

ORIGINAL RESEARCH

COVID-19-induced hyperinflammation, immunosuppression, recovery and survival: how causal inference may help draw robust conclusions

Robert B M Landewé ^{1,2}, Sofia Ramiro,^{3,4} Rémy L M Mostard⁵

To cite: Landewé RBM, Ramiro S, Mostard RLM. COVID-19-induced hyperinflammation, immunosuppression, recovery and survival: how causal inference may help draw robust conclusions. *RMD Open* 2021;**7**:e001638. doi:10.1136/rmdopen-2021-001638

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/rmdopen-2021-001638>).

Received 19 February 2021
Revised 18 March 2021
Accepted 22 March 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Amsterdam Rheumatology Center, AMC, Amsterdam, The Netherlands

²Rheumatology, Zuyderland MC, Heerlen, The Netherlands

³Rheumatology, Leiden University Medical Center, Leiden, The Netherlands

⁴Rheumatology, Zuyderland Medical Center, Heerlen, The Netherlands

⁵Department of Pulmonology, Zuyderland Medical Centre Heerlen, Heerlen, The Netherlands

Correspondence to

Dr Robert B M Landewé;
landewe@rlandewe.nl

ABSTRACT

Background The CHIC study (COVID-19 High-intensity Immunosuppression in Cytokine storm syndrome) is a quasi-experimental treatment study exploring immunosuppressive treatment versus supportive treatment only in patients with COVID-19 with life-threatening hyperinflammation. Causal inference provides a means of investigating causality in non-randomised experiments. Here we report 14-day improvement as well as 30-day and 90-day mortality.

Patients and methods The first 86 patients (period 1) received optimal supportive care only; the second 86 patients (period 2) received methylprednisolone and (if necessary) tocilizumab, in addition to optimal supportive care. The main outcomes were 14-day clinical improvement and 30-day and 90-day survival. An 80% decline in C reactive protein (CRP) was recorded on or before day 13 (CRP >100 mg/L was an inclusion criterion). Non-linear mediation analysis was performed to decompose CRP-mediated effects of immunosuppression (defined as natural indirect effects) and non-CRP-mediated effects attributable to natural prognostic differences between periods (defined as natural direct effects).

Results The natural direct (non-CRP-mediated) effects for period 2 versus period 1 showed an OR of 1.38 (38% better) for 14-day improvement and an OR of 1.16 (16% better) for 30-day and 90-day survival. The natural indirect (CRP-mediated) effects for period 2 showed an OR of 2.27 (127% better) for 14-day improvement, an OR of 1.60 (60% better) for 30-day survival and an OR of 1.49 (49% better) for 90-day survival. The number needed to treat was 5 for 14-day improvement, 9 for survival on day 30, and 10 for survival on day 90.

Conclusion Causal inference with non-linear mediation analysis further substantiates the claim that a brief but intensive treatment with immunosuppressants in patients with COVID-19 and systemic hyperinflammation adds to rapid recovery and saves lives. Causal inference is an alternative to conventional trial analysis, when randomised controlled trials are considered unethical, unfeasible or impracticable.

Key messages

What is already known about this subject?

- We have previously shown in a quasi-experimental treatment study with non-equivalent treatment arms (the CHIC study, COVID-19 High-intensity Immunosuppression in Cytokine storm syndrome) that a brief course of high-dose methylprednisolone plus tocilizumab if necessary may accelerate recovery and save lives of patients with COVID-19-associated cytokine storm syndrome.
- The CHIC study was a non-randomised study, which leaves room for potential selection biases explaining treatment differences and precludes causal conclusions.

What does this study add?

- Causal inference with non-linear mediation analysis allows under certain circumstances a causal interpretation of the results of quasi-randomised experiments with a control group.
- We showed that the beneficial effects of 5–7 days of intense immunosuppressive therapy in patients with COVID-19-associated CSS on 14-day improvement and 90-day survival are mediated via an immediate and profound C reactive protein response.

How might this impact on clinical practice?

- Results of quasi-experimental studies conducted under difficult circumstances or time pressure, or when randomised controlled trials are considered unfeasible or unethical, may gain credibility by application of the principles of causal inference.

