Bias in Epidemiological Studies: the big picture

Madhukar Pai, MD, PhD
Professor
Department of Epidemiology & Biostatistics
McGill University, Montreal, Canada
Email: madhukar.pai@mcgill.ca

A family of biases

GLOSSARY

Bias

Miguel Delgado-Rodríguez, Javier Llorca

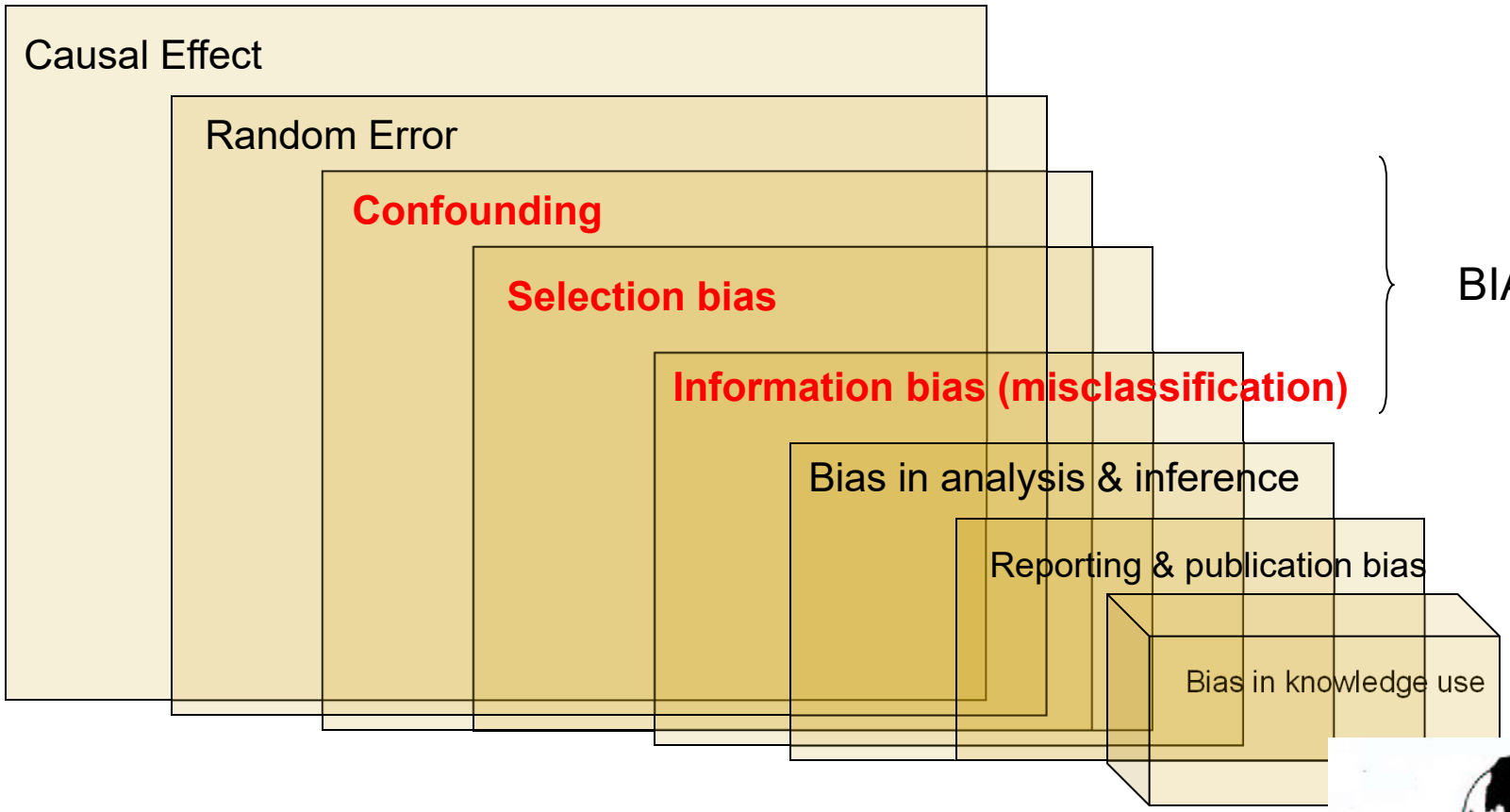
J Epidemiol Community Health 2004;58:635–641. doi: 10.1136/jech.2003.008466

