Causal inference and accountability research:
evidence synthesis, triangulation, and algorithms

Neil Pearce
Department of Medical Statistics
London School of Hygiene and
Tropical medicine

Evidence synthesis, triangulation and algorithms

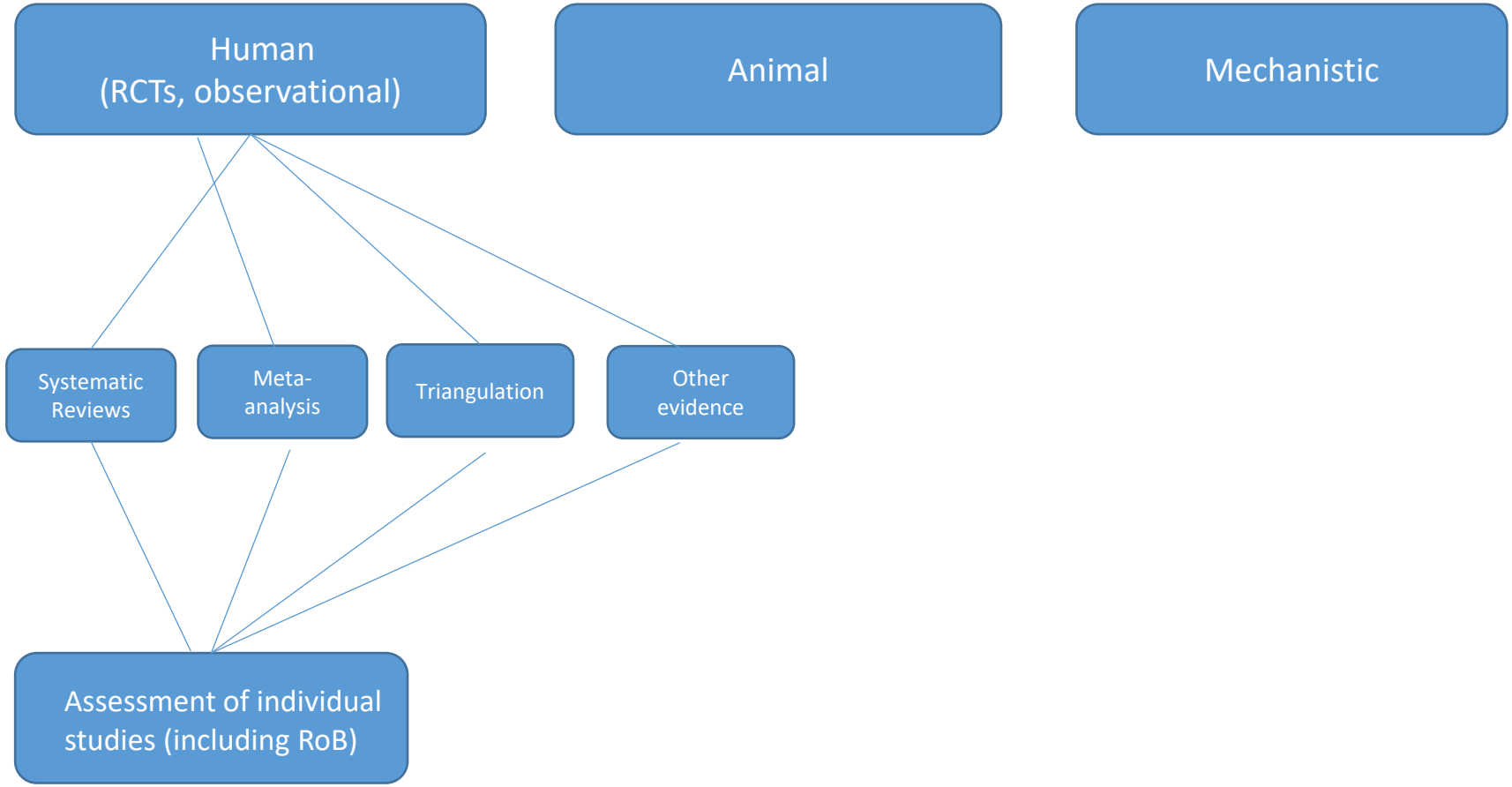
Evidence synthesis and triangulation

Three examples

Use and misuse of algorithms

What should we do?

EVIDENCE SYNTHESIS
[Bradford Hill, IARC, etc]



Bradford Hill's viewpoints

Nine "viewpoints" for assessing causality

1. Strength
2. Consistency
3. Specificity
4. Temporality
5. Biological gradient
6. Plausibility
7. Coherence
8. Experiment
9. Analogy

"What I do not believe—and this has been suggested—is that we can usefully lay down some hard-and-fast rules of evidence that *must* be obeyed before we can accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non. What they can do, with greater or less strength, is to help us to make up our minds on the fundamental question—is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?"

Developments from Bradford-Hill

- Ruling out alternatives
  - E.g. the time trends in lung cancer in the 20th century rule out the possibility that there is a gene which both causes smoking and (independently) causes lung cancer
- Interlocking arguments from different areas of science
  - E.g. the IARC Monographs integrate information from animal, human and mechanistic studies
- Triangulation
  - E.g. negative controls, estimating effects in different populations with different confounding structures

Evidence synthesis, triangulation and algorithms

Evidence synthesis and triangulation

Three examples

Use and misuse of algorithms

What should we do?