INTRODUCTION

Patients with COVID-19 pneumonia may develop a syndrome characterised by severe systemic hyperinflammation, sometimes called cytokine storm syndrome (CSS). Patients with COVID-19-associated CSS are at an increased risk of severe pulmonary thrombosis, respiratory insufficiency and death.¹

Recently, we have shown in a quasi-experimental study² with matched historical control patients (the so-called non-equivalent group design with matched controls, NEGD) that a strategy of immunosuppression with high-dose methylprednisolone (MP) plus the interleukin 6 receptor blocker tocilizumab (TCZ) for those who did not immediately respond may accelerate recovery and reduce mortality in patients with predefined signs of CSS.³ All study patients fulfilled the criteria for hyperinflammation, but the patients of the historical control group were admitted between 7 March and 31 March (period 1) and the patients of the immunosuppressive treatment group starting from 1 April (period 2). None of the patients in period 1 received (components of) the immunosuppressive treatment strategy, while all of the eligible patients in period 2 were treated according to this strategy. This strict assurance of period specificity excluded bias by indication. Further, conventional confounding analysis revealed that the treatment contrast between the intervention group and the historical control group remained fully intact after adjustment for baseline differences between both groups.³

It was argued by ourselves³ and by others⁴ that this quasi-experimental design did not preclude the scenario that the prognosis for patients in period 1 was worse than for patients in period 2. This is a valid argument since hospitals were overwhelmed by critically ill patients in the first most hectic weeks of the pandemic and optimal supportive care for COVID-19 had to be fully developed. Only a randomised controlled trial (RCT), it was argued, would have provided unambiguous proof. However, in the design phase of the study, we felt ethical and practical arguments precluded the conduct of a randomised clinical experiment with an untreated control arm in our setting.

In the absence of RCTs, *causal inference* has been proposed for studies with NEGD to further elaborate potentially causal relationships.⁵ Here we have applied the principles of causal inference to further substantiate the effects of immunosuppressive therapy in COVID-19-induced hyperinflammation. We have used non-linear mediation analysis, a form of causal inference, to decompose the effects due to differences in prognosis between different time periods and the effects due to the treatment itself, by introducing a mediator variable *CRP response*. The hypothesis, based on prior knowledge, was that part of the observed beneficial effect of immunosuppressive therapy during period 2 was mediated by active suppression of inflammation and another part by a better prognosis in period 2 than in period 1. The analysis described here is an example of how, in the absence of a randomised controlled experiment, the credibility of quasi-experimental studies can be increased by the principles of causal reasoning.

METHODS

The experimental study

Patients from the CHIC study (COVID-19 High-intensity Immunosuppression in Cytokine storm syndrome), described elsewhere in more detail,³ were analysed. The CHIC study was a quasi-experimental treatment study that used matched historical control patients. In brief, patients had to have COVID-19 pneumonia, complicated by signs of hyperinflammation, as defined by rapid respiratory deterioration on or during admission, plus fulfilment of at least two out of three biomarker criteria (C reactive protein (CRP) >100 mg/L, serum ferritin >900 µg/L, D-dimer >1.500 µg/L). Treatment was strictly time-separated. Patients assigned to the historical control group had to be admitted in period 1 and were selected retrospectively using the same above-mentioned (biomarker) criteria. Patients eligible for the intervention had to be admitted in period 2 and received treatment with MP 250 mg intravenously on day 1, followed by MP 80 mg intravenously for at least 4 consecutive days. Single-dose TCZ, 8 mg/kg body weight, maximum of 800 mg, was administered intravenously if the clinical situation did not improve, or instead worsened between day 2 and day 5.

In the original CHIC study report, three primary outcome variables were defined: discharge from the hospital or at least two stages of improvement according to the WHO improvement scale designed for the purpose of influenza pneumonia; hospital mortality; and the need to start invasive ventilation.³ Outcomes were analysed on the basis of time-to-event. In comparison with patients in period 1 (on standard care), patients in period 2 (on treatment) had a 79% higher likelihood of two-stage WHO improvement, 65% less hospital mortality and 71% less need for mechanical ventilation.

For this causal inference analysis, three outcomes were analysed: WHO improvement on day 14 of CSS, survival on day 30 and survival on day 90. We refrained from analysing mechanical ventilation as an outcome since mechanical ventilation was considered another intermediate outcome variable between WHO improvement (in hospital) and survival on day 30 (partially in hospital) and day 90 (outside the hospital).

Causal inference

Causal inference, or causal reasoning, is a process of drawing conclusions based on the conditions of the occurrence of a certain effect.^{6,7} Causal inference shares the elements of a Bayesian analysis⁸ and incorporates prior knowledge and information about the pathophysiology of a condition, as well as knowledge about the mechanism of action that is presumably at the basis of a certain effect.⁶ According to Pearl and MacKenzie⁶, causal inference presumes an association between exposure (X) and outcome (Y) (the question of correlation, forming the first layer of causal hierarchy), insight into how an intervention on X (reflected by the operator $do(X)$) changes Y (the intervention question, forming the second layer of

