

NAM: Updating Standards for Systematic Reviews Perspectives from *Annals of Internal Medicine*

NAM February 2026

Christine Laine MD, MPH
Editor in Chief, *Annals of Internal Medicine*

Annals of Internal Medicine®



Disclosures

Christine Laine, MD, MPH

Editor in Chief, *Annals of Internal Medicine*

Senior VP, American College of Physicians

Full time employee of the American College of Physicians,
the owner and publisher of *Annals of Internal Medicine*

About *Annals of Internal Medicine*

- Systematic Reviews are a priority area for publication
- Dedicated Deputy Editor for systematic reviews
- Among the first clinical journals to publish systematic reviews
- Considered among the top clinical general medical journals
- >20 million page views and >2.5 billion media hits annually
- Highly selective, <4% submissions accepted in 2025

Systematic Reviews: Common Problems to Consider When Updating Standards



Background and Introduction: Common Problems

- Poorly articulated question(s)
 - What is motivating the systematic review?
 - Why is it importance
- Lack of justification for doing a systematic review
 - What is current practice?
 - Is the question already answered? If so, is there substantial new evidence that might change that answer?
 - If prior reviews, is this systematic review likely to contribute meaningful new info? a



Methods: Common Problems

- Lack of prospective registration in Prospero
- Search date is not current
- Failure to follow methods standards (IOM, GIN), reporting guidelines (PRISMA) or evidence rating system (GRADE, Cochrane Risk of Bias)... often even when stating these were used
- Using “modified methods” without specifying exactly how modified

“We evaluated risk of bias using a modified Cochrane Risk of Bias tool.”

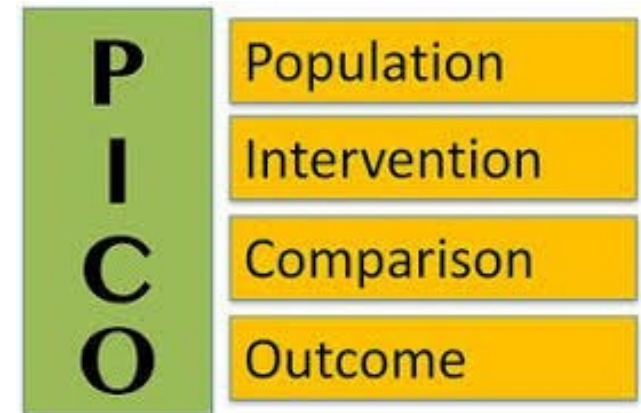


Does this really mean?

” We know we should have used the Cochrane Risk of Bias tool but we made a lot of shortcuts.”

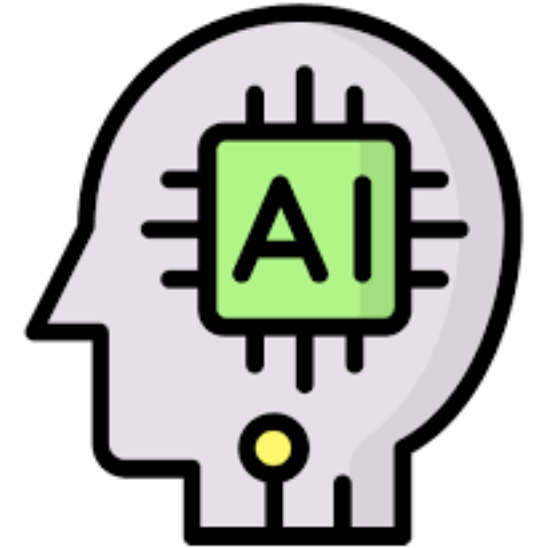
Methods: Common Problems

- PICO inclusion/exclusion PICO criteria poorly described
- Ensure PICO criteria reflect the clinical question
 - Population - Are controversies in a particular subgroup?
 - Intervention(s) - How are interventions defined?
 - Comparator - Is it a question of comparative effectiveness?
What is the standard comparison?
 - Outcomes – Consider both benefits and harms, time course



Methods: Common Problems

- Failure to report that artificial intelligence tools were used
- IF AI used to assist review process, need to provide detail of how they were used, whether humans checked AI contributions



Artificial Intelligence

“At submission, *Annals* requires authors to attest whether they used Artificial Intelligence (AI)-assisted technologies (such as Large Language Models (LLMs), chatbots or image creators) in the production of submitted work. Authors who use such technology should describe, in both the cover letter and the submitted work, how they used it. Chatbots (such as ChatGPT) should not be listed as authors because they cannot be responsible for the accuracy, integrity, and originality of the work, and these responsibilities are required for authorship <https://www.icmje.org/recommendations/>. Therefore, human authors are responsible for any submitted material that included the use of AI-assisted technologies.”

