# A new instrument for assessing the credibility of effect modifiers

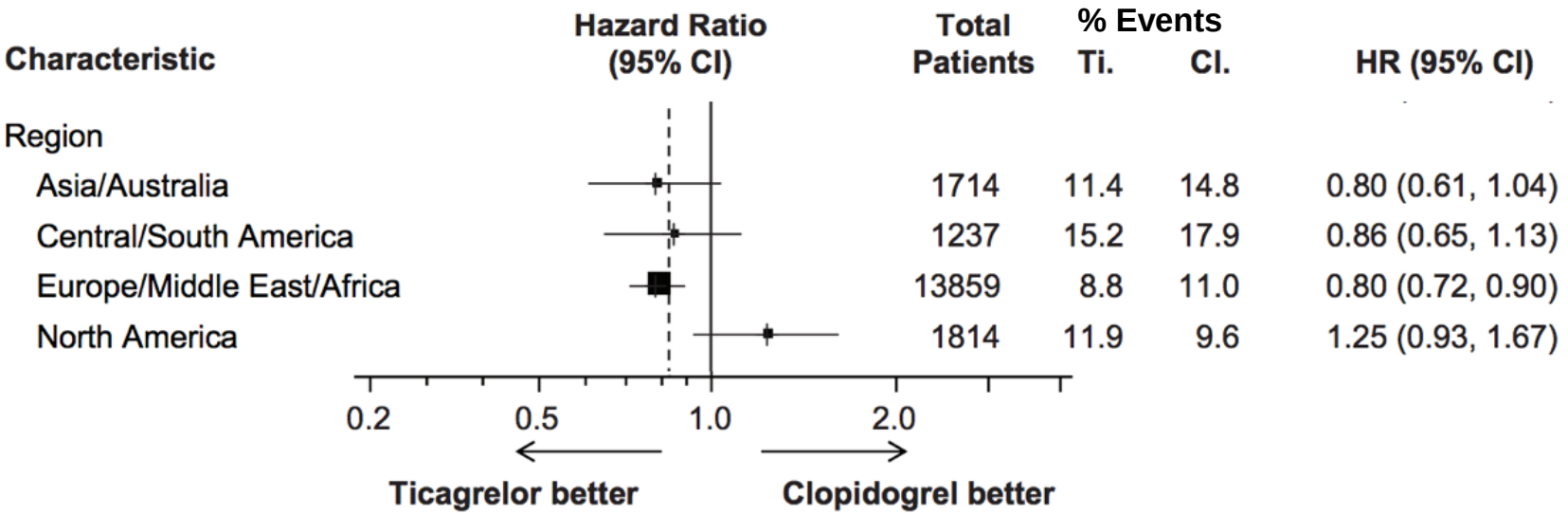
Stefan Schandelmaier, Matthias Briel, Xin Sun, Yaping Chang, Behnam Sadeghirad, Wojtek Wiercioch, Tahira Devji, Romina Brignardello, Farid Foroutan, Gordon Guyatt

McMaster University

# PLATO trial

N Engl J Med. 2009
AstraZeneca

Ticagrelor versus Clopidogrel
>18000 Patients with acute coronary syndrome
Cardiovascular events: HR 0.84 (0.77 to 0.92)




Low risk of bias
Interaction test P=0.045

No a priori hypothesis
33 subgroup analysis (3 P≤0.05)

| | | |
|---|---|---|
| <1980 | Rothman, Greenland, Walter, et al. | **Concept** of effect modification |
| 1987 | Pocock et al. | Statistical problems |
| 1991 | Yusuf et al. | **Criteria** for critical appraisal |
| 1992 | Guyatt, Oxman | **Criteria** for critical appraisal |
| 2002 | Higgins, Thompson | Limitations of **meta-regression** |
| 2005 | Rothman et al. | Series in Lancet |
| 2010 | Kent, Rothwell, Ioannidis, Altman, et al. | Framework **reporting** and analysis |
| 2010 | Sun et al. | More criteria |
| 2011 | Guyatt, Oxman, Schünemann et al. | **GRADE** guidance subgroup analysis |
| 2014 | Koch, Keene, Wang et al. | Series in J Biopharm Stat |
| 2015 | VanderWeele et al. | **Causal** interaction |
| 2015 | Burke et al. | **Bayesian** credibility assessment |
| 2016 | Wallach et al. | **Empirical evidence** of spurious findings |