Example 1: lung cancer time trends

- The fact that lung cancer rates increased dramatically mid-20th century effectively refutes the hypothesis that the association between smoking and lung cancer was due to unknown genes which caused smoking and also caused lung cancer – if this were the case, then lung cancer rates would have remained stable over time, since genes are relatively stable over time

Example 2: different control groups in case-control studies of NHL and pesticide exposure

General population

- Represents source population
- May be more prone to recall bias if cases are more likely to recall exposures
- Difficult to keep interviewer blind, and may get interviewer bias

"Other cancers"

- Other diseases may be caused by exposure (selection bias)
- Equal motivation and recall in cases and controls
- Easier to keep interviewer blind

NB: a general population control group is expected to yield an OR which is 'too high' whereas an 'other cancer' control group is expected to yield an OR which is 'too low'

Example 2: different control groups

Table 5 *Estimates of odds ratio for non-Hodgkin's lymphoma for various categories of exposure to phenoxyherbicides**

Exposure	*Exposed cases*	*Other cancer*				*General population*			
		No Exp	*OR*	*90% limits*	*p-Value*	*No Exp*	*OR*	*90% limits*	*p-Value*
Ever sprayed an agricultural chemical	51	92	1·3	0·8–2·1	0·22	126	1·6	0·8–3·1	0·18
Ever potentially exposed before cancer registration	21	36	1·3	0·7–2·2	0·30	62	1·1	0·6–2·2	0·44
Potentially exposed for more than 1 day not in the 5 years before cancer registration	17	27	1·3	0·7–2·3	0·29	41	1·0	0·5–2·0	0·52
Probably or definitely exposed for more than 1 day not in the 5 years before cancer registration	16	24	1·3	0·7–2·5	0·28	40	1·0	0·5–2·1	0·49
Probably or definitely exposed for at least 5 days not in the 10 years before cancer registration	14	21	1·3	0·7–2·4	0·33	25	1·3	0·6–2·8	0·40

Source: Pearce et al (1986)

Example 3: confounding in occupational studies

Comparison of crude and smoking-adjusted relative risks

Reference	Exposure	Crude RR	Smoking-adjusted RR
Blot et al., 1978 (20)	Shipbuilding	1.5	1.6
Blot et al., 1980 (21)	Shipbuilding	1.5	1.7
Blot et al., 1982 (27)	Shipbuilding	1.5	1.4
Blot et al., 1983 (22)	Steel industry	2.2	1.9
Breslow et al., 1954 (25)	Welders	7.2	7.7
Buiatti et al., 1985 (26)	Stone, clay and glass production	1.8	1.8
	Welding	3.3	2.8
Damber and Larson, 1985 (23)	Mining 20+ years	5.1	8.9
Hinds et al., 1985 (12)	Asbestos, high exposure	15.5	12.6
Kjuus et al., 1986 (19)	Definite exposure	2.8	2.3
Kvale et al., 1986 (4)	”Lung carcinogens”	2.1	1.7
Martischning et al. 1977 (14)	Asbestos	2.4	2.4
Pastorino et al., 1984 (16)	“Posible” lung carcinogens	2.1	2.1
Riboli et al., 1983 (17)	Phthalate factory	4.1	5.6
	Other exposures	2.0	1.7
Rothschild and Mulvey, 1982 (24)	Sugar-cane farming	2.3	2.5

Evidence synthesis, triangulation and algorithms

Evidence synthesis and triangulation

Three examples

Use and misuse of algorithms

What should we do?

The false hierarchy of study designs

Premier league	RCTs
Championship (but desperate for promotion!)	'Causal inference' Mendelian Randomization
Division 1	Cohort studies
Division 2	Case-control studies
Vauxhall Conference	Cross-sectional studies
Amateurs	Ecologic studies, case series

The false hierarchy of study designs

- None of these designs provides definitive evidence on its own; all depend on ‘auxiliary information’
- Causal inference always involves putting all of the evidence together
- ‘On the average’ a premier league team will defeat a ‘lower league’ team, but there are important exceptions
- Many questions cannot be tested in RCTs (or in observational studies that look like RCTs)
- Other study designs can provide key information