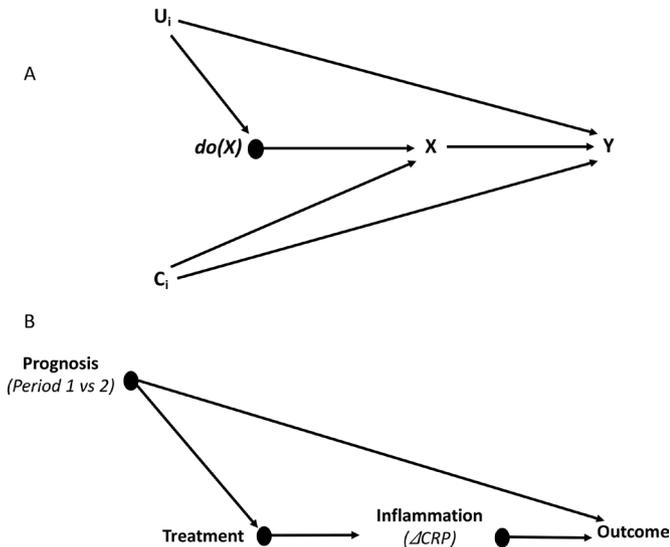


Figure 1 Structural causal model of how an intervention ($do(X)=1$) in patients fulfilling certain selection criteria (X) may have impact on a certain outcome (Y) in the context of a non-equivalent group design (A). (B) The situation worked out for the CHIC study in particular, where *treatment* refers to immunosuppressive treatment for patients with COVID-19 who are in a state of life-threatening hyperinflammation (X), *outcome* to WHO improvement or survival (30 days and 90 days, respectively) (Y), and Δ CRP to a mediator 80% CRP response (M). CHIC, COVID-19 High-intensity Immunosuppression in Cytokine storm syndrome; CRP, C reactive protein.

causal hierarchy), and a theory about—or insight into—how a different intervention on X than was really applied would have modified Y (the counterfactual question, forming the third layer of causal hierarchy).⁶ In an RCT, the counterfactual question is addressed by interpreting the effect in the control group (sometimes placebo). In studies with NEGD, other sources for counterfactual information should be sought and justified. In the CHIC study, counterfactual information was obtained using the untreated patients in period 1.

The structural causal model

The model of interest in this study is the relation between the COVID-19-induced state of hyperinflammation (X) on the one hand, and worsening of disease followed by respiratory insufficiency and death (outcome Y) on the other hand. According to Pearl and MacKenzie⁶, causality assumes that an intervention on X ($do(X)$) leads to a change in Y . Intensive immunosuppressive treatment with high doses of MP (or TCZ) administered intravenously is widely known to suppress inflammation in systemic rheumatological diseases, such as rheumatoid arthritis, systemic lupus erythematosus and vasculitis, and to improve their outcomes.^{9,10} The most widely available and sensitive biomarker to follow the state of systemic inflammation (X) is the patient's serum CRP level. CRP can respond quickly (within days) to high-dose immunosuppressive treatment. A true causal relationship therefore not only requires an association between the

treatment and the CRP response, but also between the CRP response and the outcome (Y). In addition, causality requires *irreversibility*, implying here that drug-induced suppression of inflammation (a positive CRP response) precedes a change in the clinical situation and not vice versa (namely that the clinical situation of the patient determines (precedes) the choice of treatment (which would cause bias by indication)).

The hypothesised structural causal graph (or directed acyclic graph) is visualised in figure 1A.

The association between X and Y (figure 1A) can be influenced by a set of variables (C_i) that can impact both X and Y and may confound the relationship of X on Y . Some of these confounders have been measured and adjusted for; others have not and can be responsible for residual confounding. One consequence of applying the causal inference theory is that this *back-door* via C_i is closed by conditioning on $do(X)$ by a phenomenon called *d-separation*¹¹: arrows descending from $do(X)$ and C_i collide in X (C_i and $do(X)$ are the *colliders*), and the effects of C_i for the interpretation of the effects of X on Y can be ignored. Another (imaginary) set of variables (U_i) can be formulated that may have an impact on both $do(X)$ and Y , and for which *d-separation* is irrelevant because there are no common causes. Examples of such disturbing variables, often found in studies with NEGD, are variables reflecting the severity of the disease, which physicians take into consideration when deciding about the start, change or intensity of a treatment. Such variables may cause confounding by indication. In the CHIC study (figure 1B), classic confounding by indication can be ruled out since treatment assignment was strictly time-separated (treatment=0 for period 1 and treatment=1 for period 2), but prognostic differences between period 1 and period 2 (a period effect) could still exist, and confounding of this type may contribute to explaining the observed treatment effects. In order to decompose the specific effect of treatment and the non-specific effects of period (1 and 2) (referred to by Pearl¹² as *front-door adjustment*), a mediator variable (Δ CRP) is introduced that best reflects the pathophysiological mechanism by which immunosuppressive treatment is supposed to reduce hyperinflammation (X) and influence the outcome (Y). Non-linear mediation analysis (according to Pearl¹² and VanderWeele¹³) is used to estimate what part of the total observed effect is attributable to period (natural direct effect, NDE) and what part is attributable to treatment-induced suppression of inflammation, mediated via Δ CRP (natural indirect effect, NIE) (see online supplemental text S1 and S2 for a more detailed discussion).