# What's new?

Formal instrument

- Items, response options, overall
- RCTs and MAs
- Compatible with GRADE

Systematic survey of credibility criteria

Expert consensus

User testing, reliability study

# Systematic survey

1730 journal articles + 56 text book chapters

150 reviewed in detail

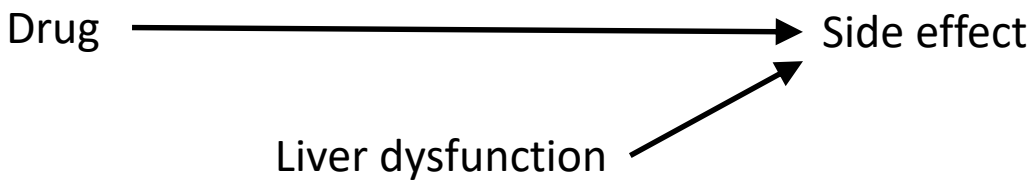
55 articles reporting 35 candidate items

- Chance
- Bias
- Rationale

# Chance (random error)

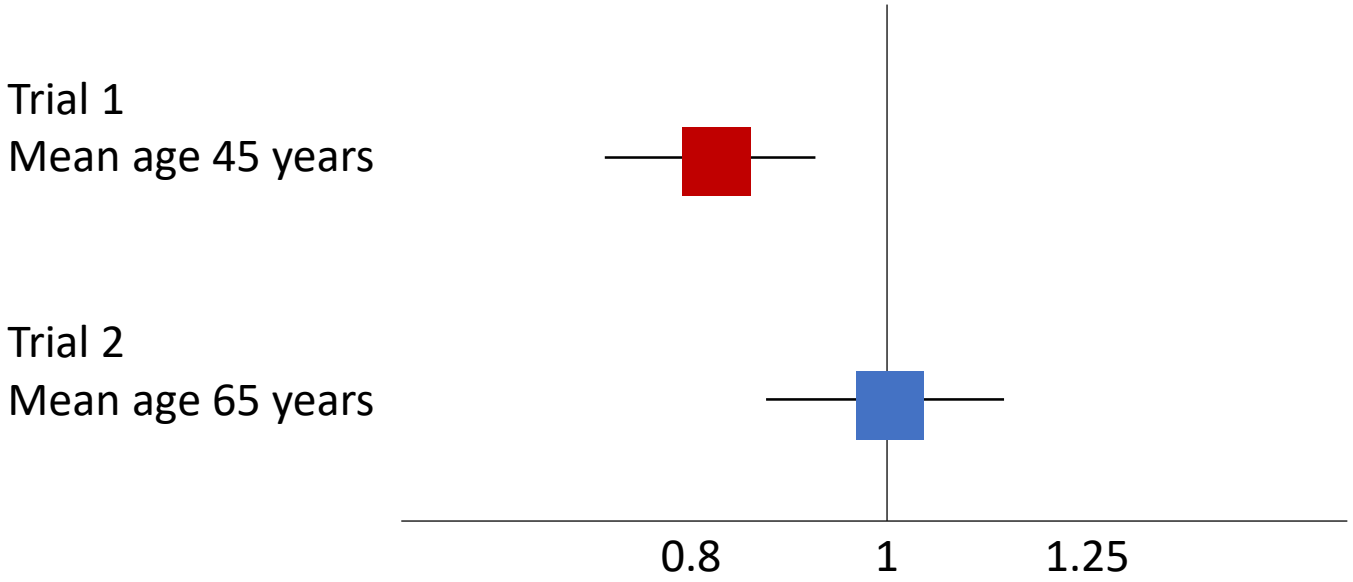
| | |
|---|---|
| Replication across studies | 13 (of 55) |
| Significant interaction test | 15 |
| Small number of subgroup analyses | 17 |
| Pre-specified analytic details | 15 |
| Adjustment for multiplicity | 8 |
| Reporting of all analysis | 6 |

