Lancaster Environment Centre | Lancaster University

Systematic reviews and preventing the misuse of Bradford Hill criteria

14 October 2016 | Royal Society of Medicine, London

Paul Whaley | Lancaster Environment Centre

p.whaley@lancaster.ac.uk

About me

- Background in environmental health advocacy and science communication
- Introduced to systematic reviews as gold-standard approach to evidence synthesis in early 2010
- Advocating use of SR methods to advance validity of results of chemical risk assessments
- Associate Editor for Systematic Reviews at *Environment International* (submissions please!)
- Research into quality assurance and control in conduct and publication of evidence syntheses: how do we ensure only high quality reviews get published?

Bradford Hill “use and misuse”

- How do we ensure that, when people are evaluating the strength of a body of evidence, they are doing so appropriately?

What I want when I read evidence syntheses

- As reader, I want to know:
  - Has everything been considered which ought to have been?
  - Has it been considered properly?
- To ensure that it’s the evidence, not the reviewer, which is causative in the outcome of the review
  - Like a lab experiment: it should be the change in conditions between intervention and control groups which causes the change in outcome
- BH gives us a list of stuff which we ought to be considering, and guidance on how to consider it
- But on its own, it’s not a process: sports equipment without a rulebook

Don’t want naïve processes

- For example, BH checklist and the Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses
- Shown empirically that scores and scales don’t work (1)
  - » Results contingent on choice of scale, not evidence reviewed
- Shown theoretically that they don’t work (2)
  - » Effect of error should be contingent on study context, not choice of scale
- Plus, arbitrarily simple and can conceal important information (3)

(1) Juni et al. 1999, *BMJ*
(2) Greenland & O’Rourke 2001, *Biostatistics*
(3) Higgins et al. 2011, *BMJ*

NG, OHAT, SYRINA: non-naïve processes

- NG, OHAT, GRADE and SYRINA are not checklists but processes for systematically accounting for important features of a body of evidence, and consistently interpreting those features into a description of how compelling that evidence is

