Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

RMD Open

Online Supplementary Text S1

Mediation vs. interaction according to *VanderWeele*[1]

VanderWeele (2014) handles non-linear mediation and interaction by decomposing the total effect of an exposure on an outcome into four separable components:

Let A be the exposure (a=1, a=0), M the mediator (m=1, m=0), Y the outcome (y=1, y=0) and *p* the probability on the outcome (Y). Let p_{am} be the likelihood of Y, given A=a and M=m (conditional probability).

Decomposed effect	Empirical expression of the decomposed effect	Interpretation
Controlled Direct Effect	$= (p_{10} - p_{00})$	Due neither to mediation nor interaction
Reference Interaction	$= (p_{11}-p_{10}-p_{01}+p_{00})\{P(M=1|A=0)\}$	Due to interaction only
Mediated Interaction	$= (p_{11}-p_{10}-p_{01}+p_{00})\{P(M=1|A=1)-P(M=1|A=0)\}$	Due to mediation and interaction
Pure Indirect Effect	$= (p_{01}-p_{00})\{P(M=1|A=1)-P(M=1|A=0)\}$	Due to mediation only
Total Observed Effect	= sum of components	

The empirical expressions in the table can be transformed into risk ratios (or ORs) by dividing the expressions in the table by $p_{a=0}$ (*ie.* the likelihood of the outcome Y, given exposure A (a=0)). See *vanderWeele (2014)* for the expressions for OR and RR, and *Online Supplementary Text S2* for an elaboration in the CHIC study

According to *vanderWeele (2014)*, the four-way decomposition of effects holds under the following four assumptions:
1. The effect of exposure A is unconfounded conditional on the set of baseline covariates (C)
2. The effect of the mediator M on the outcome Y is unconfounded conditional on (C,A)
3. The effect of the exposure A on the mediator M is unconfounded conditional on C
4. None of the mediator-outcome confounders are themselves affected by the exposure A

The research question underlying the CHIC-study involves an estimation of the effect of drugs on outcomes mediated via the inhibition of hyperinflammation, irrespective of whether this effect was purely due to mediation or to mediation plus interaction. Note that interaction here refers to an interaction between the exposure (immunosuppressive treatment) and the mediator (80% CRP-decline) on the outcomes. Such an interaction would formally imply that the treatment-effect in those *with* an early 80% CRP-decline is different from those *without* an early 80% CRP-decline. From a theoretical point of view this is arguable, since the absence of an 80% CRP-decline in a patient with a favorable outcome does not mean that there has been no CRP-decline at all; the CRP-decline may not have ticked the 80% threshold.

Solving the equations in the table by the probabilities shown in Table 1 yields small (and partly negative) estimations for components 2 (reference interaction) and 3 (mediated interaction), that are negligible in comparison to component 4 (pure indirect effect). According to *VanderWeele* this situation is not uncommon, due to statistical instability (*eg.* low numbers), and it is advised to report all components separately only if they point in the same direction (all positive or all negative). This is why we refrained from interpreting components 2 and 3 separately, and present the sum of components 2, 3 and 4 as the (aggregated) natural indirect effect (NIE).

[1] *VanderWeele TJ.* A unification of mediation and interaction: a four-way decomposition. Epidemiology 2014, 25:749-761

Online Supplementary Text S2

Natural direct and indirect effects expressed as odds ratios in the CHIC-study

1.The OR for the NDE (OR_{NDE}) of the treatment (not via CRP) on the outcome (14-day WHO-improvement or 30-day survival or 90-day survival) is determined only in patients <u>without a CRP-response (M=0)</u>:

OR_{NDE} = likelihood of the outcome in the treated / likelihood of the outcome in the untreated, given M=0

(*or:* how more often does the outcome occur in the treated without a CRP-response than in the untreated without a CRP response)

2.The OR for the NIE (OR_{NIE}) of the treatment (via CRP) on the outcome (14-day WHO-improvement, 30-day survival or 90-day survival) is determined only in patients <u>who have received the treatment (A=1)</u>:

OR_{NIE} = likelihood of the outcome in CRP-responders / likelihood of the outcome in non-CRP-responders, given A=1

(*or:* how more often have CRP-responders the good outcome than non-CRP-responders?)

The OR for the total causal effect of treatment on the outcome (direct and indirect) (14-day WHO-improvement or 30-day survival or 90-day survival) is the product of OR_{NDE} and OR_{NIE}

$OR_{TCE} = OR_{NDE} * OR_{NIE}$