# Bias (is it what we think it is?)

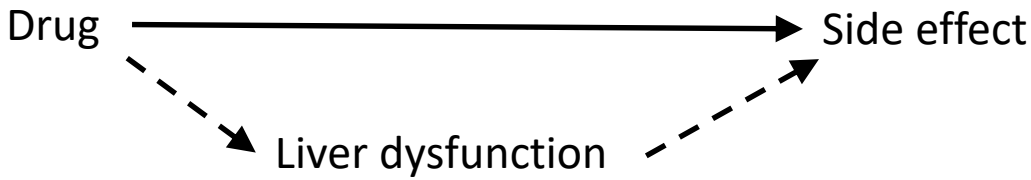



Baseline characteristic 7

Within vs between 7

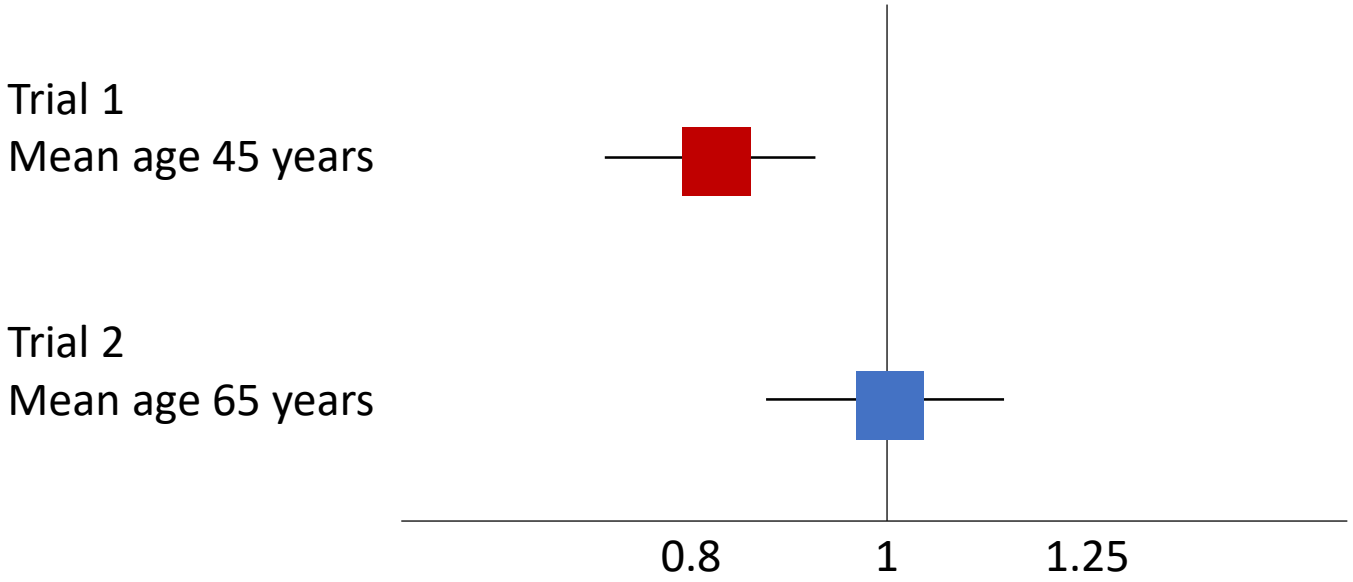


# Bias (is it what we think it is?)

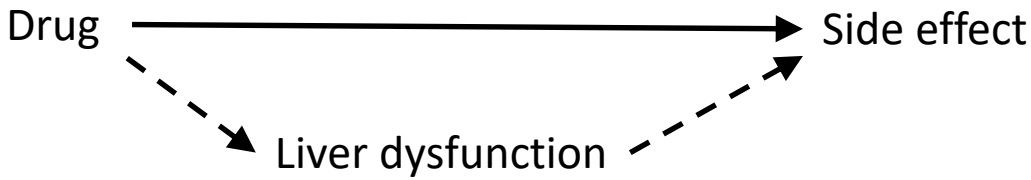



Baseline characteristic 7

Within vs between 7

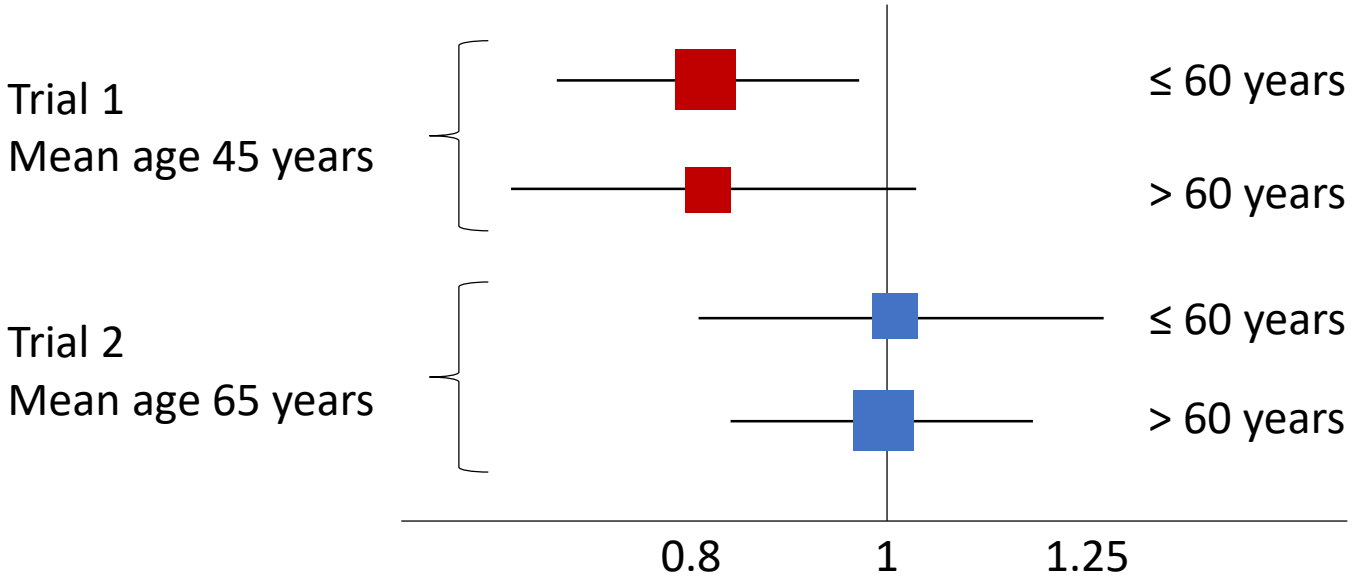


# Bias (is it what we think it is?)

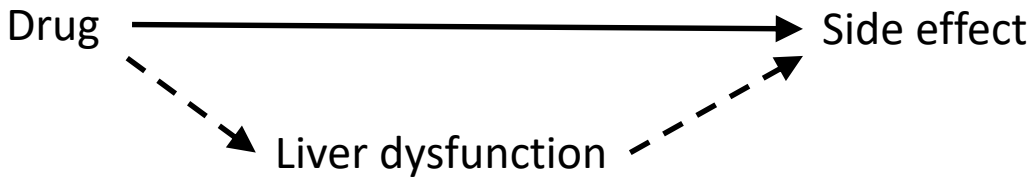