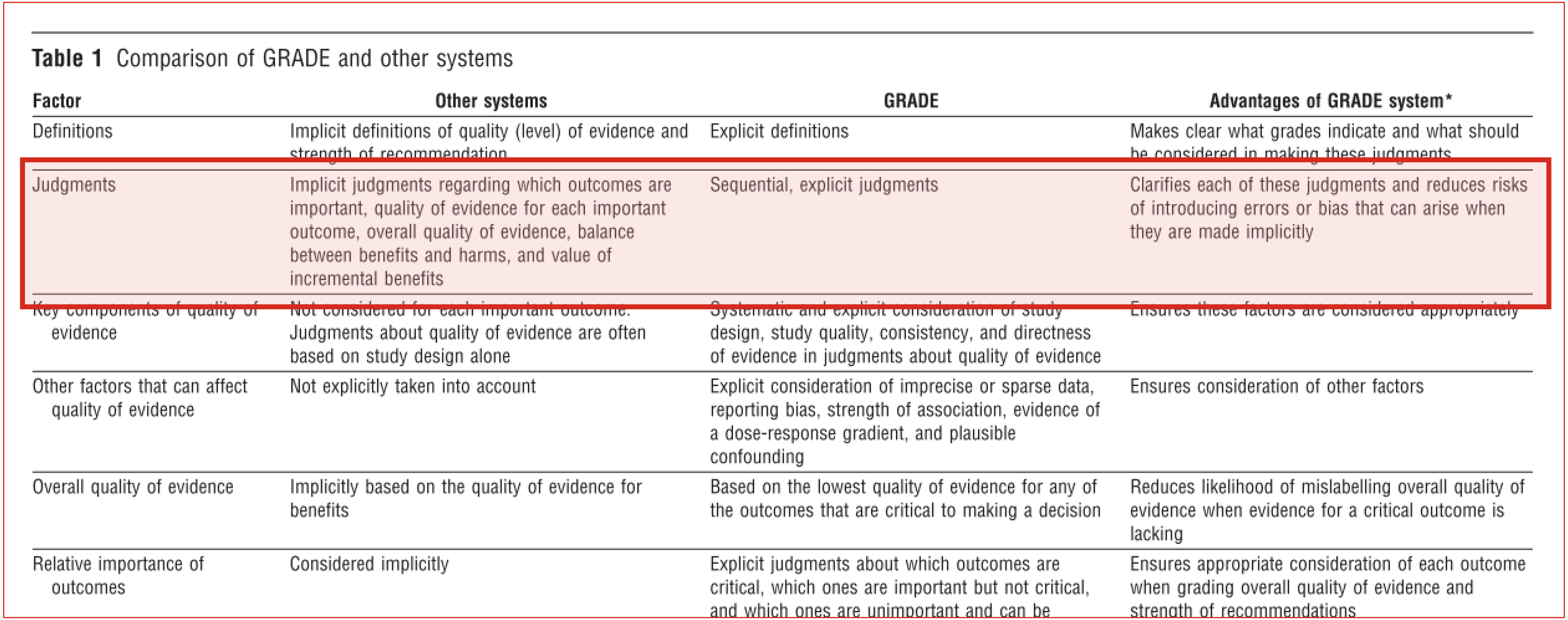
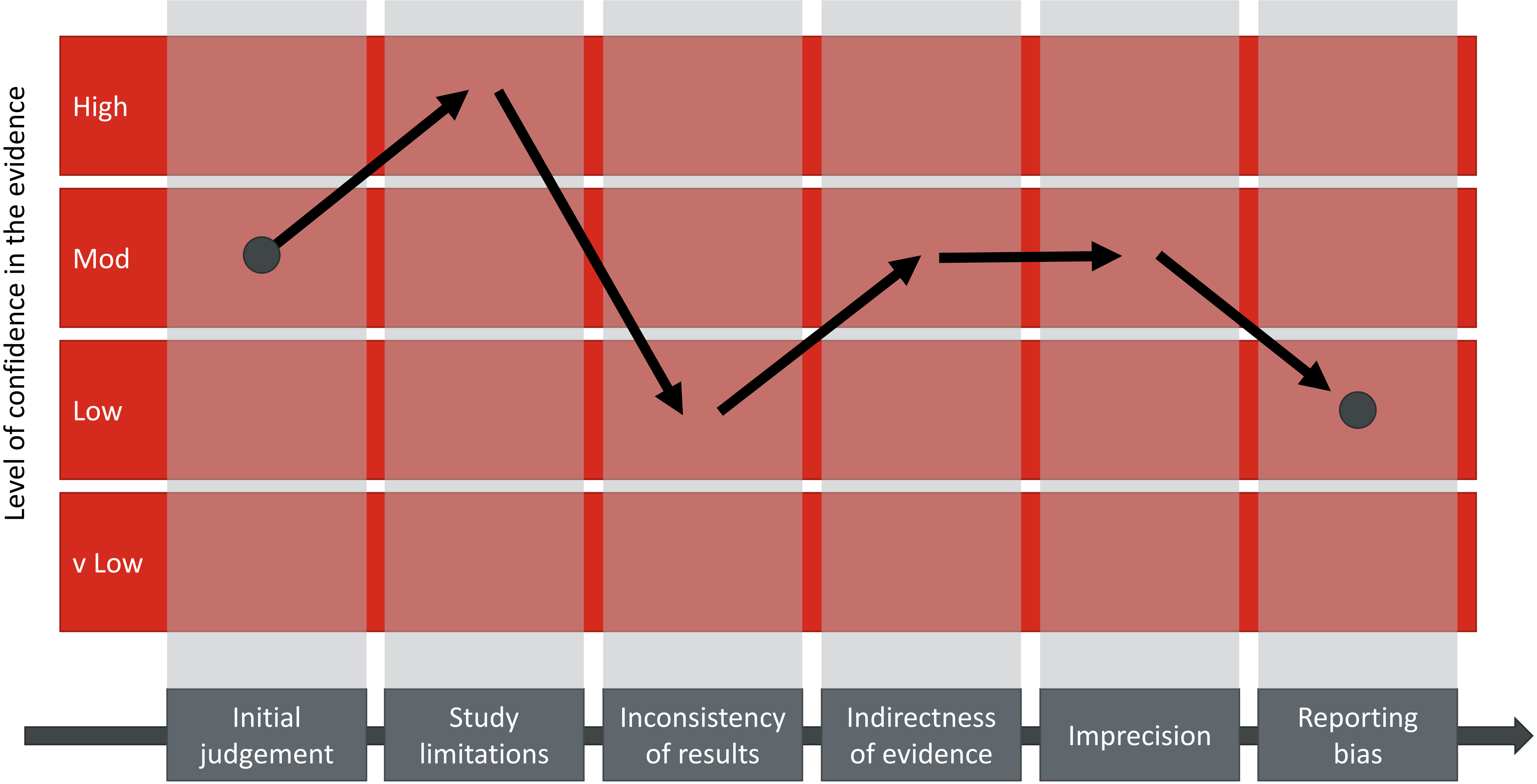


