] Dahabreh IJ, Sheldrick RC, Paulus JK, Chung M, Varvarigou V, Jafri H, et al. Do observational studies using propensity score methods agree with randomized trials? A systematic comparison of studies on acute coronary syndromes. Eur Heart J 2012;33:1893–901.
[3] Lodi S, Phillips A, Lundgren J, Logan R, Sharma S, Cole SR, et al. INSIGHT START Study Group and the HIV-CAUSAL Collaboration. Effect estimates in randomized trials and observational studies: comparing apples with apples. Am J Epidemiol 2019;188:1569–77.
[4] Laguna-Goya R, Utrero-Rico A, Talayero P, Lasa-Lazaro M, Ramirez-Fernandez A, Naranjo L, et al. IL-6-based mortality risk model for hospitalized patients with COVID-19. J Allergy Clin Immunol 2020;146:799–807.
[5] Kotch C, Barrett D, Teachey D. Tocilizumab for the treatment of chimeric antigen receptor T cell-induced cytokine release syndrome. Expert Rev Clin Immunol 2019;15:813–22.
[6] Xu Xiaoling, Han Mingfeng, Li Tiantian, Sun Wei, Wang Dongsheng, Fu Binqing, et al. Effective treatment of severe COVID-19 patients with tocilizumab. Proc Natl Acad Sci USA 2020;117:10970–5.
[7] Guaraldi G, Meschiari M, Cozzi-Lepri A. Tocilizumab in patients with severe COVID-19: a retrospective cohort study. Lancet Rheumatol 2020;2:e474–84.
[8] Gupta S, Wang W, Hayek SS, Chan L, Mathews KS, Melamed ML al. STOP-COVID Investigators. Association between early treatment with tocilizumab and mortality among critically ill patients with COVID-19. JAMA Intern Med 2021;181:41–51.
[9] Hermine O, Mariette X, Tharaux P-L, Resche-Rigon M, Porcher R, Ravaud P, et al. Effect of tocilizumab vs usual care in adults hospitalized with Covid-19 and moderate or severe pneumonia: a randomized clinical trial. JAMA Intern Med 2021;181:32–40.
[10] Salvarani C, Dolci G, Massari M, Merlo DF, Cavuto S, Savoldi L, et al., RCT-TCZ-COVID-19 Study Group. Effect of tocilizumab vs standard care on clinical worsening in patients hospitalized with COVID-19 pneumonia: a randomized clinical trial. JAMA Intern Med 2021;181:24–31.
[11] Veiga VC, Prats JAGG, Farias DLC, Rosa RG, Dourado LK, Zampieri FG, et al. Effect of tocilizumab on clinical outcomes at 15 days in patients with severe or critical coronavirus disease 2019: randomised controlled trial. BMJ 2021;372:n84.
[12] Gordon AC, Mouncey PR, Al-Beidh F, Rowan KM, Nichol AD, Arabi YM, et al., REMAP-CAP Investigators. Interleukin-6 receptor antagonists in critically ill patients with Covid-19. N Engl J Med 2021;384:1491–502.
[13] RECOVERY Collaborative Group. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet 2021;397:1637–45.
[14] Salama C, Han J, Yau L, Reiss WG, Kramer B, Neidhart JD, et al. Tocilizumab in patients hospitalized with Covid-19 pneumonia. N Engl J Med 2021;384:20–30.
[15] Rosas IO, Bräu N, Waters M, Go RC, Hunter BD, Bhagani S, et al. Tocilizumab in hospitalized patients with severe Covid-19 pneumonia. N Engl J Med 2021;384:1503–16.
[16] McCreary EK, Meyer NJ. Covid-19 controversies: the tocilizumab chapter. BMJ 2021;372:n244.
[17] Stone JH, Frigault MJ, Serling-Boyd NJ, Fernandes AD, Harvey L, Foulkes AS, et al., BACC Bay Tocilizumab Trial Investigators. Efficacy of tocilizumab in patients hospitalized with Covid-19. N Engl J Med 2020;383:2333–44.
[18] Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 2016;31:337–50.
[19] Mussini C, Falcone M, Nozza S, Sagnelli C, Parrella R, Meschiari M, et al., The Italian Society of Infectious and Tropical Diseases. Therapeutic strategies for severe COVID-19: a position paper from the Italian society of infectious and tropical diseases (SIMIT). Clin Microbiol Infect 2021;27:389–95.
[20] NICE covid-19 guidelines. https://www.nice.org.uk/advice/es33/chapter/Product-overview. [Accessed 17 March 2021].
[21] NIH COVID treatment guidelines. https://www.covid19treatmentguidelines.nih.gov/statement-on-tocilizumab/. [Accessed 17 March 2021].
[22] Published by IDSA on 4/11/ IDSA guidelines on the treatment and management of patients with COVID-19. 2020. Last updated, 4/14/2021.
[23] Dahabreh IJ, Hernán MA. Extending inferences from a randomized trial to a target population. Eur J Epidemiol 2019;34:719–22.
[24] Howick J, Glasziou P, Aronson JK. The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute? J R Soc Med 2009;10:186–94.
[25] Shepshelovich D, Yahav D, Ben Ami R, Goldvaser H, Tau N. Concordance between the results of randomized and non-randomized interventional clinical trials assessing the efficacy of drugs for COVID-19: a cross-sectional study. J Antimicrob Chemother 2021;76:2415–8.
[26] Gautret P, Lagier JC, Parola P, Hoang VT, Meddeb L, Mailhe M, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrob Agents 2020;56:105949.
[27] Arshad S, Kilgore P, Chaudhry ZS, Jacobsen G, Wang DD, Huitsing K, et al. Treatment with hydroxychloroquine, azithromycin, and combination in patients hospitalized with COVID-19. Int J Infect Dis 2020;97:396–403.
[28] Hill A, Garratt A, Levi J, Falconer J, Ellis L, McCann K, et al. Meta-analysis of randomized trials of ivermectin to treat SARS-CoV-2 infection. Open Forum Infect Dis 2021.
[29] Elgazzar A, Hany B, Youssef SA, Hafez M, Moussa H, et al. https://www.researchsquare.com/article/rs-100956/v1.
[30] Lawrence JM, Meyerowitz-Katz G, Heathers JAJ, Brown NJL, Sheldrick KA. The lesson of ivermectin: meta-analyses based on summary data alone are inherently unreliable. Nat Med 2021. https://doi.org/10.1038/s41591-021-01535-y. Sep 22.