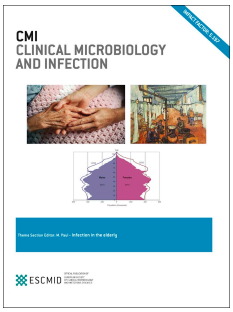
Journal Pre-proof



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

Alessandro Cozzi-Lepri, Colette Smith, Cristina Mussini

PII: S1198-743X(21)00620-0

DOI: https://doi.org/10.1016/j.cmi.2021.10.018

Reference: CMI 2734

To appear in: *Clinical Microbiology and Infection*

Received Date: 19 August 2021

Revised Date: 25 October 2021

Accepted Date: 25 October 2021

Please cite this article as: Cozzi-Lepri A, Smith C, Mussini C, Signals were broadly positive for months, but never definitive: the tocilizumab story, *Clinical Microbiology and Infection*, https://doi.org/10.1016/j.cmi.2021.10.018.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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

Journal Pre-proof

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

Alessandro Cozzi-Lepri[1], Colette Smith[1] and 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

Corresponding author:

Alessandro Cozzi-Lepri
Associate Professor in 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

Tel.: +44 (0)20801 68048 (Ext. 68048)
Email: a.cozzi-lepri@ucl.ac.uk

Word count 2,306

Journal Pre-proof

Abstract

Background
Most COVID-19 treatment guidelines currently recommend tocilizumab in combination with dexamethasone in critically ill patients who are exhibiting rapid respiratory decompensation.

Objectives
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 over July 2020-July 2021.

Content
Two large 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 small sample size randomized trials (RCTs) 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 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.

Abstract word count: 300

Journal Pre-proof

The clinical picture

The standard of care of patients with severe 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

Randomised clinical trials (RCTs) are considered the gold 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 organise 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 trust evidence from non-randomised 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, adequately powered, with 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 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, randomised studies are placed higher in the hierarchy of EBM because bias due to unmeasured confounding is minimised. Therefore, over the last two decades an increasing number of epidemiologists and statisticians have advocated for 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 randomised 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 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 tolerability of the drug in humans with rheumatoid arthritis and CAR-T led clinicians, initially in China followed by Europe and USA, to administer tocilizumab to treat 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 multicenter 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 the 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) was0·53 (95% CI: 0·31–0·89, p=0·016), thus showing high effectiveness of tocilizumab in reducing the risk of these events. The effect of tocilizumab was even greater in people with baseline PaO_2/FiO_2 value of less than 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 vs. 37% in the non-tocilizumab group with an aHR for death of0.71 (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 symptoms 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 hypoxemia and higher levels of inflammatory markers.

After August 2020, randomised 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 comprised mainly critically ill patients (12). Exact dosing of tocilizumab also varied by trial. In the RECOVERY, EMPACTA and REMAP-CAP (12-14) patients received 1 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 (10,15).

By the summer of 2020, because of the conflicting nature of the results of these 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 the need to re-synthesize all the evidence coming from interventional studies (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, albeit of a small magnitude, of tocilizumab 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 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, only included patients with mild disease severity with a median PaO2/FiO2 ratio of 265 mmHg (10). This is clearly a population with no ongoing cytokine storm for

whom tocilizumab should not be indicated. Secondly, 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 publically 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 patients (IP) data rather than an assemblage of summary statistics. Indeed, 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].

At this point, we should also note 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 replications 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 criticised to be affected by confounding and immortal-time bias [26,27]. Similarly, invermectim'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 put 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 randomised trials in isolation, and to ensure that meta-analyses consider differences in study designs and patient populations before pooling the results (23).

Nevertheless, there were a number of anomalies in this 'tocilizumab story' which seems 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 focussed 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, randomised trials and subsequent trigger of FDA EUA and inclusion in treatment guidelines. This includes 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 randomised and non-randomised 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 bias is more important than the actual study design (experimental vs. 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 under-estimated 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 D 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-randomised 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 guidelines 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.