The mediator variable

It was hypothesised that active treatment ($do(X)=1$) in patients with hyperinflammation (X) will lead to a measurable CRP response (M), which will in turn lead to a change in the outcome (Y). The discriminating capacity of M in predicting Y was tested by conventional receiver

operating characteristic analysis. The optimal CRP response, measured up to 13 days after inclusion, based on the best balance between sensitivity and specificity for discriminating between clinical improvement (assessed after 14 days) or not, or survival (after 30 days) or not, appeared to be 80% decrease from baseline (data not shown).

Definition of natural indirect and direct effects

NIE is defined here as the effect that is mediated via *M* (the CRP response). NDE is defined here as all other effects (positive or negative) that are *not* mediated via *M*. Such effects may include adverse effects of the treatment (eg, patients on immunosuppressive treatment may get complications such as bacterial infections) as well as unmeasured prognostic differences between patients in period 1 and patients in period 2, either existing at baseline or occurring during follow-up (unmeasured confounding) and not related to the mediator.

Modelling of the impact of treatment

After decomposition of the NDE and NIE, the ORs for the NIE (OR_{NIE} in table 2) were assumed to reflect the unbiased effects of immunosuppressive treatment, and the ORs for the NDE (OR_{NDE} in table 2) were assumed to reflect the unbiased effect of changed (improved) prognosis in period 2 versus period 1. Only OR_{NIE} was used to calculate the post-treatment likelihood for all three outcomes, as a function of their pretreatment likelihood (probability plots), using Bayes' rule.¹⁴ Bayes' rule implies post-treatment likelihood (expressed as odds) is proportional to the product of OR for the treatment effect and pretreatment likelihood (expressed as odds):

$$\text{post-treatment odds} = OR_{NIE} \times \text{pretreatment odds}$$

Observed pretreatment odds were assumed to be the odds on the outcomes (*Y*) actually observed in period 1 (the control period). The number needed to treat (NNT) with intensive immunosuppressive treatment in order to find one additional patient with a favourable outcome (*Y*) was calculated as $NNT = 1 / \text{absolute risk reduction (ARR)}$. ARR was defined as the likelihood of *Y* in period 2 minus the likelihood of *Y* in period 1.

RESULTS

Due to missing CRP values at follow-up, a proper CRP response could not be calculated in 6 of the 86 patients from period 1. Data on a CRP response were available for all 86 patients from period 2. In total 166 patients were included in this analysis. The mean (SD) baseline CRP level was 167 (98) mg/L for patients in period 1 and 159 (73) mg/L for patients in period 2. The mean (SD) CRP response was 20% (51%) in period 1 and 85% (83%) in period 2. Table 1 proves that the likelihood of WHO clinical improvement on day 14 and survival on days 30 and 90 was strongly associated with the occurrence of an early CRP response (last column); those with a CRP response had better outcomes than those without. This effect was found both in period 1 (the period without treatment) and in period 2 (with treatment) (see columns 3 and 4). However, the likelihood of a positive CRP response was far higher for patients in period 2 (who had received treatment) than for patients in period 1 (who had not received treatment). Still, in period 1, a few positive CRP responses (8 of 80 had it) were documented, while not all patients in period 2 had a positive CRP response (21 of 86 did not have it).

Table 1 Likelihood of clinical improvement (day 14) and survival (days 30 and 90) in both periods, stratified by CRP response* (up to day 13)

		Likelihood (P)		
		Period 1, control patients	Period 2, patients on active treatment	Periods 1 and 2, all patients
Improvement at 14 days	$\emptyset CRP_{no}$	0.24 (17/72)†	0.33 (7/21)	0.26 (24/93)
	$\emptyset CRP_{yes}$	0.88 (7/8)	0.75 (49/65)	0.77 (56/73)
	All	0.30 (24/80)	0.65 (56/86)	0.48 (80/166)
Survival at 30 days	$\emptyset CRP_{no}$	0.49 (35/72)	0.57 (12/21)	0.51 (47/93)
	$\emptyset CRP_{yes}$	1.0 (8/8)	0.91 (59/65)	0.92 (67/73)
	All	0.54 (43/80)	0.83 (71/86)	0.69 (114/166)
Survival at 90 days	$\emptyset CRP_{no}$	0.49 (35/72)	0.57 (12/21)	0.51 (47/93)
	$\emptyset CRP_{yes}$	1.0 (8/8)	0.85 (55/65)	0.86 (63/73)
	All	0.54 (43/80)	0.78 (67/86)	0.66 (110/166)

Figures in brackets reflect the number of patients (n/N).

