

With great interest did we read the much needed commentary by Poole, Shrier, and VanderWeele¹ on the *empirical* evidence behind claims that the risk difference is more heterogeneous than the odds ratio (i.e., the difference in effects between studies or groups of patients). In their contribution the authors show that the previously reported higher rejection rates of risk difference homogeneity may be explained by differences in power between measurement scales, an issue we previously addressed as well². Without detracting from their contribution or conclusions, with which we agree, we were surprised that the authors omitted the *theoretical* grounds why the odds ratio is thought to be a less heterogeneous measure.

Essentially, our comment is that the odds ratio (OR) can be homogeneous given any conceivable distribution of risk in patient subgroups, while the risk ratio (RR) and risk difference (RD) cannot. Take, for example, the hypothetical study mentioned by Poole et al., where the *average* risk in the control group was 0.27, and the *average* risk in the treated group 0.46 (scenario1), resulting in an OR of 2.30, a RR of 1.70, and a RD of 0.19. If we now think it acceptable that the risk for any subgroup of control subjects is only bounded by 0 and 1, given the RD of 0.19, there cannot be any subgroup of control subjects with a risk higher than 0.81 unless the RD is heterogeneous. In general the RD can be homogenous if the control group risk is bounded between $\max(0, 0-RD)$ and $\min(1-RD, 1)$. Similarly, for the RR to be homogenous, the risk in any subgroup of control subjects should be bounded between $0/RR$ and $1/RR$, for the current example 0 and 0.59. The OR, however, can be homogeneous for any control group risk between 0 and 1. Given that the OR never “forces” heterogeneity, one might expect the OR to be the least heterogeneous in *empirical* settings as well. However, one may question whether these bounds are actually violated often in *empirical* settings; therefore we agree with Poole et al. that there is insufficient *empirical* evidence to claim any effect measure induced heterogeneity, and furthermore that comparisons between scales are difficult. In the end, sound biological reasoning on potential pathways may provide the most suitable grounds for choosing

an effect measure, not mathematical or statistical properties. However, to facilitate such a choice, knowledge of these properties is, we feel, essential.

Reference List

- (1) Poole C, Shrier IF, VanderWeele TJ. Is the Risk Difference Really a More Heterogeneous Measure?(1531-5487 (Electronic)).

- (2) Schmidt AF, Groenwold RH, Knol MJ et al. Exploring interaction effects in small samples increases rates of false-positive and false-negative findings: results from a systematic review and simulation study. *J Clin Epidemiol.* 2014;67(7):821-829.