Acknowledgments

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

Conflict of interest declaration

None of the authors declare conflicts of interest and no external funding was received for this work

Authors contributions

Cozzi-Lepri A: data analysis, data interpretation, writing and revising for intellectual content.
Smith C: writing and revising for intellectual content.
Mussini C: writing and revising for intellectual content.

References

1. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000 Sep;11(5):550-60.
2. Dahabreh IJ, Sheldrick RC, Paulus JK,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, 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;188:1569-1577.
4. Laguna-Goya R, Utrero-Rico A, Talayero P, et al. IL-6–based mortality risk model for hospitalized patients with COVID-19. Journal of Allergy and Clinical Immunology 2020;146:799–807.
5. Kotch C, Barrett D, Teachey D, et al. Tocilizumab for the treatment of chimeric antigen receptor T cell-induced cytokine release syndrome. Expert Rev Clin Immunol. 2019;15 (8):813-822.
6. Xiaoling Xu1, Mingfeng Han, Tiantian Li et al. Effective Treatment of Severe COVID-19 Patients with Tocilizumab. Proc Natl Acad Sci U S A 2020;117:10970-10975.
7. Guaraldi G, Meschiari M, Cozzi-Lepri A, Tocilizumab in patients with severe COVID-19: a retrospective cohort study. Lancet Rheumatol. 2020;2:e474-e484.
8. Gupta S, Wang W, Hayek SS, et 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, 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 2020;e206820.
10. Salvarani C, Dolci G, Massari M 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(1):24-31.
11. Veiga VC, Prats JAGG, Farias DLC et al. Coalition covid-19 Brazil VI Investigators. 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. REMAP-CAP Investigators, Gordon AC, Mouncey PR, Al-Beidh F. et al. Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19. N Engl J Med. 2021;384:1491-1502.
13. RECOVERY Collaborative Group. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet. 2021;397(10285):1637-1645.

14. Salama C, Han J, Yau L, 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. et al. Tocilizumab in Hospitalized Patients with Severe Covid-19 Pneumonia. N Engl J Med. 2021;384:1503-1516.
16. McCreary EK, Meyer NJ. Covid-19 controversies: the tocilizumab chapter. BMJ. 2021;372:n244.
17. BACC Bay Tocilizumab Trial Investigators, Stone JH, Frigault MJ, Serling-Boyd NJ et al. Efficacy of Tocilizumab in Patients Hospitalized with Covid-19. N Engl J Med. 2020;383:2333-2344.
18. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. 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, et al for 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: S1198-743X(20)30770-9.
20. NICE Covid-19 Guidelines accessed on 17 March **2021** at https://www.nice.org.uk/advice/es33/chapter/Product-overview
21. NIH COVID treatment guidelines accessed on 17 March 2021 at https://www.covid19treatmentguidelines.nih.gov/statement-on-tocilizumab/
22. IDSA Guidelines on the Treatment and Management of Patients with COVID-19 Published by IDSA on 4/11/2020. Last updated, 4/14/2021
23. Dahabreh IJ, Hernán MA. Extending inferences from a randomized trial to a target population. Eur J Epidemiol. 2019;34:719-722.
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 Aug 12;76(9):2415-2418
26. Gautret P, Lagier JC, Parola P, Hoang VT, Meddeb L, Mailhe M, Doudier B, Courjon J, Giordanengo V, Vieira VE, Tissot Dupont H, Honoré S, Colson P, Chabrière E, La Scola B,

Rolain JM, Brouqui P, Raoult D. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrob Agents. 2020 Jul;56(1):105949.

27. Arshad S, Kilgore P, Chaudhry ZS, et al. Treatment with hydroxychloroquine, azithromycin, and combination in patients hospitalized with COVID-19. *Int J Infect Dis*. 2020;97:396-403.
28. Andrew Hill, Anna Garratt, Jacob Levi, Jonathan Falconer, Leah Ellis, Kaitlyn McCann, Victoria Pilkington, Ambar Qavi, Junzheng Wang, Hannah Wentzel, Meta-analysis of randomized trials of ivermectin to treat SARS-CoV-2 infection, *Open Forum Infectious Diseases*, 2021
29. Elgazzar 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 Sep 220.

Journal Pre-proof