The concept of bias is the lack of internal validity or incorrect assessment of the association between an exposure and an effect in the target population in which the statistic estimated has an expectation that does not equal the true value. Biases can be classified by the research

Table 1 Alphabetical list of biases, indicating their type and the design where they can occur

Specific name of bias	Group of bias	Subgroup of bias (next level to specific name)	Type of design affected
Allocation of intervention bias	Execution of an intervention		Trial
Apprehension bias	Information bias	Observer bias	All studies
Ascertainment bias	Selection bias	Inappropriate definition of the eligible population	Observational study
Berkson's bias	Selection bias	Inappropriate definition of the eligible population	Hospital based case-control study
Centripetal bias	Selection bias	Healthcare access bias	Observational study
Citation bias	Selection bias	Lack of accuracy of sampling frame	Systematic review/meta-analysis
Competing risks	Selection bias	Ascertainment bias	All studies
Compliance bias	Execution of an intervention		Trial
Confounding by group	Confounding		Ecological study
Confounding by indication	Confounding		Case-control study, cohort study
Contamination bias	Execution of an intervention		Trial, mainly community trials
Detection bias	Selection bias	Uneven diagnostic procedures in the target population	Case-control study
Detection bias	Information bias	Misclassification bias	Cohort study
Diagnostic/treatment access bias	Selection bias	Healthcare access bias	Observational study
Diagnostic suspicion bias	Selection bias	Detection bias	Case-control study
Diagnostic suspicion bias	Information bias	Detection bias	Cohort study
Differential maturing			Trial
Differential misclassification bias	Information bias	Misclassification bias	All studies
Dissemination bias	Selection bias	Lack of accuracy of sampling frame	Systematic review/meta-analysis
Ecological fallacy	Information bias		Ecological study
Exclusion bias	Selection bias	Inappropriate definition of the eligible population	Case-control study
Exposure suspicion bias	Information bias	Recall bias	Case-control study
Family aggregation bias	Information bias	Reporting bias	Observational study
Friend control bias	Selection bias	Inappropriate definition of the eligible population	Case-control study
Hawthorne effect	Information bias		Trial
Healthcare access bias	Selection bias	Ascertainment bias	Observational study
Healthy volunteer bias	Selection bias	Non-response bias	Observational study
Healthy worker effect	Selection bias	Inappropriate definition of the	Cohort study (mainly

The long road to causal inference (the "big picture")