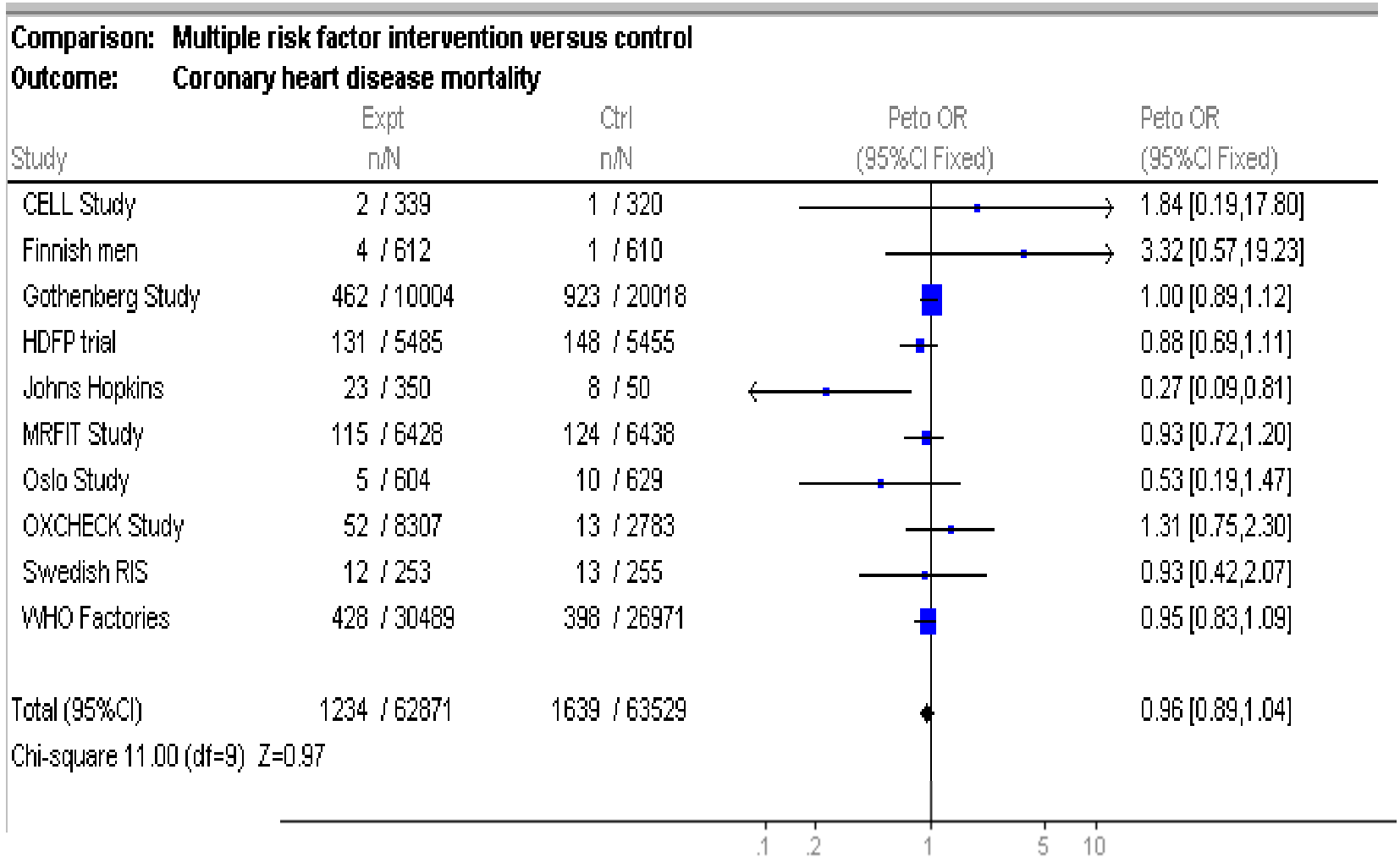
No RCT can prove causality on its own

- RCTs are rarely used in ‘hard sciences’ such as physics; many studies are observational, and those that are trials usually involved controlled rather than randomized experiments
- NO RCT can be interpreted without auxiliary information

Limitations of randomized trials

- Only certain questions can be asked
  - States cannot be studied directly (e.g. sex, obesity)
  - Many actions (e.g. climate change, smoking) are impossible or difficult to randomize
- Only simplistic questions get asked
  - Most interventions on poverty do not address structural inequalities
- The intervention doesn't work
  - Many interventions are not applied properly and/or there is 'spillover' to the non-intervention group
- The intervention works but it is not clear why
  - Often it is not clear which component of the intervention accounts for its success

Systematic review of multiple risk factor interventions: effect on CHD mortality



Ebrahim and Davey Smith BMJ 1997 and Cochrane Library

When RCTs go bad: the MRFIT trial

- RCT for the prevention of coronary heart disease including an extensive smoking cessation programme
- Largest, ‘best’ and most expensive study of its type
- Small increased risk of lung cancer in the smoking cessation intervention group
- Explained as being different from the large body of observational study evidence ‘due to chance’
- Major problem of ‘spillover’ of the intervention
- When the randomization is dropped, then similar estimates of smoking cessation benefit to those of observational studies

Curr Epidemiol Rep (2015) 2:263–270
DOI 10.1007/s40471-015-0054-4

ENVIRONMENTAL EPIDEMIOLOGY (G WELLENIUS, SECTION EDITOR)

Are Randomized Trials Necessary to Advance Epidemiologic Research on Household Air Pollution?

Jennifer L. Peel[1] · **Jill Baumgartner**[2] · **Gregory A. Wellenius**[3] · **Maggie L. Clark**[1] · **Kirk R. Smith**[4]

Published online: 28 September 2015

Abstract Nearly three billion people burn solid fuels in inefficient stoves for cooking and space heating. The resulting household air pollution is the third leading risk factor for mor-

Exposure assessment · Indoor air pollution · Biomass · Cookstoves · Solid fuel · Coal

Example: RCTs for indoor air pollution

- Policy makers and funding agencies often call for more randomized trials of interventions to reduce household air pollution, randomized
- Trials for household air pollution are not feasible for certain health endpoints, may not provide the information that is needed for advancing policy, and may even lead to improper causal inference.
- A variety of study designs, both observational and randomized, may be useful if they include quantitative exposure measurements and appropriately track and measure stove use and other important confounders over time.

