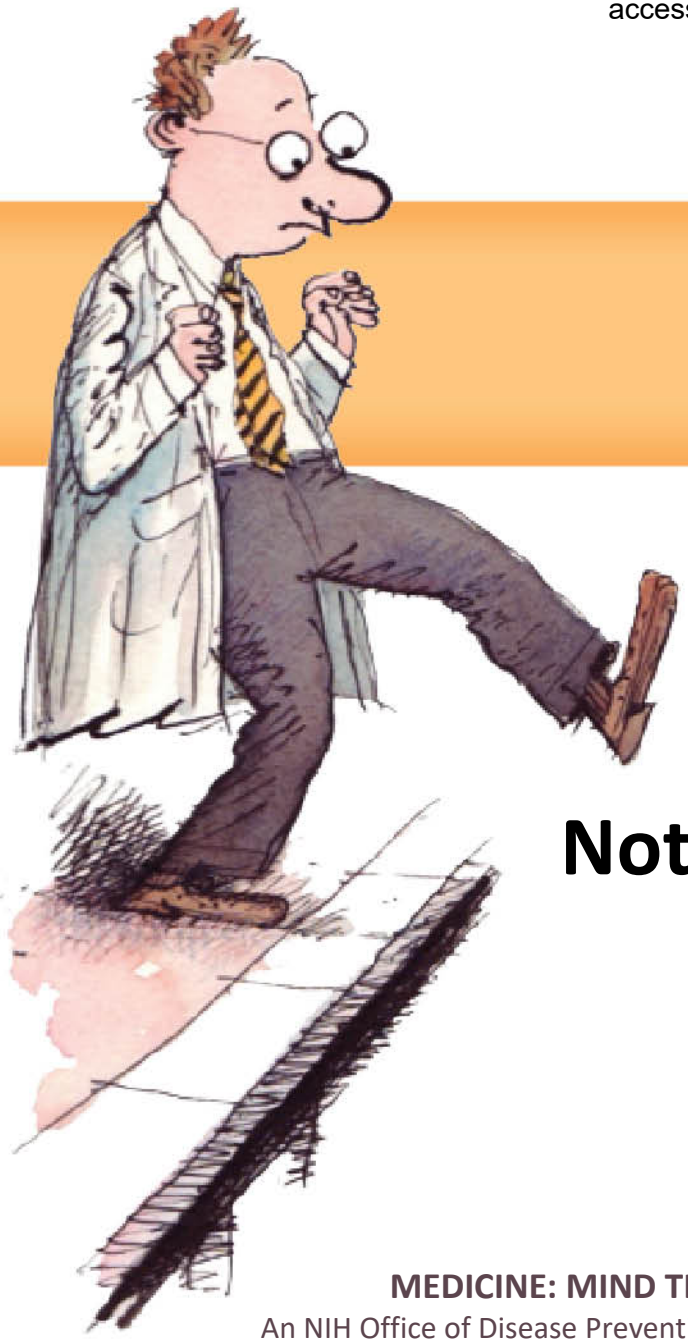
If you're a person with a disability or using assistive technology and are having difficulty accessing any of the content in this file, please contact Marie Rienzo at marie.rienzo@nih.gov.



Mind the Gap
NIH Webinar Series

Sizing Up Systematic Reviews: Not All Syntheses Are Created Equal

Presented by

Jennifer Croswell, M.D., M.P.H.

Patient-Centered Outcomes Research Institute (PCORI)