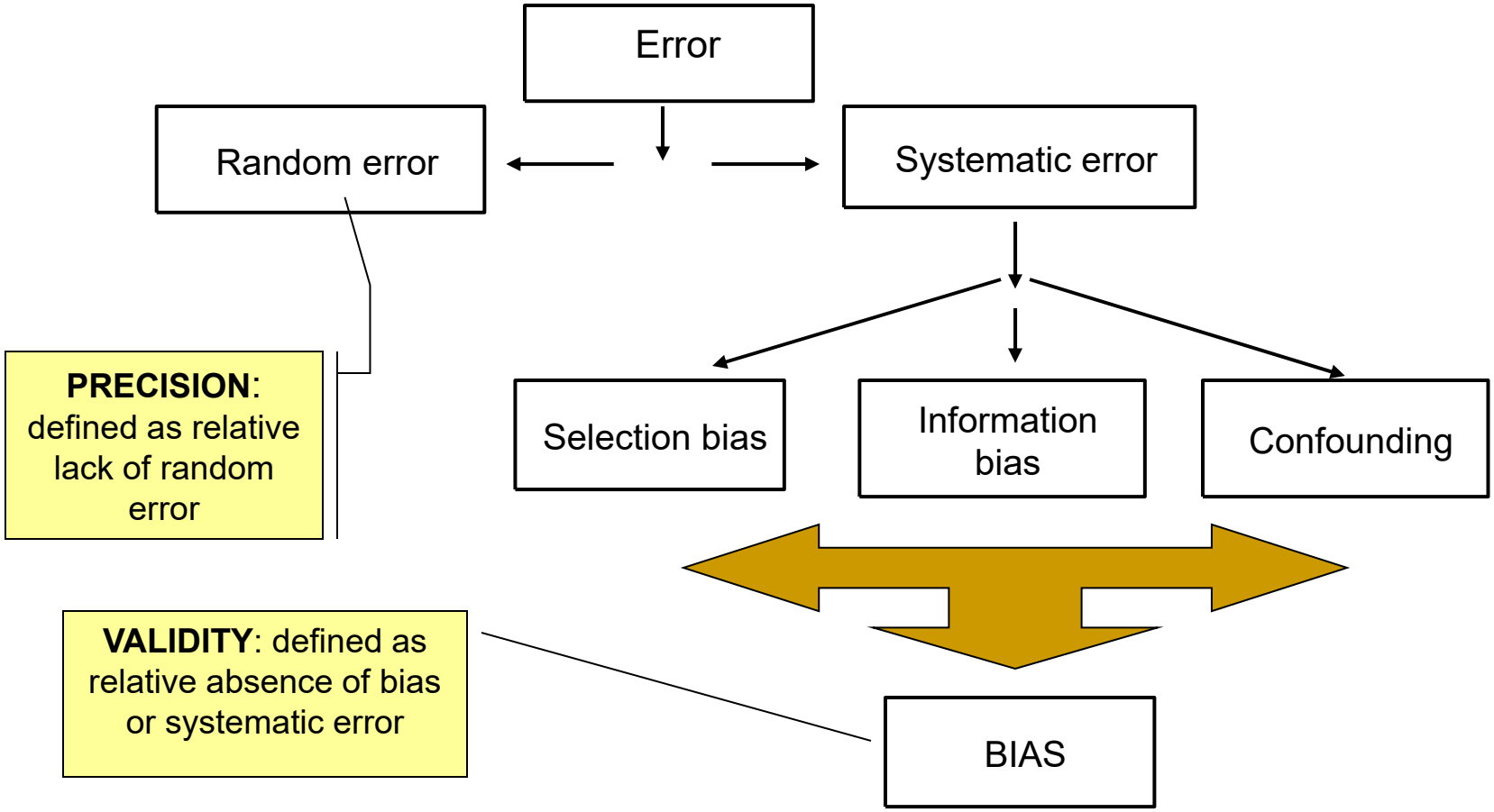
RR_{causal} "truth" [counterfactual]