Analysis: Common Problems

- Just because data can be meta-analyzed doesn't mean that it should be meta-analyzed
- If evidence is heterogeneous, a qualitative synthesis is likely more useful than a quantitative synthesis
- Failure to prospectively specify minimally important differences

Presenting Results: Avoiding Common Problems

- Don't just summarize individual studies, clearly convey the summary conclusion and strength of evidence
- Describe the body of evidence by relevant subgroups
- Describe how the evidence supports the conclusion
- If evidence is “insufficient” or “very low” certainty make that clear—simply describing effect estimates can give the impression of evidence that is stronger than it is
- Report results in both absolute and relative terms
- Well conceived and organized tables and figures

Over-reliance on Statistical Significance

- Statistical significance/imprecision is only part of assessing the certainty of evidence
- Need to Consider minimally important difference and decision thresholds for interpreting estimates and confidence intervals
- Need to convey overall conclusions with certainty of evidence

Example:

Mean difference -9.7, 95% CI -20.4 to 1.1; low certainty)



“low certainty of a trivial to no difference”

Mean difference -9.1, 95% CI -17.2 to -1.1; moderate certainty)



“bisphosphonates probably reduce pain intensity compared with placebo”

Tables and Figures: Avoiding Common Problems

- Table and figures should enable the reader to understand the evidence supporting summary conclusions
- Organize text, tables, and figures in similar groupings and order
- Summary Results table should
 - Report ALL primary outcomes
 - Report absolute AND relative effects
 - Describe evidence supporting conclusions
 - Include clear conclusion considering effect estimates, CI, certainty of evidence

Outcomes	Anticipated absolute effects* (95% CI)			No of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with absence of vaccination	Risk with COVID-19 vaccination	Relative effect (95% CI)			
Incidence of SARS-CoV-2 infection	46 per 1.000	22 per 1.000 (13 to 36)	OR 0.46 (0.28 to 0.76)	142527 (8 observational studies)	⊕○○○ Very low ^{a,b,c}	The evidence is very uncertain about the effect of COVID-19 vaccination on the incidence of SARS-CoV-2 infection.
COVID-19 related hospitalisation	9 per 1.000	4 per 1.000 (3 to 4)	OR 0.41 (0.33 to 0.51)	129911 (4 observational studies)	⊕○○○ Very low ^{a,c}	The evidence is very uncertain about the effect of COVID-19 vaccination on COVID-19 related hospitalisation.
COVID-19 related admission in ICU	Absolute AND Relative effects			Description of evidence supporting conclusion		The evidence is very uncertain about the effect of COVID-19 vaccination on COVID-19 related admission in ICU.
Incidence of SARS-CoV-2 infection in fully vaccinated	52 per 1.000	17 per 1.000 (9 to 31)	OR 0.31 (0.16 to 0.59)	104091 (4 observational studies)	⊕○○○ Very low ^{a,c,d}	The evidence is very uncertain about the effect of COVID-19 vaccination on incidence of SARS-CoV-2 infection in fully vaccinated.
COVID-19 related hospitalisation in fully vaccinated	10 per 1.000	2 per 1.000 (1 to 2)	OR 0.15 (0.10 to 0.21)	93129 (2 observational studies)	⊕○○○ Very low ^{a,c}	The evidence is very uncertain about the effect of COVID-19 vaccination on COVID-19 related hospitalisation in fully vaccinated.
COVID-19 related admission in ICU in fully vaccinated		1 per 1.000	OR 0.53	105223 (4 observational studies)	⊕○○○ Very low ^{a,c,d,e}	Clear conclusion considering effect estimates, CI, certainty of evidence
Summary of all primary outcomes – benefits and harms						

Study, Year (Reference)	Events/Participants, n/N		Mean Age, y	Mean Baseline BMI, kg/m ²	PCOS, %	Risk of Bias	Mean Difference in Weight Change, kg
	Intervention	Comparator					

Diet and/or physical activity vs. no/minimal intervention

Becker et al, 2015 (27)	3/16	0/19	31.3	28.7	15.4	High	-5.23
Moran et al, 2011 (26)	0/21	1/25	33.1	34	0	Some concerns	-3.30
Wang et al, 2023 (30)	2/18	0/20	30	27.8	0	High	-3.20
Subgroup, DL+HKSJ (<i>I</i> ² = 0.0%, <i>P</i> = 0.54)							
Subgroup, MH							