Baseline characteristic 7

Within vs between 7

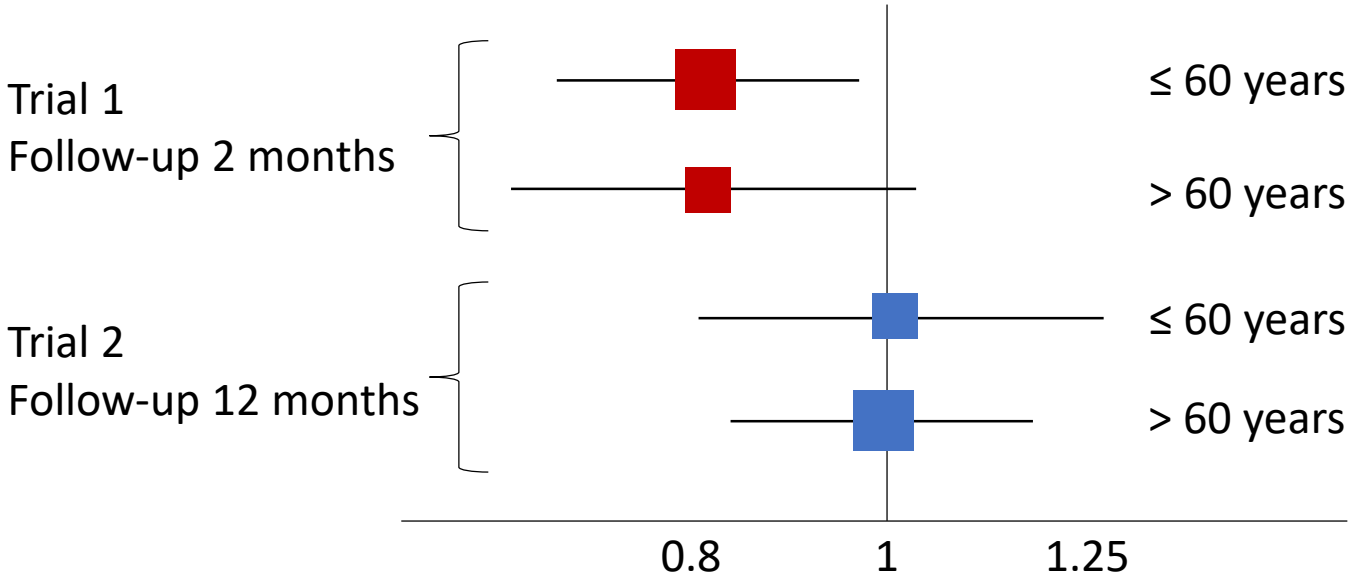


# Bias (is it what we think it is?)

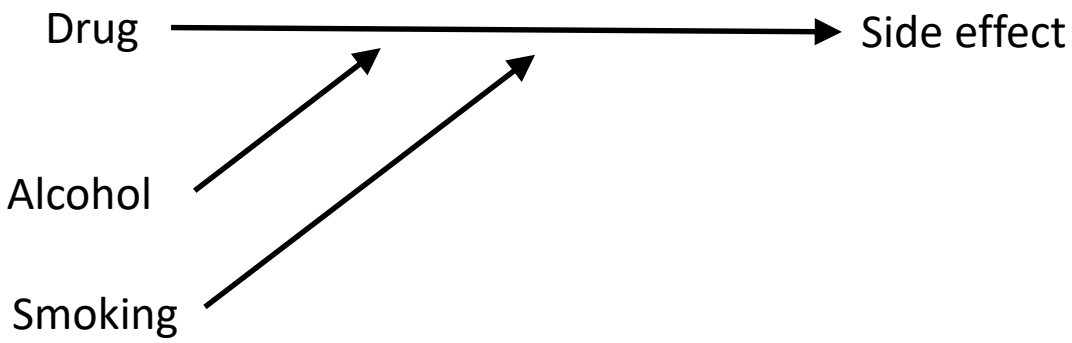



Baseline characteristic 7

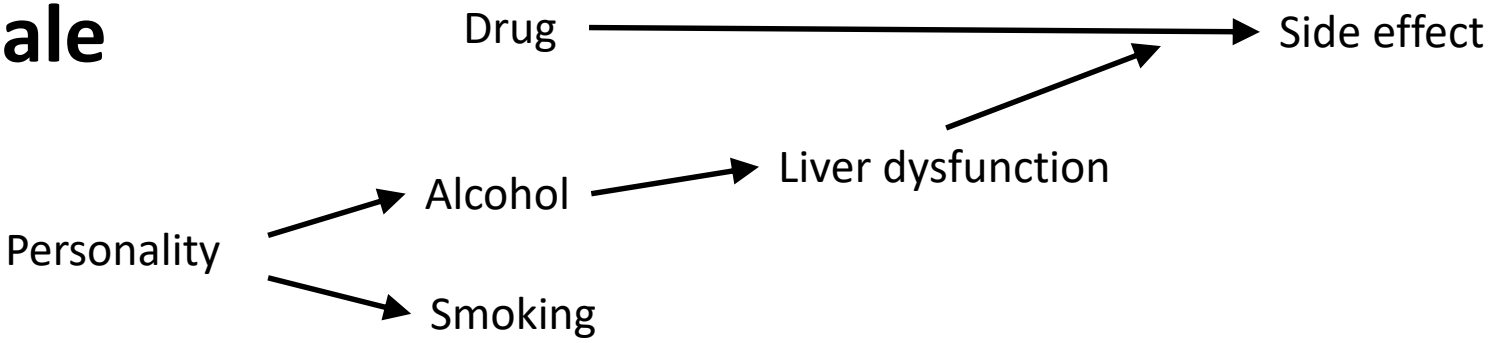
Within vs between 7



# Rationale

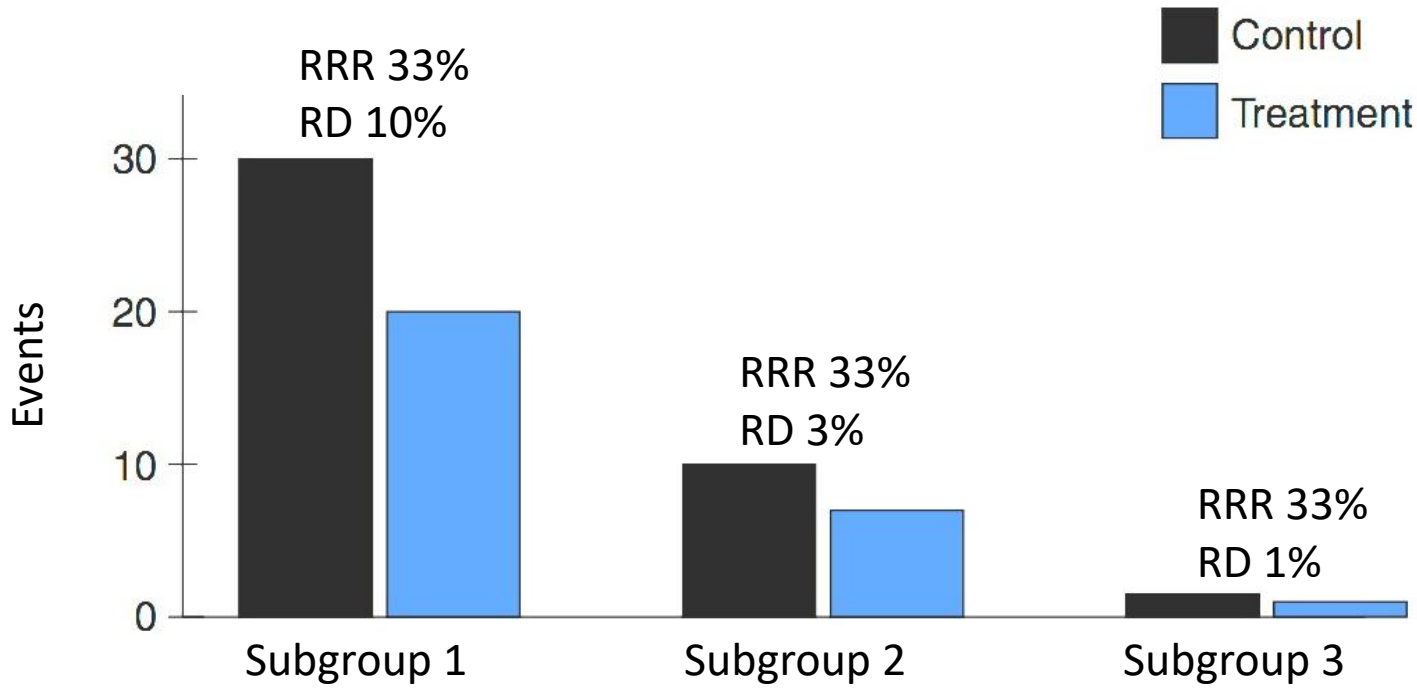


# Rationale



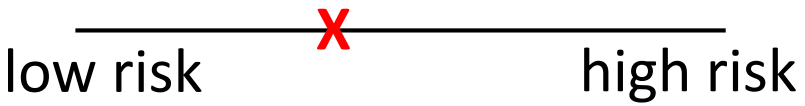