the long road to causal inference…

$RR_{association}$

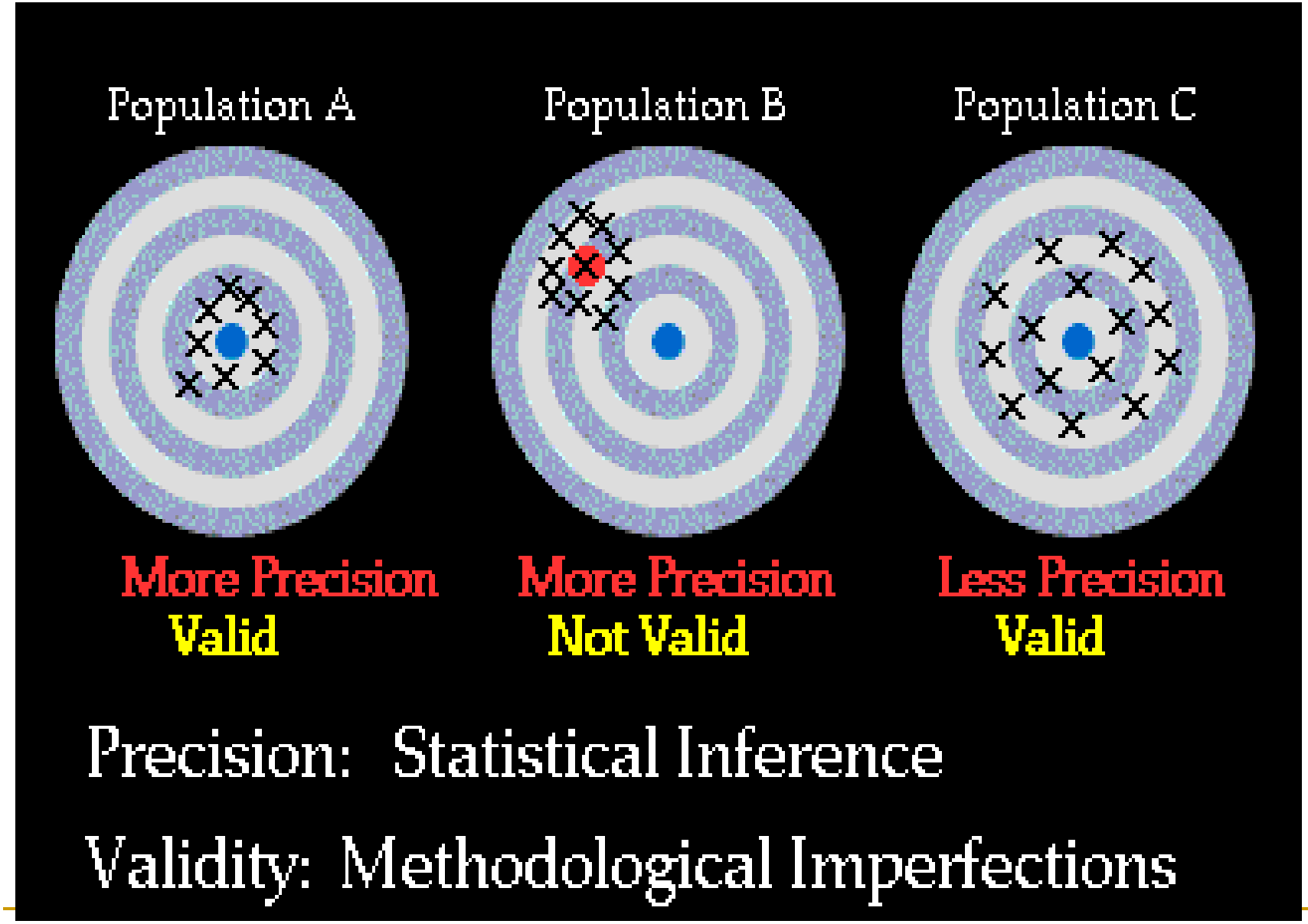
Adapted from: Maclure, M, Schneeweis S. Epidemiology 2001;12:114-122.

Errors in epidemiological inference



"Bias is any process at any stage of inference which tends to produce results or conclusions that differ systematically from the truth" – Sackett (1979)

"Bias is systematic deviation of results or inferences from truth." [Porta, 2008]


Population A
Population B
Population C
More Precision
Valid
More Precision
Not Valid
Less Precision
Valid
Precision: Statistical Inference
Validity: Methodological Imperfections

Quantitative definition of bias

- Effect estimate in the source population (parameter of interest – "the truth") = θ
- Effect estimate in the actual study sample (effect estimate from study) = $\hat{\theta}$

$\hat{\theta}$ is a biased estimate of θ if $\hat{\theta} - \theta \neq 0$

Direction of bias

- ***Positive bias*** – observed effect is higher than the true value (causal effect)
- ***Negative bias*** – observed effect is lower than the true value (causal effect)