[Peel et al, 2015]

Limitations of RCTs for indoor air pollution

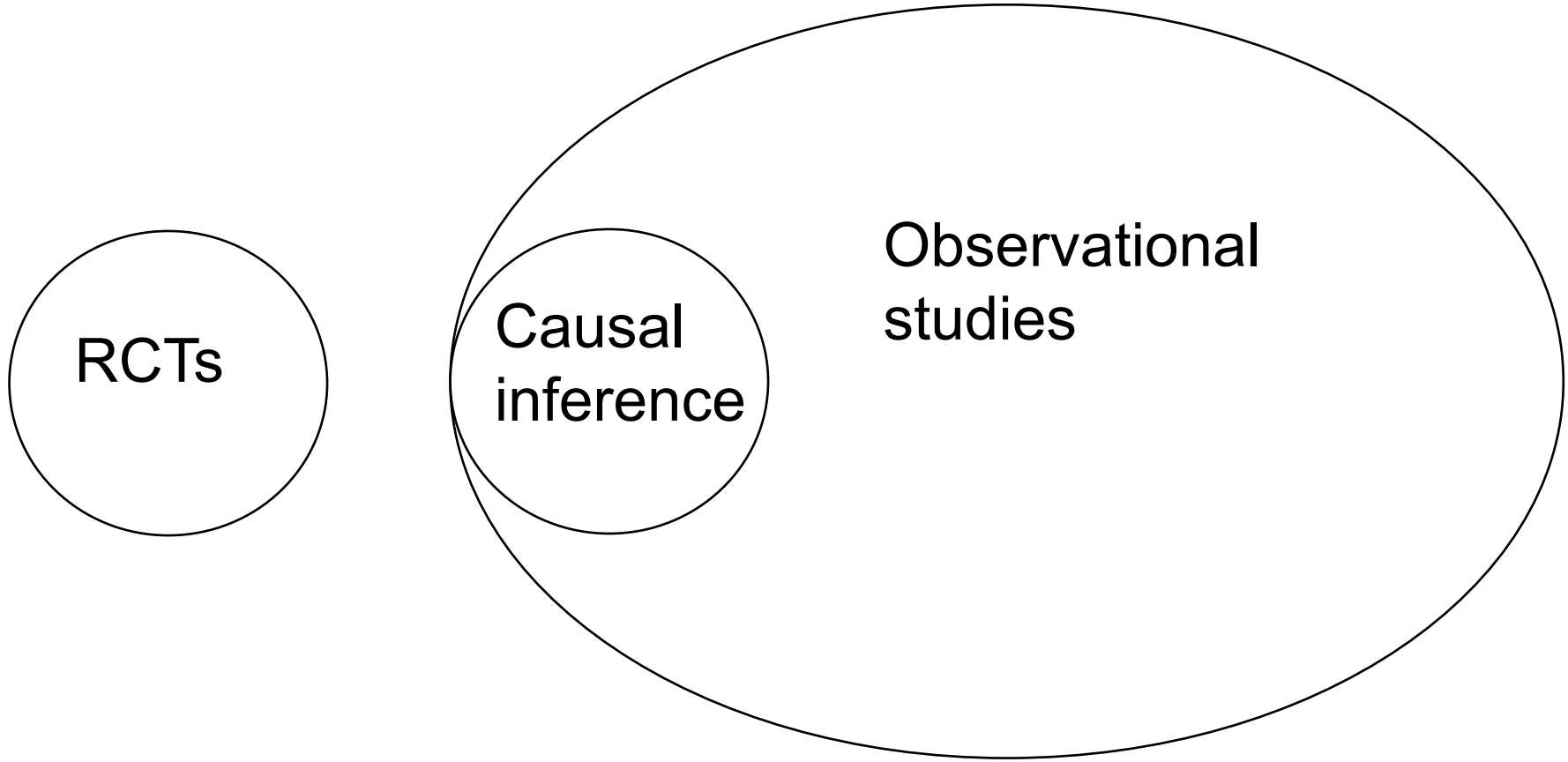
- Not suited for long latency or rare diseases
- Biomarkers may show early effects but scientific and policy relevance is unclear
- Blinding of participants is impossible
- Blinding of investigators may be difficult
- Low adoption of intervention
- Spillover effects
- Ethical considerations
- Assess ‘ideal conditions’ not ‘real world’
- Generalizability over place and time

[Peel et al, 2015]

Old debates in new bottles

Old version: RCTs versus observational studies

New version: 'Causal inference' versus other types of observational studies



The revenge of the algorithms

A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE

Rebecca L. Morgan[a], Kristina A. Thayer[b], Nancy Santesso[a], Alison C. Holloway[c], Robyn Blain[d], Sorina E. Eftim[d], Alexandra E. Goldstone[d], Pam Ross[d], Mohammed Ansari[e], Elie A Akl[a,f], Tommaso Filippini[g], Anna Hansell[h,i,j], Joerg J. Meerpohl[k], Reem A. Mustafa[a,l], Jos Verbeek[m], Marco Vinceti[g,n], Paul Whaley[o], Holger J. Schünemann[a,p,*], GRADE Working Group

