

1. Credibility is higher if the investigators stated a hypothesis prior to performing the study
2. Credibility increases with the prior probability of the effect modification being real
3. Credibility is higher if prior knowledge was integrated through a formal Bayesian analysis
4. Credibility is higher if investigators correctly anticipated the direction of the subgroup effect
5. Credibility is higher if there is a compelling causal model explaining the effect modification
6. Credibility is higher if indirect evidence supports the effect modification
7. Credibility is higher if content experts were involved in the selection of effect modifiers
8. Credibility is higher if the effect modification is consistent across independent studies
9. Credibility is higher if the effect modification is consistent across related outcomes
10. Credibility is higher if analytic details have been pre-specified
11. Credibility is higher if investigators accounted formally or informally for multiplicity
12. Credibility is higher if all performed analyses of effect modification and results are reported
13. Credibility is higher if only a small number of effect modifiers have been tested
14. Credibility is higher for effect modification of the primary rather than secondary outcomes
15. Credibility is higher if an interaction test suggests a small likelihood for a chance finding
16. Credibility is higher if an effect modification is independent from another effect modification
17. Credibility increases with the power to detect effect modification
18. Credibility is lower if the overall treatment effect is at risk of bias
19. Credibility is higher if, within each subgroup, prognostic factors are balanced
20. Credibility is higher if the effect modifier was used as a stratification variable at randomization
21. Credibility is higher if the sample size is large and balanced across subgroups under consideration
22. Credibility is higher if the effect modification is large
23. Credibility is higher if the effect modifier is a characteristic measured at randomization
24. Credibility is higher if the effect modifier is a cause of the outcome as opposed to being a proxy
25. Credibility is higher if the effect modification was identified in individual participant data
26. Credibility is higher if the effect modifier was measured without error
27. Credibility is higher if continuous outcomes were not categorized
28. Credibility is lower if a threshold is unnatural and not justified
29. Credibility is higher if the researchers considered the appropriateness of the model, such as non-linearity
30. Credibility is higher if there is a dose-response relationship across ordered levels of an effect modifier
31. Credibility is higher if the effect modification is quantitative rather than qualitative
32. Credibility is higher if the model allows for true variation within a subgroup
33. Credibility is lower if the effect modifier is the control group risk of an outcome
34. Credibility is lower if presence of effect modification depends on the scale
35. Credibility is higher if a sensitivity analysis suggests robustness to relevant assumptions
36. Credibility is higher if the overall effect is statistically significant

Table 1. Candidate credibility criteria identified in a systematic survey of the methodological literature