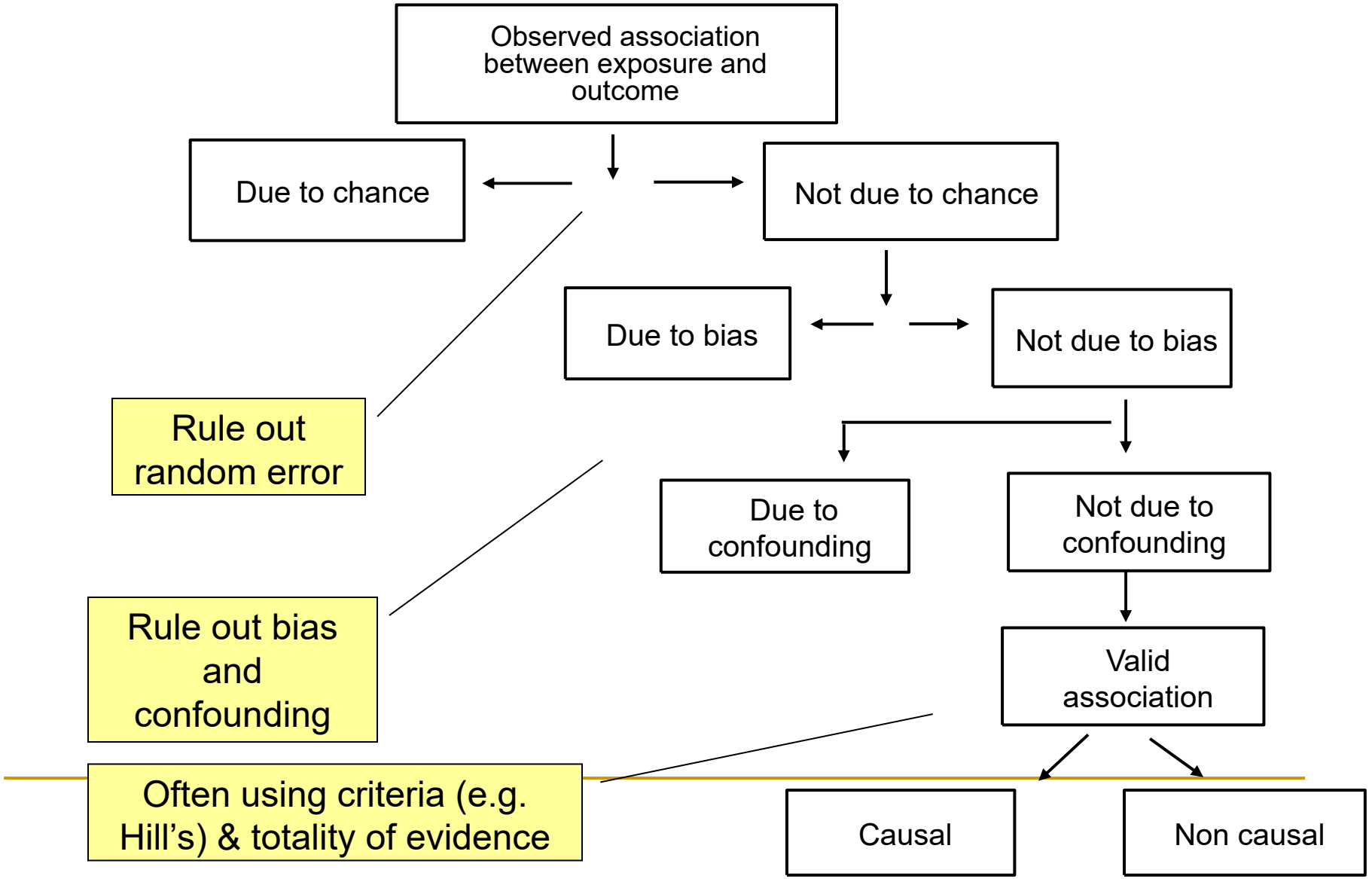
A better approach would be:

- ***Bias towards the null*** – observed value is closer to 1.0 than is the true value (causal effect)*

- ***Bias away from the null*** – observed value is farther from 1.0 than is the true value (causal effect)*

*Note: 1 is the null value for ratio measures (e.g. OR, RR), but not for risk difference Measures (where null value is 0)