[a] Department of Health Research Methods, Evidence, and Impact, McMaster University, Health Sciences Centre, Room 2C14, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada
[b] Integrated Risk Information System (IRIS) Division, National Center for Environmental Assessment (NCEA), Office of Research and Development, US Environmental Protection Agency, Building B (Room 2111), Research Triangle Park, NC 27711, USA
[c] Department of Obstetrics and Gynecology, McMaster University, Health Sciences Centre, Room 3N52A, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada
[d] ICF International Inc., 9300 Lee Highway, Fairfax, VA, USA
[e] School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, ON K1H 8M5, Canada
[f] Department of Internal Medicine, Faculty of Health Sciences, American University of Beirut, P.O. Box: 11-0236, Riad-El-Solh Beirut 1107 2020, Lebanon
[g] Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Italy
[h] MRC-PHE Centre for Environment and Health, Imperial College London, St Mary's Campus, Praed St, Paddington, London W2 1PG, UK
[i] Public Health Directorate, Imperial College Healthcare NHS Trust, St Mary's Hospital, Paddington, London, W2 1PG, UK
[j] Centre for Environmental Health and Sustainability, University of Leicester, George Davies Building, University Road, Leicester LE1 7RH, UK
[k] Institute for Evidence in Medicine (for Cochrane Germany Foundation), Medical Center - University of Freiburg, Breisacher Strasse 153, 79110 Freiburg, Ge
[l] Division of Nephrology and Hypertension, Department of Medicine, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160,
[m] Finnish Institute of Occupational Health, Cochrane Work, Neulaniementie 4, 70701 Kuopio, Finland
[n] Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA
[o] Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK
[p] Department of Medicine, McMaster University, Health Sciences Centre, Room 2C14, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada

ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions

Jonathan AC Sterne,[1] Miguel A Hernán,[2] Barnaby C Reeves,[3] Jelena Savović,[1,4] Nancy D Berkman,[5] Meera Viswanathan,[6] David Henry,[7] Douglas G Altman,[8] Mohammed T Ansari,[9] Isabelle Boutron,[10] James R Carpenter,[11] An-Wen Chan,[12] Rachel Churchill,[13] Jonathan J Deeks,[14] Asbjørn Hróbjartsson,[15] Jamie Kirkham,[16] Peter Jüni,[17] Yoon K Loke,[18] Theresa D Pigott,[19] Craig R Ramsay,[20] Deborah Regidor,[21] Hannah R Rothstein,[22] Lakhbir Sandhu,[23] Pasqualina L Santaguida,[24] Holger J Schünemann,[25] Beverly Shea,[26] Ian Shrier,[27] Peter Tugwell,[28] Lucy Turner,[29] Jeffrey C Valentine,[30] Hugh Waddington,[31] Elizabeth Waters,[32] George A Wells,[33] Penny F Whiting,[34] Julian PT Higgins[35]

Non-randomised studies of the effects of interventions are critical to many areas of healthcare evaluation, but their results may be biased. It is therefore important to understand and appraise their strengths and weaknesses. We developed ROBINS-I

such as cohort studies and case-control studies in which intervention groups are allocated during the course of usual treatment decisions, and quasi-randomised studies in which the method of allocation falls short of full randomisation. Non-randomised studies can provide evidence additional to that available from randomised trials about long term outcomes, rare events, adverse effects and populations that are typical of real world practice.[1 2] The availabil-

Why is it a problem?

- In many ways this is a re-run of the old debate about RCTs versus observational studies ('old wine in new bottles')
- We used to be told that we could only establish causality with RCTs; now we are being told that we can only establish causality with RCTs, or with observational studies which closely mimic RCTs
- Exclusion and/or neglect of other causes which are states, even though many of these are important clinically (e.g. dynamic states such as obesity, hypercholesterolemia, high blood pressure) or socioeconomically (e.g. 'fixed' states such as sex)
- Suggests that causality can be established with a single 'perfect' study, and that we should always strive to do such a study rather than taking a more comprehensive (triangulation) approach
- Neglect or elimination of many of the things that make epidemiology unique and important (the population perspective)

How Monsanto manipulates journalists and academics

Carey Gillam

Monsanto's own emails and documents reveal a disinformation campaign to hide its weedkiller's possible links to cancer

▲ Monsanto's weedkiller Roundup, one of the world's most popular herbicides, may cause cancer. Photograph: Mike Blake/Reuters

Over the past year, evidence of Monsanto's deceptive efforts to defend the safety of its top-selling Roundup herbicide have been laid bare for all to see. Through three civil trials, the public release of internal corporate communications has revealed conduct that all three juries have found so unethical as to warrant punishing punitive damage awards.

Much attention has been paid to Monsanto conversations in which company scientists casually discuss ghostwriting scientific papers and suppressing science that conflicts with corporate assertions of Roundup's safety. There has also been public outrage over internal records illustrating cozy relationships with friendly regulators which border on - and possibly cross into - collusion.

Uproar after research claims red meat poses no health risk

One expert says findings by international experts represent 'egregious abuse of evidence'

▲ Critics of the study say many of the participants were young and unlikely to succumb to illness during the trial period. Photograph: Getty

New research that claims red and processed meat is probably not harmful to our health has caused controversy among experts who maintain people should cut down.

The World Health Organization has classified red and processed meats as

Scott Pruitt is administrator of the Environmental Protection Agency. GAGE SKIDMORE/FLICKR (**CC BY-SA 2.0**)

EPA unveils new industry-friendlier science advisory boards

By Sean Reilly, E&E News, Kevin Bogardus, E&E News | Nov. 3, 2017 , 1:30 PM

Originally published by E&E News

The U.S. Environmental Protection Agency (EPA) officially unveiled new membership rosters today for several key science advisory panels that give more weight to representatives of industry and state governments at the expense of university researchers.