*The $\emptyset CRP$ threshold for a 'good CRP response' was defined as 80% decrease from baseline based on receiver operator characteristic analysis.

†0 ≤ likelihood (P) ≤ 1.

CRP, C reactive protein.

Table 2 Decomposition of effects: OR for the natural direct effect (NDE), the natural indirect effect (NIE) and the total causal effect (TCE)

	Between-group effect <i>not</i> mediated via CRP	Between-group effect mediated via CRP	Total effect
	OR_{NDE}	OR_{NIE}	OR_{TCE}
Improvement at 14 days	1.38*	2.27	3.14
Survival at 30 days	1.16	1.60	1.87
Survival at 90 days	1.16	1.49	1.73

Two types of contrasts can be distinguished: contrasts *not* mediated by CRP response and contrasts mediated by CRP response. Contrasts *not* mediated by CRP response refer to ‘natural’ prognostic differences between period 1 and period 2 and may include differences existing already at baseline (measured and unmeasured) or differences occurring during follow-up (eg, differences in ventilatory support, or differences in antithrombotic treatment or unwarranted effects due to bacterial superinfection). Contrasts mediated by CRP response include differences specifically due to immunosuppressive therapy under the assumption that these cause suppression of inflammation (as assessed by an 80% CRP response).

*ORs refer to the contrast in outcome between period 2 and period 1. OR <1 indicates less effect on the outcome in period 2 than in period 1; OR >1 indicates more effect in period 2 than in period 1.
CRP, C reactive protein.

When looking in the strata of ‘no CRP response’ (the lines with $\emptyset CRP_{no}$), there was also some effect of period 2 versus period 1, independent of CRP response, but this contrast was smaller than the contrast between positive CRP response and no positive CRP response for all three outcomes. So far, it looks as if immunosuppressive treatment had an effect on outcome via CRP response (an indirect effect), but that there was also a period effect (a direct effect).

Direct and indirect effects were further decomposed and quantified using non-linear mediation analysis (table 2). For all three outcomes, direct (CRP response-independent) period effects could be demonstrated. The likelihood of non-treatment-related WHO clinical improvement on day 14 was 38% higher (OR=1.38) for patients in period 2 than in period 1, and survival was 16% higher (OR=1.16) (both on day 30 and day 90). This means that, irrespective of immunosuppressive treatment, the overall prognosis of period 2 patients for improvement and survival was better than of period 1 patients.

For all three outcomes, substantial indirect effects could also be demonstrated. The likelihood of CRP response-mediated clinical improvement on day 14 was 127% (OR=2.27) higher for patients in period 2 than in period 1, 30-day survival was 60% (OR=1.60) higher and 90-day survival was 49% (OR=1.49) higher.

Figure 2 extrapolates the impact of the treatment effects (the decomposed NIE) to different virtual pretreatment levels. Note that the figure reflects modelled data based on measured effects in the CHIC study and not real data. The treatment effects visualised in this graph can be considered as unbiased. All curves lying above the diagonal represent beneficial treatment effects. The more the curve deviates from the diagonal the more impressive the treatment effect is. More treatment effects are seen in 14-day clinical improvement than in 30-day and 90-day survival. The symbols representing the actually measured data in the CHIC study are plotted in the graph. The

NNTs calculated on the basis of these really observed data are 5 for 14-day clinical improvement, 9 for 30-day survival, and 10 for 90-day survival.

DISCUSSION

The data indicate that, in comparison with period 1, the immunosuppressive therapy administered to patients in period 2 has increased the 14-day clinical improvement rate over and above the level that could be expected based on the prognostic advantages that patients in period 2 (apparently) had. Similar effects were found for (30-day and 90-day) survival. It was possible to decompose the period effect (direct) from the CRP-mediated effect (indirect) by using mediation analysis. The period effect includes non-CRP-related improvements in management over time, such as better ventilatory support techniques,¹⁵ more focus on anticoagulation,¹⁶ and putatively a better natural prognosis. While it cannot be entirely excluded that these improvements have also had some impact on CRP response, it is hard to believe that they could reduce baseline CRP levels by more than 80% within 13 days. The estimate of the immunosuppressive treatment effect is nevertheless conservative. It is possible that part of the specific effect by immunosuppressive treatment was not captured by us as indirect effect. As said, the threshold of 80% improvement in CRP before a CRP response was counted as present was quite high, and it cannot be excluded that some patients with a beneficial outcome due to immunosuppressive treatment had in reality a CRP response slightly lower than 80%. If true, this would mean that the real contribution of immunosuppressive therapy is even higher than estimated by us.