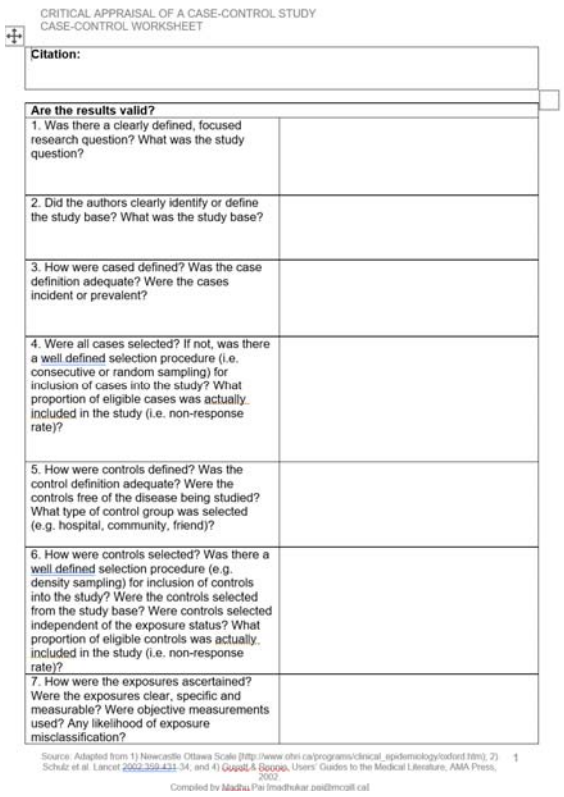
A Skeptic's Algorithm for Associations



The key biases we look for when we read a paper, depends on the study design

- Sources of bias in RCTs:
  - Improper randomization
  - Lack of blinding
  - Attrition
- Sources of bias in case-control studies
  - How were cases and controls selected?
  - Was information collected using same methods in both cases and controls?
  - Was confounding addressed?

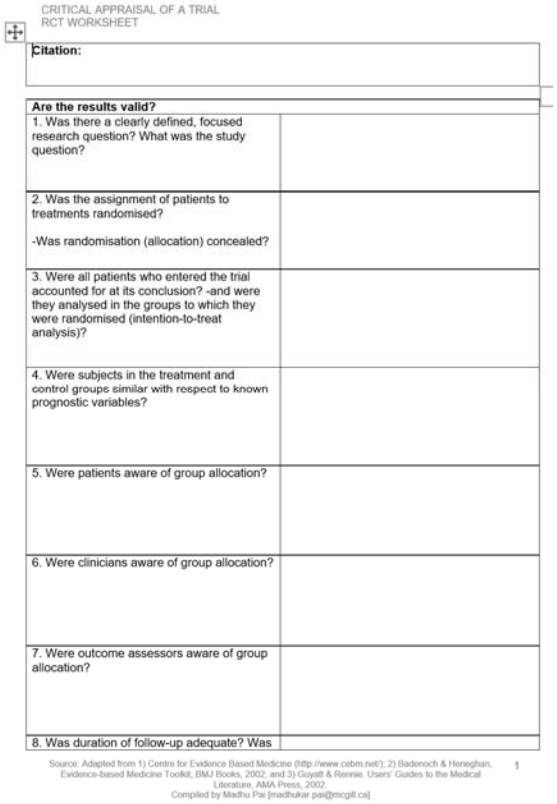
We have critical appraisal worksheets for each study design



CRITICAL APPRAISAL OF A CASE-CONTROL STUDY
CASE-CONTROL WORKSHEET

Citation:	
Are the results valid?	
1. Was there a clearly defined, focused research question? What was the study question?	
2. Did the authors clearly identify or define the study base? What was the study base?	
3. How were cased defined? Was the case definition adequate? Were the cases incident or prevalent?	
4. Were all cases selected? If not, was there a well defined selection procedure (i.e. consecutive or random sampling) for inclusion of cases into the study? What proportion of eligible cases was actually included in the study (i.e. non-response rate)?	
5. How were controls defined? Was the control definition adequate? Were the controls free of the disease being studied? What type of control group was selected (e.g. hospital, community, friend)?	
6. How were controls selected? Was there a well defined selection procedure (e.g. density sampling) for inclusion of controls into the study? Were the controls selected from the study base? Were controls selected independent of the exposure status? What proportion of eligible controls was actually included in the study (i.e. non-response rate)?	
7. How were the exposures ascertained? Were the exposures clear, specific and measurable? Were objective measurements used? Any likelihood of exposure misclassification?	