The New York Times

Opinion

Don't Let a Killer Pollutant Loose

The Trump administration is moving to ease standards on a particularly deadly air contaminant.

By John Balmes
Dr. Balmes is a medical professor and member of the California Air Resources Board.

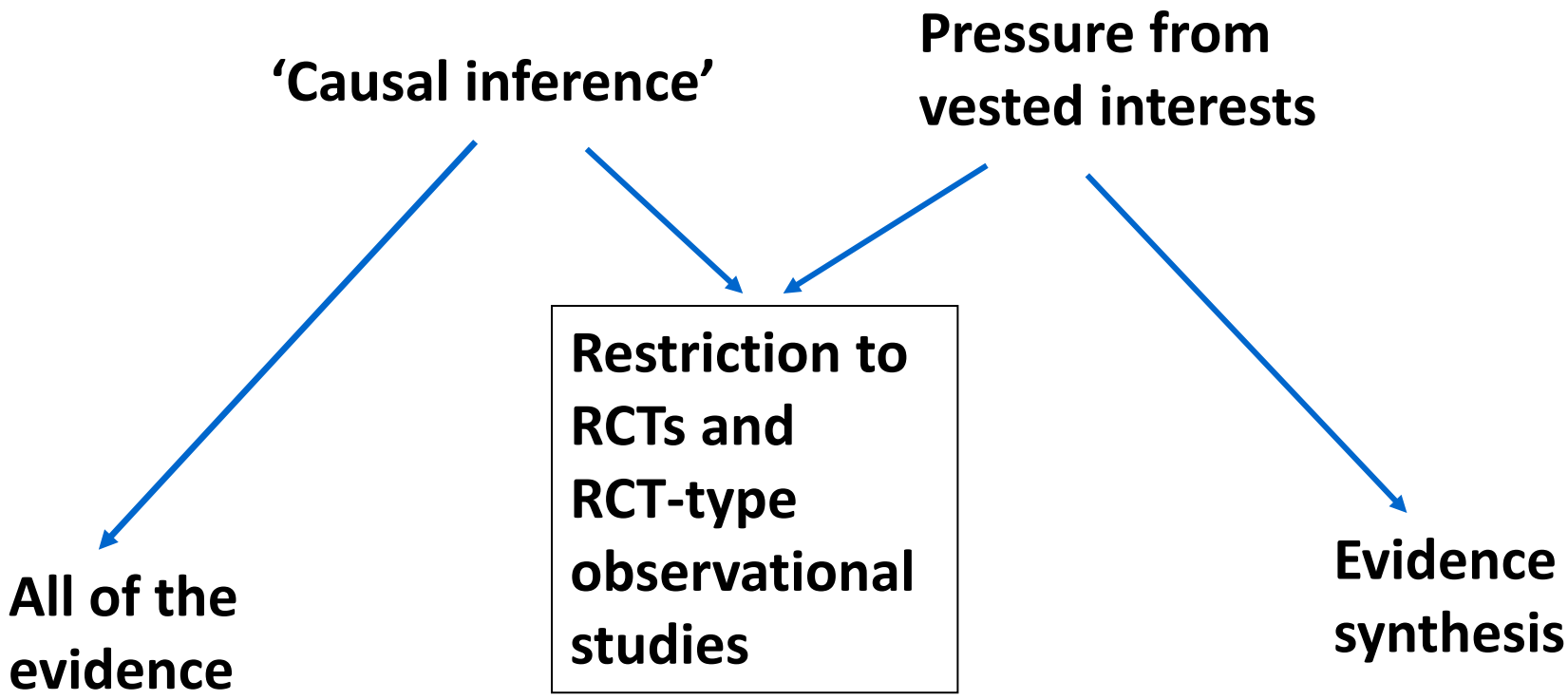
April 14, 2019

Kim Ryu

PM 2.5 kills people. There has been little dispute that microscopic particulate matter in air pollution penetrates into the deepest parts

"Rather than relying on the weight-of-evidence approach that the EPA has traditionally used to infer causation, [the Clean Air Science Advisory Committee] wants to rely on studies that use… 'manipulative causality'. This theory restricts epidemiologic evidence that may be considered acceptable to assess causality to results from intervention studies or studies that have been analysed with the use of causal inference statistical methods". [NEJM 2019; 381:8]

The ‘back door’ selection bias in evidence synthesis



Evidence synthesis, triangulation and algorithms

Evidence synthesis and triangulation

Three examples

Use and misuse of algorithms

What should we do?

Solutions 1: triangulation

International Journal of Epidemiology, 2017, 1–21
doi: 10.1093/ije/dyw314
Original Article

Original Article

Triangulation in aetiological epidemiology

Debbie A. Lawlor,[1,2,*] Kate Tilling[1,2] and George Davey Smith[1,2]

[1]MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK and [2]School of Social and Community Medicine, University of Bristol, Bristol, UK

*Corresponding author. MRC IEU, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. E-mail: d.a.lawlor@bristol.ac.uk

Accepted 3 October 2016

Abstract

Triangulation is the practice of obtaining more reliable answers to research questions through integrating results from several different approaches, where each approach has different key sources of potential bias that are unrelated to each other. With respect to

Triangulation is the practice of obtaining more reliable answers to research questions through integrating results from several different approaches, where each approach has different key sources of potential bias that are unrelated to each other… We emphasize the importance of being explicit about the expected direction of bias within each approach, whenever this is possible, and seeking to identify approaches that would be expected to bias the true causal effect in different directions.