The difference in prognosis between period 1 and period 2 (the period effect), here reflected by an $OR_{NDE} > 1$ in the decomposition analysis, was obvious. The primary (conventional) analysis of the CHIC study already showed that the existing (measured) differences at baseline could not be held responsible for the observed treatment

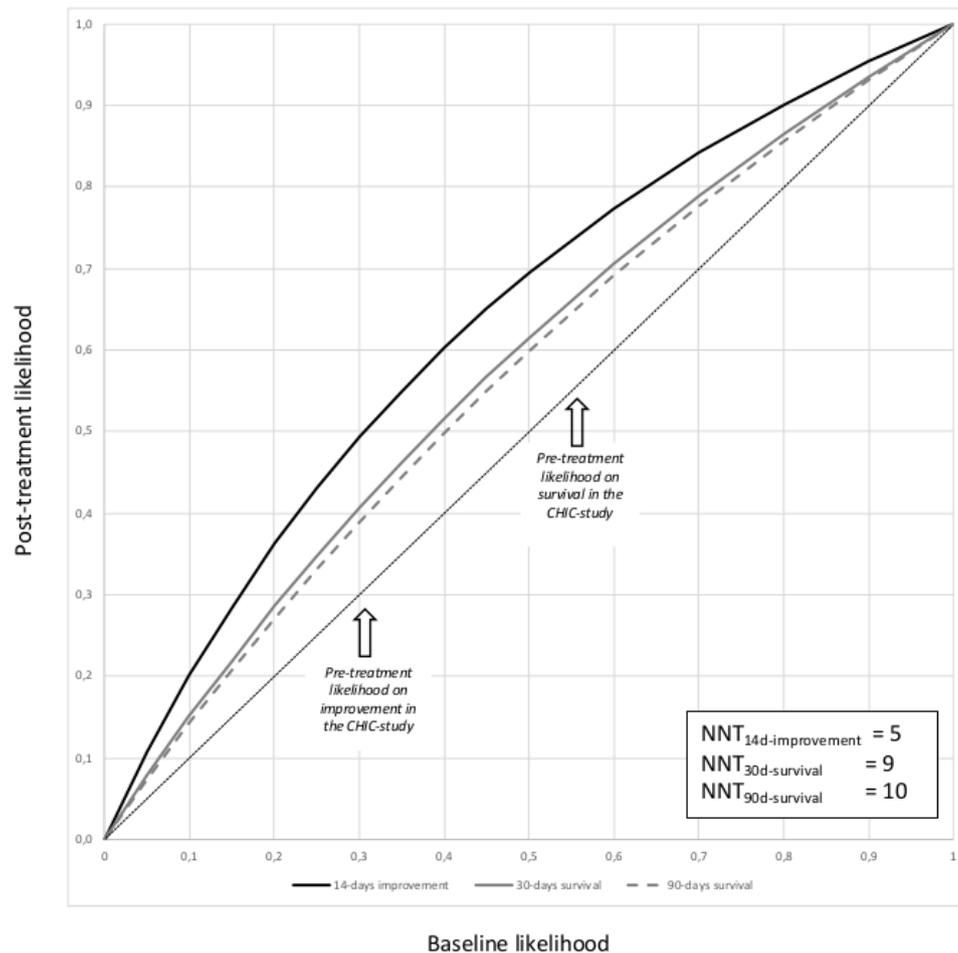


Figure 2 Probability curves for the post-treatment likelihood of clinical improvement and survival, plotted against different (virtual) pretreatment likelihoods. Post-treatment likelihood is the product of prior likelihood (expressed as *odds*) and OR for the natural indirect effects (OR_{NIE}), as determined in table 2. The diagonal reflects the situation of *zero treatment effect*. Curves above the diagonal reflect a beneficial effect on outcome (Y) by treatment ($do(X)=1$). Pretreatment likelihood for 14-day clinical improvement, 30-day survival and 90-day survival, as observed in period 1 of the CHIC study, is marked with arrows in the figure. The number needed to treat (NNT) was calculated for these rates actually observed in the CHIC study by taking the reciprocal of the difference between pretreatment and post-treatment likelihood ($NNT=1/\text{absolute risk reduction}$). CHIC, COVID-19 High-intensity Immunosuppression in Cytokine storm syndrome.

effects.³ Still, as brought up by others, differences in unmeasured variables can be responsible for differences in prognosis in a study with NEGD. Not only patients' baseline prognosis may have differed, prognostic differences could also have occurred during follow-up through non-inflammation-mediated mechanisms (eg, differences in mechanical ventilation or anticoagulation therapy, or adverse effects due to immunosuppressive therapy).