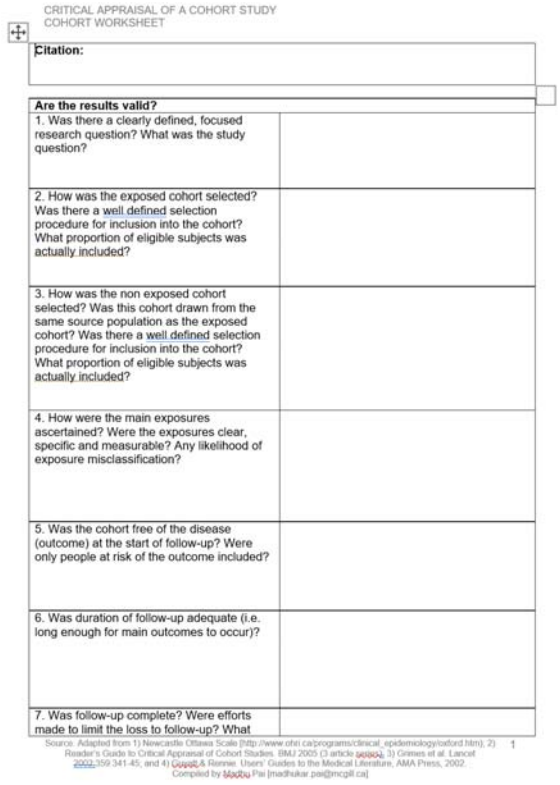
Source: Adapted from 1) Newcastle Ottawa Scale [http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm); 2) Schulz et al. Lancet 2002;359:431-34; and 4) Guyatt & Rennie, Users' Guides to the Medical Literature, AMA Press, 2002.
Compiled by Madhu Pai [madhukar.pai@mcgill.ca] 1



CRITICAL APPRAISAL OF A TRIAL
RCT WORKSHEET

Citation:	
Are the results valid?	
1. Was there a clearly defined, focused research question? What was the study question?	
2. Was the assignment of patients to treatments randomised? -Was randomisation (allocation) concealed?	
3. Were all patients who entered the trial accounted for at its conclusion? -and were they analysed in the groups to which they were randomised (intention-to-treat analysis)?	
4. Were subjects in the treatment and control groups similar with respect to known prognostic variables?	
5. Were patients aware of group allocation?	
6. Were clinicians aware of group allocation?	
7. Were outcome assessors aware of group allocation?	
8. Was duration of follow-up adequate? Was	

Source: Adapted from 1) Centre for Evidence Based Medicine (http://www.cebm.net/); 2) Badenoch & Heneghan, Evidence-based Medicine Toolkit, BMJ Books, 2002, and 3) Guyatt & Rennie. Users' Guides to the Medical Literature, AMA Press, 2002.
Compiled by Madhu Pai [madhukar.pai@mcgill.ca] 1



CRITICAL APPRAISAL OF A COHORT STUDY
COHORT WORKSHEET

Citation:	
Are the results valid?	
1. Was there a clearly defined, focused research question? What was the study question?	
2. How was the exposed cohort selected? Was there a well defined selection procedure for inclusion into the cohort? What proportion of eligible subjects was actually included?	
3. How was the non exposed cohort selected? Was this cohort drawn from the same source population as the exposed cohort? Was there a well defined selection procedure for inclusion into the cohort? What proportion of eligible subjects was actually included?	
4. How were the main exposures ascertained? Were the exposures clear, specific and measurable? Any likelihood of exposure misclassification?	
5. Was the cohort free of the disease (outcome) at the start of follow-up? Were only people at risk of the outcome included?	
6. Was duration of follow-up adequate (i.e. long enough for main outcomes to occur)?	
7. Was follow-up complete? Were efforts made to limit the loss to follow-up? What	

Source: Adapted from 1) Newcastle Ottawa Scale [http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm); 2) Reader's Guide to Critical Appraisal of Cohort Studies. BMJ 2005 (3 article series); 3) Grimes et al. Lancet 2002;359:341-45; and 4) Guyatt & Rennie. Users' Guides to the Medical Literature, AMA Press, 2002.
Compiled by Madhu Pai [madhukar.pai@mcgill.ca] 1

https://www.teachepi.org/teaching-resources/worksheets/

Case studies of bias

Compiled by

Madhukar Pai, MD, PhD

Jay S Kaufman, PhD

Real life case studies of how things went wrong and what we can learn from them!

Available at: https://www.teachepi.org/teaching-resources/bias-case-studies/

Every single epidemiological study will have bias: we can try and reduce the amount & adjust for it in our analyses

Try and prevent + handle it in analysis

- Prevent
  - Pick a stronger study design
  - Randomization, blinding, etc
  - Random sampling to reduce bias during selection
  - Collect information with accurate tools
  - Matching to reduce confounding
- Handle during analysis
  - Adjust for confounding
  - Adjust for selection probabilities
  - Adjust for imperfect tools
  - Adjust for misclassification, etc