Diet and/or physical activity vs. active comparator

Kiel et al, 2018 (31)	2/8	1/10	32.3	30.2		High	0.10
-----------------------	-----	------	------	------	--	------	------

Low-energy diet with/without physical activity advice vs. active comparator

Price et al, 2020 and 2021 (34, 35)	49/85	29/79	32.3	38.7	39	Some concerns	-9.80
Sim et al, 2014 (33)	3/27	0/22	32.9	36.4	30.6	Some concerns	-5.00
Subgroup, DL+HKSJ (<i>I</i> ² = 0.0%, <i>P</i> = 0.69)							
Subgroup, MH							

Pharmacotherapy with some diet and physical activity advice vs. active comparator

Li et al, 2022 (38)	21/80	11/80	28	29.1	100	High	-1.56
Wang et al, 2021 (37)	11/445	9/443	30.7	29.4		Low	-1.18
Salamun et al, 2018 (36)	3/13	3/14	30.6	36.6	100	Some concerns	-0.52
Subgroup, DL+HKSJ (<i>I</i> ² = 0.0%, <i>P</i> = 0.56)							

Organize tables so that readers can understand differences across studies

Make headings understandable to someone not familiar with the review

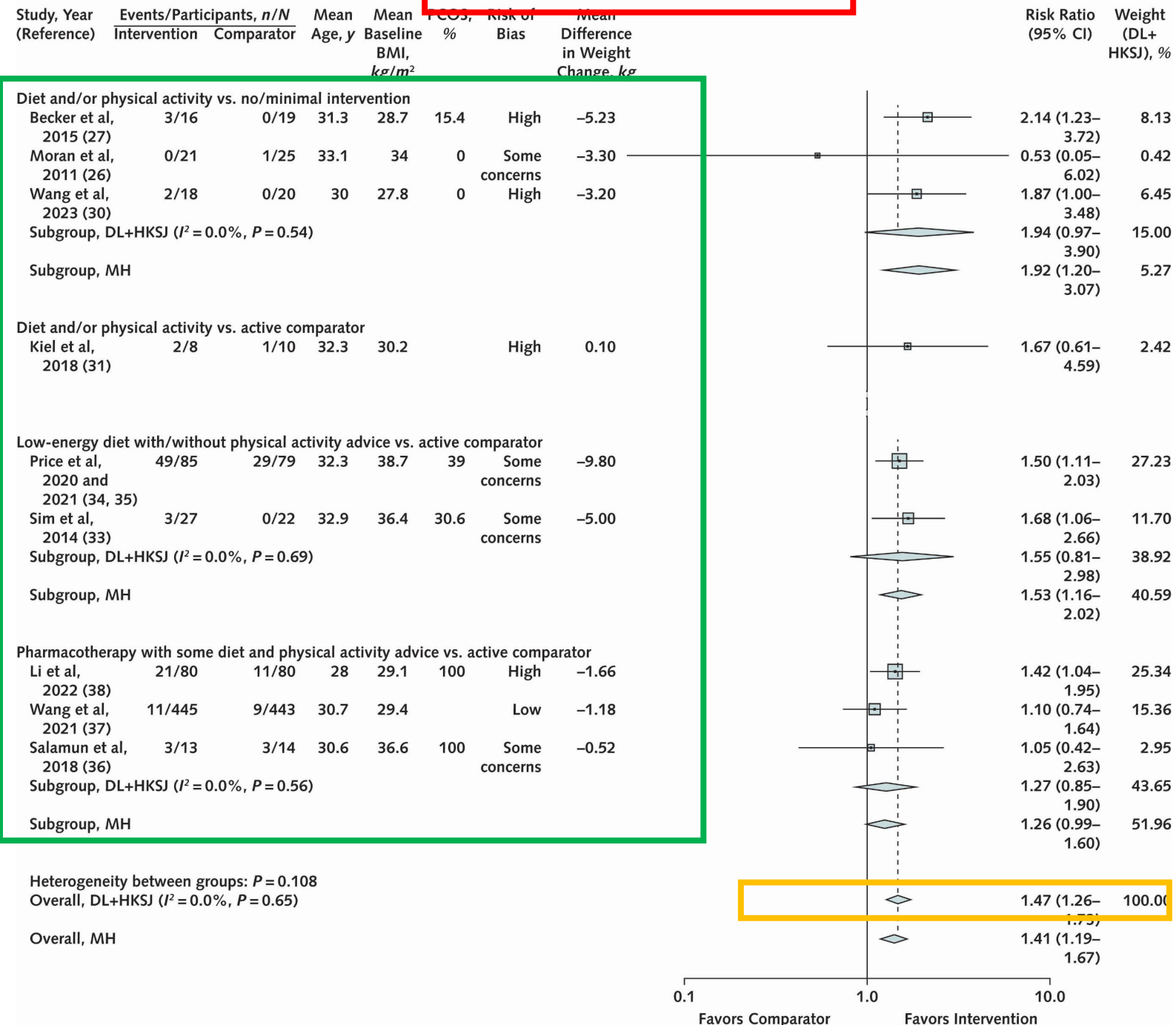
Separate by important subgroups

Provide information on potential effect modifiers

Organize studies within subgroups in a meaningful way

<https://www.acpjournals.org/doi/10.7326/ANNALS-24-01025>

Unassisted Pregnancy



Organize forest plot to support the summary results

Forest plot clearly labelled with outcome that matches the Summary Table

Reader can:

- 1) See how many studies contribute to the effect estimate
- 2) Understand differences across studies

Reader can see where effect estimate from summary table is derived

Conclusions: Common Pitfalls

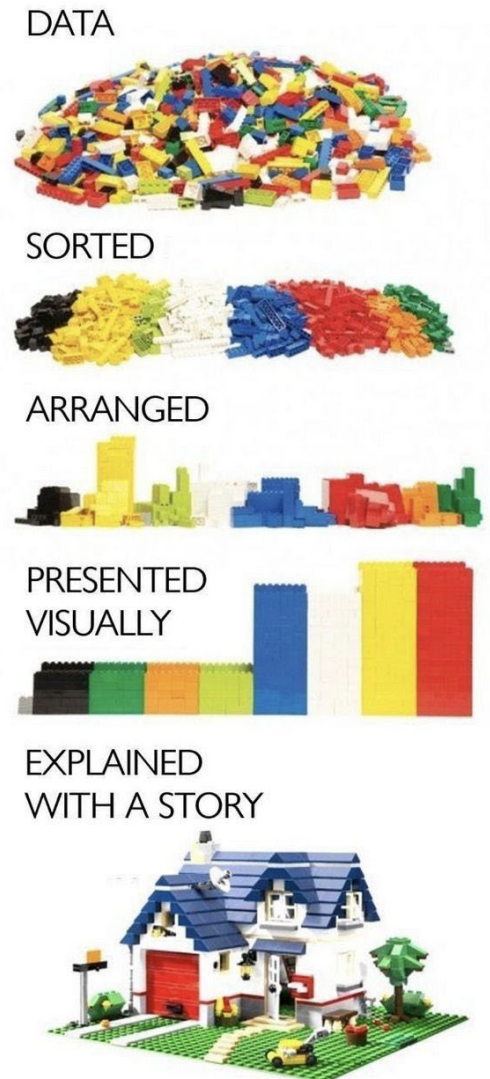
Often reviews give much attention to process and too little attention for the synthesis of findings