Table 1 Comparison of GRADE and other systems

Factor	Other systems	GRADE	Advantages of GRADE system*
Definitions	Implicit definitions of quality (level) of evidence and strength of recommendation	Explicit definitions	Makes clear what grades indicate and what should be considered in making these judgments
Judgments	Implicit judgments regarding which outcomes are important, quality of evidence for each important outcome, overall quality of evidence, balance between benefits and harms, and value of incremental benefits	Sequential, explicit judgments	Clarifies each of these judgments and reduces risks of introducing errors or bias that can arise when they are made implicitly
Key components of quality of evidence	Not considered for each important outcome. Judgments about quality of evidence are often based on study design alone	Systematic and explicit consideration of study design, study quality, consistency, and directness of evidence in judgments about quality of evidence	Ensures these factors are considered appropriately
Other factors that can affect quality of evidence	Not explicitly taken into account	Explicit consideration of imprecise or sparse data, reporting bias, strength of association, evidence of a dose-response gradient, and plausible confounding	Ensures consideration of other factors
Overall quality of evidence	Implicitly based on the quality of evidence for benefits	Based on the lowest quality of evidence for any of the outcomes that are critical to making a decision	Reduces likelihood of mislabelling overall quality of evidence when evidence for a critical outcome is lacking
Relative importance of outcomes	Considered implicitly	Explicit judgments about which outcomes are critical, which ones are important but not critical, and which ones are unimportant and can be	Ensures appropriate consideration of each outcome when grading overall quality of evidence and strength of recommendations