| | |
|---|---|
| Rationale preceded analysis | 15 |
| Specified direction | 15 |
| Expert input | 4 |
| Indirect evidence | 4 |
| Causal factor not proxy | 6 |

# Effect modification is scale-dependent

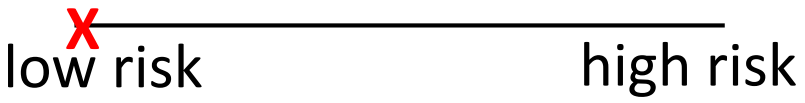


**Scale of interest:**

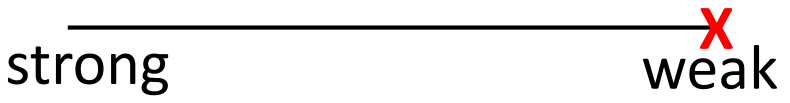
**Chance**



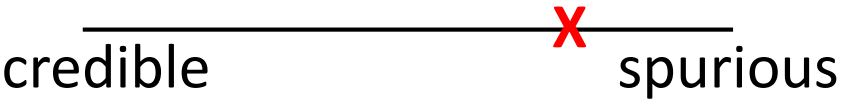

**Bias**




**Rationale**




**Overall, how credible is the effect modifier?**



# Strengths and limitations

✓ Systematic survey

✓ Expert consensus

✓ Reliability

Validity

✓ Presence

Absence

( ) Relevance

# Bottom line

- Credibility of subgroup analyses has taken up a lot of intellectual energy
- No single criterion will be sufficient to establish credibility
- Formal instrument for overall credibility