Clear conclusions require going beyond the meta-analysis findings – meta-analysis is only the first step in digesting the data

- Over-reliance on statistical significance
- Inadequate consideration of heterogeneity of available evidence and implications for findings
- Presenting conclusions with “spin”
- Inadequate discussion of applicability of findings

Consider Heterogeneity

- Enable the reader to understand differences across studies
 - Define the potential effect modifiers
 - Provide a table that summarizes the differences across studies for major effect modifiers
 - Organize forest plots by relevant effect modifiers
- Avoid “over-combining” studies that differ in important ways



Avoid “Spin”



Annals of Internal Medicine RESEARCH AND REPORTING METHODS

Development and Evaluation of a Framework for Identifying and Addressing Spin for Harms in Systematic Reviews of Interventions

Riaz Qureshi, PhD; Kevin Naaman, MS; Nicolas G. Quan, BSc; Evan Mayo-Wilson, DPhil; Matthew J. Page, PhD; Victoria Cornelius, PhD; Roger Chou, MD; Isabelle Boutron, MD, PhD; Su Golder, PhD; Lisa Bero, PhD; Peter Doshi, PhD; Matt Vassar, PhD; Reint Meursing Reynders, DDS, PhD; and Tianjing Li, MD, PhD

“Spin” refers to misleading reporting, interpretation, and extrapolation of findings in primary and secondary research (such as in systematic reviews). The study of spin primarily focuses on beneficial outcomes. The objectives of this research were threefold: first, to develop a framework for identifying spin associated with harms in systematic reviews of interventions; second, to apply the framework to a set of reviews, thereby pinpointing instances where spin may be present; and finally, to revise the spin examples, offering guidance on how spin can be rectified.