At the roots of this causal analysis was the critique that the real efficacy of immunosuppressive therapy could not be estimated properly because of these prognostic differences (see figure 1: the proposed structural causal model). Formally, the CHIC study lacks the internal validity of an RCT. That there were prognostically relevant differences between period 1 patients and period 2 patients has been proven by this analysis, but the mediation analysis also has made clear that these prognostic benefits only account for part of the observed differences between period 1 and period 2. In fact, the positive

effects of immunosuppressive therapy outweigh the positive effects of a better prognosis over time.

In the primary analysis, it has already been stipulated, using conventional statistical analysis, that the observed treatment contrast was robust to adjustment for measured (known) confounders. The structural causal model further adds credibility to this observation. Formally, by changing the state of hyperinflammation (X) through administering immunosuppressants ($do(X)$), by ruling out confounding of the relationship of X on Y by *d-separation*, and by proving that changing (X) has a positive impact on the outcome (Y) mediated by CRP (M), the relation between the 'parent' (X) and its 'descendant' (Y) proves to be causal.⁶ This means immunosuppressive therapy can be considered an effective means of improving outcome via reducing inflammation, as measured by CRP response. The requirement of *time precedence*, relevant for causation, is fulfilled since a CRP response up to day 13 follows the intervention on day 0

and precedes the clinical outcomes on days 14, 30 and 90, respectively. The formal requirement of *ignorability*, meaning that treatment is independent of the outcome, assured in an RCT by random treatment assignment, is met here by the strict time separation of period 1 and period 2 and period exclusivity for control (period 1) and immunosuppressive (period 2) treatment. We had already stipulated in the primary publication that bias by indication could be ruled out as an explanation for observed treatment differences.³

The potential implications for the treatment of patients with COVID-19 are noteworthy. Figure 2 shows the potential impact of immunosuppressive treatment on top of optimal supportive care in appropriately selected patients, namely those with hyperinflammation or CSS. Note that these effects should be considered free of bias since we separated them from the period effects. At first glance the effects visualised in figure 2 do not seem to be very impressive, but the intervention is only brief and temporary (5–10 days), applied during the acute phase of COVID-19, while its benefit in terms of improvement and better survival extends until at least 90 days. The brief intervention is assumed to help patients recovering from a temporary but dangerous period of hyperinflammation, invoked by the viral infection followed by dysregulation of the immune system.¹ We used the CHIC study data to estimate NNTs, which were understandably better for clinical improvement than for survival. These estimated NNTs compare very well with those that can be deduced from the RECOVERY trial, which has compared the effects of the glucocorticoid dexamethasone on top of supportive care only versus supportive care alone.¹⁷ The beneficial effect of dexamethasone on 30-day mortality was rather small in the entire study population (we calculated an NNT of 36). Looking at the patients who needed mechanical ventilation, more likely those with hyperinflammation, the NNT in the RECOVERY trial was 8–9, very close to what we estimated for 30-day and 90-day survival in the CHIC study. Unfortunately, the RECOVERY investigators did not look at the subgroup of patients with increased CRP at baseline. Taking both studies together, though, NNTs of approximately 10 for mortality/survival are very reasonable and broadly considered clinically relevant. It means that 10 patients with COVID-19-induced hyperinflammation should be treated in order to save one additional life.

The structural causal model we have proposed here applies to patients with COVID-19 who fulfil our criteria for hyperinflammation (generalisability). This means *not* all patients with COVID-19 admitted to the hospital should be treated with immunosuppressive drugs, even though independent RCTs have now claimed benefits for both components of our strategy, when administered separately.^{17 18} The RECOVERY investigators also concluded that dexamethasone did *not* provide benefits when administered to patients who did *not* need oxygen to be supplied (the less severe patients)¹⁷ and TCZ

seemed to be less effective in trials with patients *without* signs of hyperinflammation.^{19 20}

The structural causal model as worked out here predicts the beneficial effects of immunosuppressive drugs are non-specific and depend on the rapid suppression of hyperinflammation (measured by a rapid and profound CRP response). It does not suggest or support a compound-specific treatment effect. Further, the structural causal model implies that immunosuppressive therapy only improves the outcome on top of optimal supportive care (ventilation, anticoagulation and so on), which should not be omitted.