GRADE Working Group, *BMJ* 2004



Level of confidence in the evidence

High

Mod

Low

v Low

Initial judgement

Study limitations

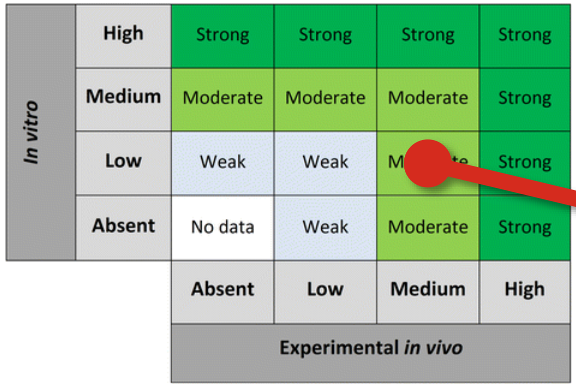
Inconsistency of results

Indirectness of evidence

Imprecision

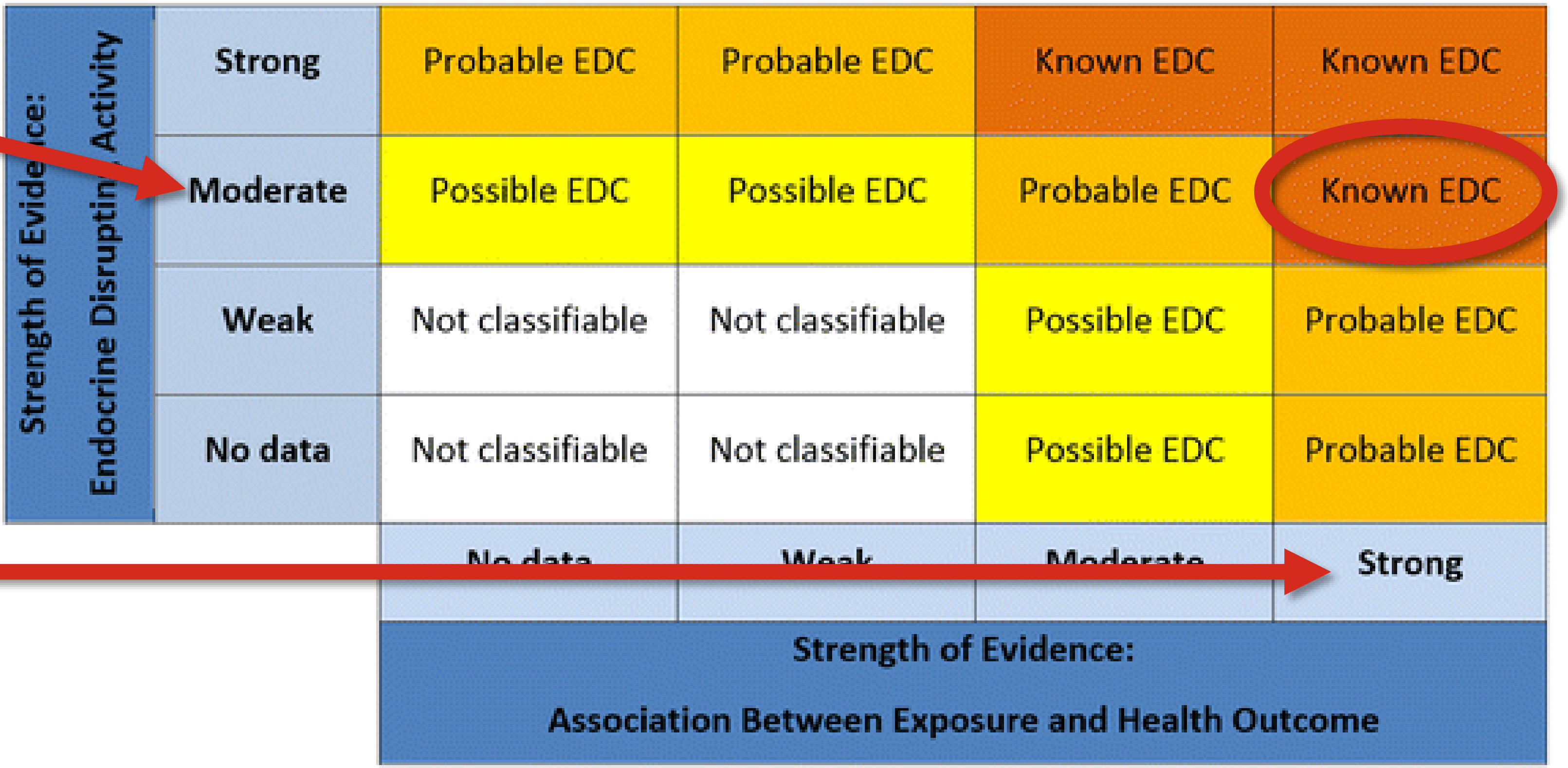
Reporting bias

SYRINA

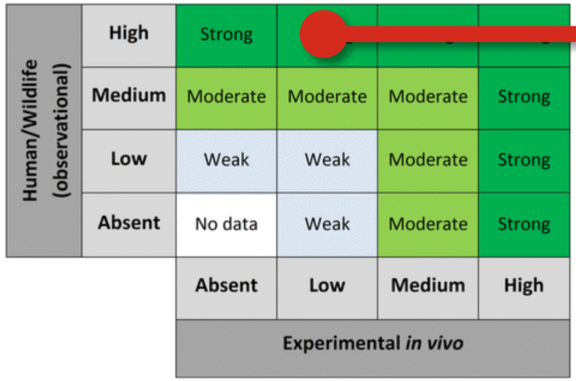


In vitro	High	Strong	Strong	Strong	Strong
	Medium	Moderate	Moderate	Moderate	Strong
	Low	Weak	Weak	Moderate	Strong
	Absent	No data	Weak	Moderate	Strong
		Absent	Low	Medium	High
		Experimental *in vivo*			

Strength of evidence: ED activity



Strength of Evidence: Endocrine Disrupting Activity	Strong	Probable EDC	Probable EDC	Known EDC	Known EDC
	Moderate	Possible EDC	Possible EDC	Probable EDC	Known EDC
	Weak	Not classifiable	Not classifiable	Possible EDC	Probable EDC
	No data	Not classifiable	Not classifiable	Possible EDC	Probable EDC
		No data	Weak	Moderate	Strong
		Strength of Evidence: Association Between Exposure and Health Outcome			



Human/Wildlife (observational)	High	Strong	Strong	Strong	Strong
	Medium	Moderate	Moderate	Moderate	Strong
	Low	Weak	Weak	Moderate	Strong
	Absent	No data	Weak	Moderate	Strong
		Absent	Low	Medium	High
		Experimental *in vivo*			

Strength of evidence: health outcome

Vandenberg et al. 2016, *Env Health*

Algorithms are scientific

- To an extent it is algorithmic, but it is not like a checklist or NOS, because the input determines the output, not the process itself.
- It is transparent: if the process is producing duff results, (a) this is scrutable, (b) the process can be critiqued and adjusted

Can’t opt out of process

- There is always a process
- If you use the BH considerations and come to a conclusion, you have followed a reasoning process, you have just kept it secret
  - » What did you put most weight on? Why?
  - » How much did it affect your conclusions?
  - » Would I or anyone else come to the same conclusions?
- Secret methods have no place in science: cannot audit them or improve them, and therefore cannot determine whether criteria are being used or misused
- If you reject “algorithms”, yet want to police the misuse of BH, then you are rejecting the very thing that will help you