Solutions 2: assess specific sources of bias – don't use algorithms

American Journal of Epidemiology

Vol. 188, No. 9
DOI: 10.1093/aje/kwz131
Advance Access publication:
May 30, 2019

Special Article

The Problem With Mechanistic Risk of Bias Assessments in Evidence Synthesis of Observational Studies and a Practical Alternative: Assessing the Impact of Specific Sources of Potential Bias

David A. Savitz*, Gregory A. Wellenius, and Thomas A. Trikalinos

* Correspondence to David A. Savitz, Department of Epidemiology, Brown University School of Public Health, 121 South Main Street, Box G-S121, Providence, RI 02912 (e-mail: david_savitz@brown.edu).

Initially submitted August 4, 2018; accepted for publication May 17, 2019.

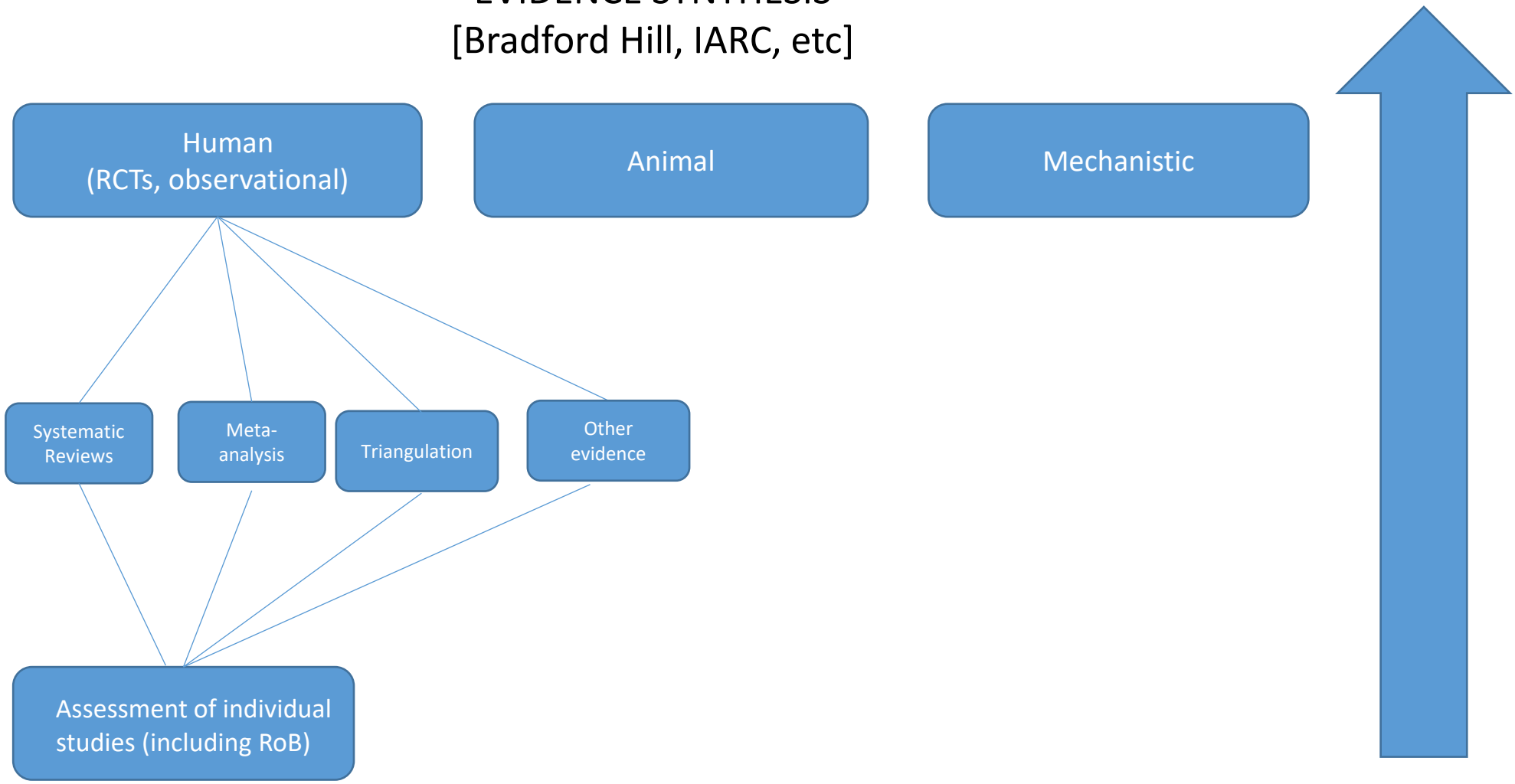
The trustworthiness of individual studies is routinely characterized in systemic reviews by evaluating risk of bias, often by mechanistically applying standardized algorithms. However, such instruments prioritize the repeatability of the process over a more thoughtful and informative but necessarily somewhat more subjective approach. In mechanistic risk of bias assessments, the focus is on determining whether specific biases are present, but these assessments do not provide insights into the direction, magnitude, and relative importance of individual biases. In such assessments, all

Mechanistic risk of bias assessments focus on assessing whether specific biases are present but fail to provide insights into the direction, magnitude, and relative importance of individual biases. Instead, risk of bias assessments should focus on identifying a small number of the most likely influential sources of bias… classifying each specific study based on how effectively it has addressed each potential bias, and determining whether results differ across studies in relation to susceptibility to each hypothesized source of bias.