Acceptability of clinical research, measured as clinical implementation, depends on a combination of several factors, including scientific rigour, comprehension and face validity. It is obvious that the results of RCTs are more credible than those of quasi-experimental studies, which is why the clinical community as well as guideline committees always *wait for the trials*. While the logic of causal inference is not fundamentally different from that underpinning RCTs, causal inference will be mistrusted, maybe since clinicians consider it *less statistical* (this article does not contain one p value) and medical journals will basically be reluctant to accept. One should for, instance, accept that a conclusion of causality depends on the acceptability of using *counterfactual* information (the third layer of causation). For the CHIC study, this means accepting the estimation of how patients in period 1 would have responded on immunosuppressive therapy, based on a biologically plausible mechanism (suppression of inflammation). In reality, however, no one in period 1 has ever received immunosuppressive therapy. Clinicians and guideline committees, used to interpret RCT results that provide a global estimation of counterfactuals on a silver platter (the control arm), will find this difficult to accept, but counterfactuals have become common in many fields of science and facets of society, such as artificial intelligence and machine learning, social sciences and econometrics.^{21–23} Societies cannot even properly function anymore without accepting counterfactual information and principles of causal inference.

We still think conventional RCTs should be done whenever possible to resolve the most burning clinical questions and especially in situations of maximum uncertainty (equipoise). However, clinical scenarios may exist in which RCTs are unrealistic, unethical or impractical. In such circumstances, causal inference provides a basis for accepting the added value of certain treatments that have not passed the filter of RCTs yet, or will never do so.

Contributors All authors are responsible for the intellectual content of the manuscript and designed it together. RL wrote the manuscript, SR did the analyses, and all authors approved the final version of this manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The design and performance of the CHIC study were approved by the ethics committee of the Zuyderland Medical Center, Heerlen/Sittard.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The database contains deidentified (anonymous) participant data. Requests for admission can be directed to landewe@rlandewe.nl.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Robert B M Landewé <http://orcid.org/0000-0002-0577-6620>

REFERENCES

- Fajgenbaum DC, June CH. Cytokine storm. *N Engl J Med* 2020;383:2255–73.
- Bärnighausen T, Röttinger J-A, Rockers P, et al. Quasi-experimental study designs series-paper 1: introduction: two historical lineages. *J Clin Epidemiol* 2017;89:4–11.
- Ramiro S, Mostard RLM, Magro-Checa C, et al. Historically controlled comparison of glucocorticoids with or without tocilizumab versus supportive care only in patients with COVID-19-associated cytokine storm syndrome: results of the chiC study. *Ann Rheum Dis* 2020;79:1143–51.
- Winthrop KL, Mariette X. To immunosuppress: whom, when and how? that is the question with COVID-19. *Ann Rheum Dis* 2020;79:1129–31.
- Pearl J. An introduction to causal inference. *Int J Biostat* 2010;6:Article 7.
- Pearl J, MacKenzie D. *The book of why: the new science of cause and effect*. New York, USA, 2018.
- Rubin D. *Matched Sampling for Causal Effects*. Cambridge: Cambridge University Press, 2006.
- Ashby D. Bayesian statistics in medicine: a 25 year review. *Stat Med* 2006;25:3589–631.
- Aletaha D, Smolen JS. Diagnosis and management of rheumatoid arthritis: a review. *JAMA* 2018;320:1360–72.
- Dörner T, Furie R. Novel paradigms in systemic lupus erythematosus. *Lancet* 2019;393:2344–58.
- Tritchler D. Reasoning about data with directed graphs. *Stat Med* 1999;18:2067–76.
- Pearl J. Interpretation and identification of causal mediation. *Psychol Methods* 2014;19:459–81.
- VanderWeele TJ. A unification of mediation and interaction: a 4-way decomposition. *Epidemiology* 2014;25:749–61.
- Westbury CF. Bayes' rule for clinicians: an introduction. *Front Psychol* 2010;1:192.
- Raof S, Nava S, Carpati C, et al. High-Flow, noninvasive ventilation and awake (Nonintubation) Prone in patients with coronavirus disease 2019 with respiratory failure. *Chest* 2020;158:1992–2002.
- Barnes GD, Burnett A, Allen A, et al. Thromboembolism and anticoagulant therapy during the COVID-19 pandemic: interim clinical guidance from the anticoagulation forum. *J Thromb Thrombolysis* 2020;50:72–81.
- Lim WSet al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. *N Engl J Med* 2020.
- Salama C, Han J, Yau L, et al. Tocilizumab in patients hospitalized with Covid-19 pneumonia. *N Engl J Med* 2021;384:20–30.
- Stone JH, Frigault MJ, Serling-Boyd NJ, et al. Efficacy of tocilizumab in patients hospitalized with Covid-19. *N Engl J Med* 2020;383:2333–44.
- Remap-Cap-Investigators GAC, Mouncey PR, et al. Interleukin-6 receptor antagonists in critically ill patients with Covid-19. *N Engl J Med* 2021.
- Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun* 2020;11:3923.
- Kleinberg J, Lakkaraju H, Leskovec J, et al. Human decisions and machine predictions. *Q J Econ* 2018;133:237–93.
- Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods* 2013;18:137–50.