The authors developed their framework through an iterative process that engaged an international group of researchers specializing in spin and reporting bias. The framework comprises 12 specific types of spin for harms, grouped by 7 categories across the 3 domains (reporting, interpretation, and extrapolation). The authors subsequently gathered instances of spin from a random sample of 100 systematic reviews of interventions. Of the 58 reviews that assessed

harm and the 42 that did not, they found that 28 (48%) and 6 (14%), respectively, had at least 1 of the 12 types of spin for harms. Inappropriate extrapolation of the results and conclusions for harms to populations, interventions, outcomes, or settings not

“Spin refers to misleading reporting, interpretation, and extrapolation of findings in primary and secondary research (such as systematic reviews).”

Ann Intern Med. 2024;177:1089-1098. doi:10.7326/M24-0771 **Annals.org**
For author, article, and disclosure information, see end of text.
This article was published at **Annals.org** on 16 July 2024.

Qureshi et al., Development and Evaluation of a Framework for Identifying and Addressing Spin for Harms in Systematic Reviews of Interventions, *Annals of Internal Medicine*, Volume 177, Number 8 <https://doi.org/10.7326/M24-07>

Annals
of Internal Medicine®

Avoiding “Spin”

- Present ALL pre-specified primary outcomes (even if no evidence or no benefit)
- Present BOTH benefits and harms (even if no evidence or no benefit)
- Do not over-interpret findings
 - Use appropriate language to convey certainty of evidence
 - Be consistent in interpretation of estimates
- If presenting secondary or post-hoc outcomes or analyses, make this clear



Discuss Applicability of Findings

Can Salt Substitutes Improve Cardiovascular Outcomes?

Study, Year (Reference) [Duration, <i>mo</i>]	Country (Setting)	Age, <i>y</i>	Male, %	SBP DBP HTN, % (CV Risk)	RR (95% CI)	RR (95% CI)
All-cause mortality						
Chang et al, 2006 (29) [30]	Taiwan (AC/R)	75	100	131 71 40 (H)		0.92 (0.61–1.40)
China Salt Substitute Study Collaborative Group, 2007 (31) [12]	China (C/R)	60	44	158 90 61 (H)		0.99 (0.25–3.95)
Neal et al, 2021 (32) [56]	China (C/R)	65	51	154 89 88 (H)		0.88 (0.82–0.95)
Yuan et al, 2023 (15) [24]	China (AC/R)	71	76	138 81 62 (H)		0.84 (0.63–1.12)
Zhang et al, 2018 (33) [36–48]	China (AC/R)	67	0	0 0 0 (H)		0.85 (0.69–1.05)
Zhou et al, 2016 (34) [36–120]	China (C/R)	46	49	152 90 75 (H)		0.81 (0.46–1.42)
Overall effect						0.88 (0.82–0.93)
Heterogeneity $I^2 = 0\%$, $\tau^2 = 0$, $\gamma^2 = 0.32$ ($P = 1.00$)						

Greenwood et al., *Annals of Internal Medicine*
Volume 177, Number 5
<https://doi.org/10.7326/M23-2626>

All 6 RCTs conducted in China or Taiwan and all provided salt substitute

All 6 RCTs in persons of older ages (average 64 yrs)

3 in residential facilities

3 in adults with increased CV risk (HTN with or without stroke)

Living/Rapid Reviews: Common Pitfalls

- Not identifying review as living/rapid in the title, abstract, and methods
- If “living,” failure to specify ongoing surveillance methods and planned periodicity of updates
- If “rapid,” failure to specify the time from start to finish and what elements were bypassed (if any) to achieve desired rapidity

Summary: Avoiding Common Pitfalls in Systematic Reviews:

- Clearly articulate the question(s), justify importance and why a systematic review is the appropriate methodology
- If saying review used methods and reporting standards, be sure to use those standards, if “modified” specify those modifications
- If the review is “rapid” need to specify time frame from conception to reporting and any shortcuts taken/ If “living” need to specify criteria/periodicity for updating
- Consider heterogeneity, just because you can meta-analyze doesn’t mean you should, don’t rely solely on statistical significance
- Thoughtful tables and figures
- Convey clear conclusions, discuss applicability, avoid “spin,” don’t be reluctant to admit uncertainty,