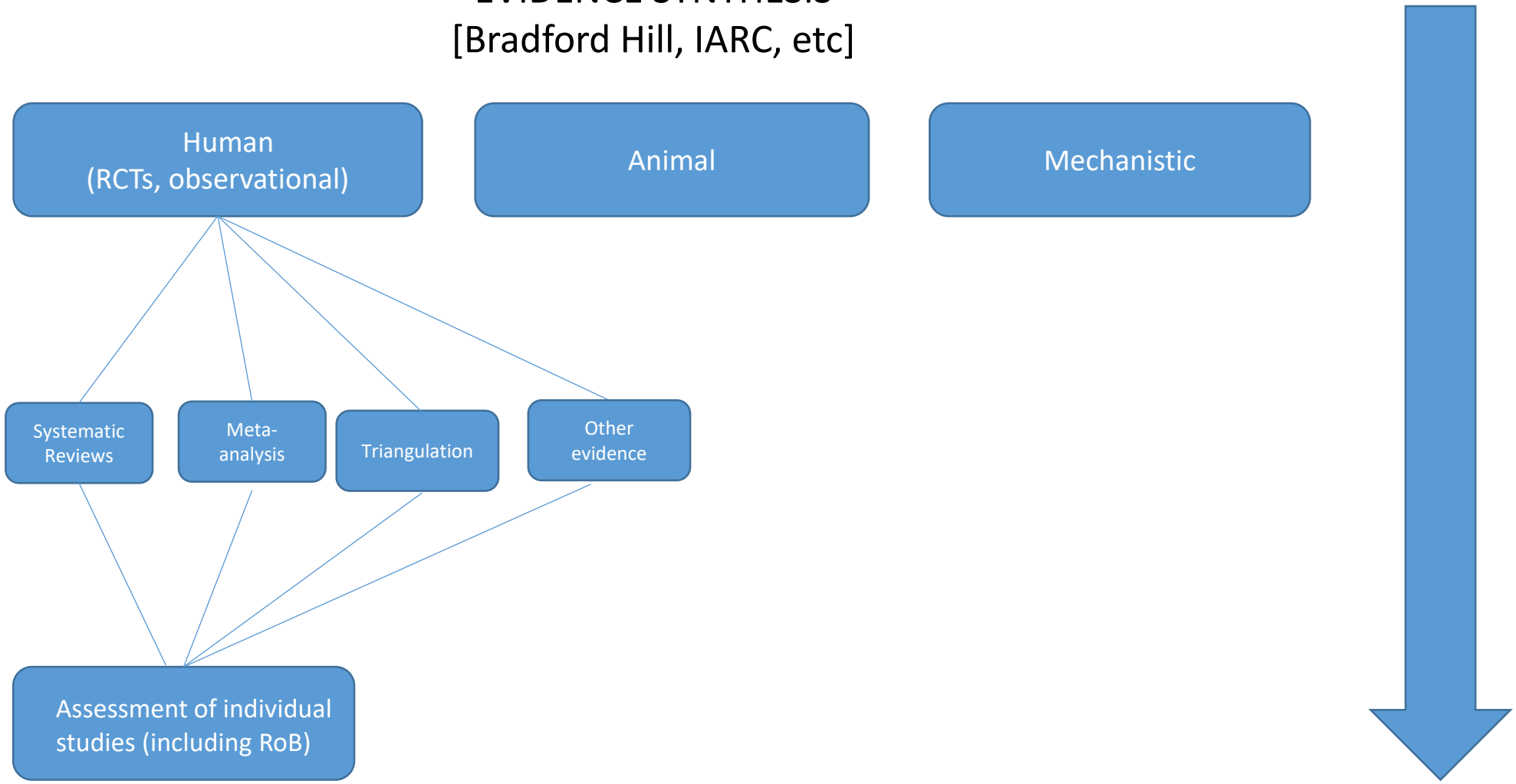
What should we do?

- Risk of bias tools can be useful in providing a ‘checklist’ of possible biases in individual studies, or groups of similar studies
- They only provide part of the information needed for evidence synthesis
- A ‘low scoring’ study may provide crucial evidence in the context of triangulation and/or systematic reviews
- Risk of bias assessment cannot be done in the abstract, but requires knowledge of the context, and the broader evidence that is to be synthesized
- Risk of bias assessments should not be used to reject ‘low scoring’ studies which may still provide useful information for evidence synthesis

EVIDENCE SYNTHESIS
[Bradford Hill, IARC, etc]



EVIDENCE SYNTHESIS
[Bradford Hill, IARC, etc]



Causal inference and accountability research:
evidence synthesis, triangulation, and algorithms

Neil Pearce
Department of Medical Statistics
London School of Hygiene and
Tropical medicine