MEDICINE: MIND THE GAP

An NIH Office of Disease Prevention Webinar Series

Sizing Up Systematic Reviews

NOT ALL SYNTHESES ARE CREATED EQUAL

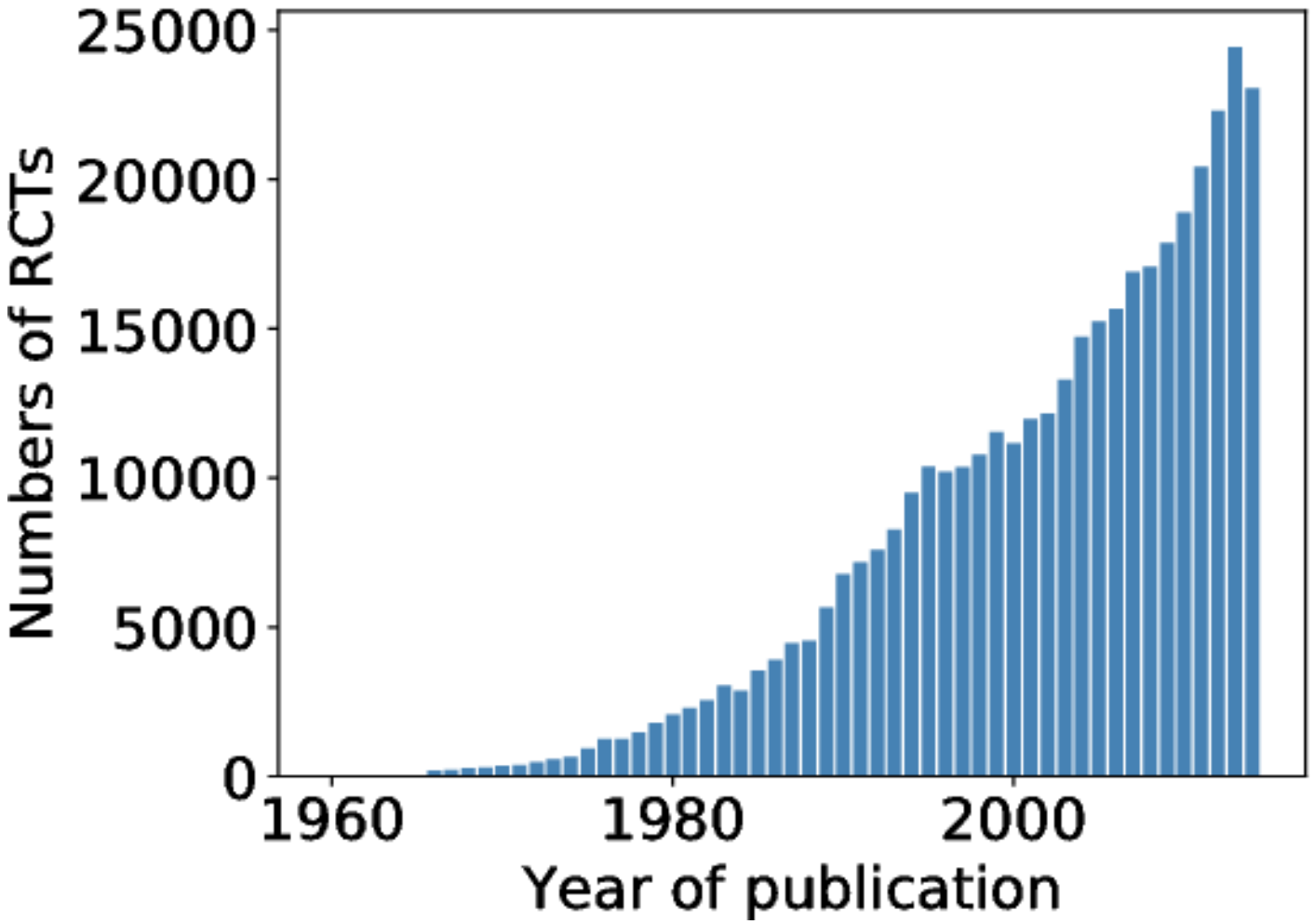
Jennifer Croswell, MD, MPH

Senior Program Officer

Research Synthesis Program

PCORI

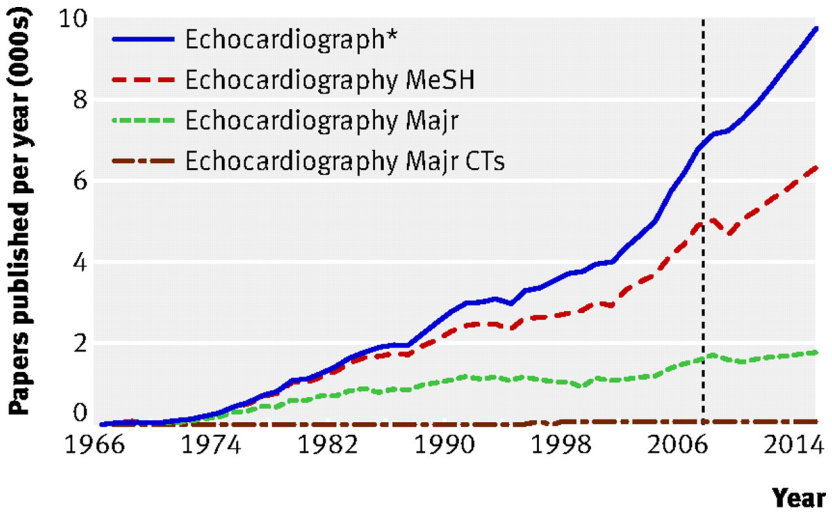
RCTs published annually in PubMed, 1960-2014



Dernoncourt F, Lee JY. PubMed 200k RCT: A dataset for sequential sentence classification in medical abstracts. *Proceedings of the 8th International Joint Conference on Natural Language Processing*. 2017. Taipei, Taiwan: 308-13.

On the impossibility of being expert

- Study estimated the amount of time it would take a new cardiologist to come up to date with the prior literature and then stay abreast of new research specifically related to cardiac diagnostic imaging
- Assuming she read 5 papers an hour for 8 hours a day, 5 days a week, 50 weeks a year, it would take her 11.5 years (during which time another 82,000 papers would have been added)
- To start reading new manuscripts published at the same time she is reading them, it would take her 40 years (and 400,000 articles)



Fraser AG, Dunstan FD. On the impossibility of being expert. *BMJ*. 2010; 341.

Differences between systematic and narrative reviews

Feature	Systematic Review	Narrative Review
Research question(s)	Focused, specific Lends itself to PICOTS	May be broad in scope
Sources and search strategy	Comprehensive Explicit search strategy	Not usually specified May be biased
Study selection	Criterion-based, transparent, uniformly applied	Not usually specified May be biased
Study appraisal	Rigorous critical appraisal according to standard criteria	Variable, not always explicit or uniform
Synthesis	Qualitative and quantitative (may include meta-analysis)	Generally restricted to qualitative

Thoma A, Eaves FF. What is wrong with systematic reviews and meta-analyses: If you want the right answer, ask the right question*! Aesthetic Surgery Journal.* 2016; 36(10): 1198-1201.

Systematic review: Definition

A SR attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made.

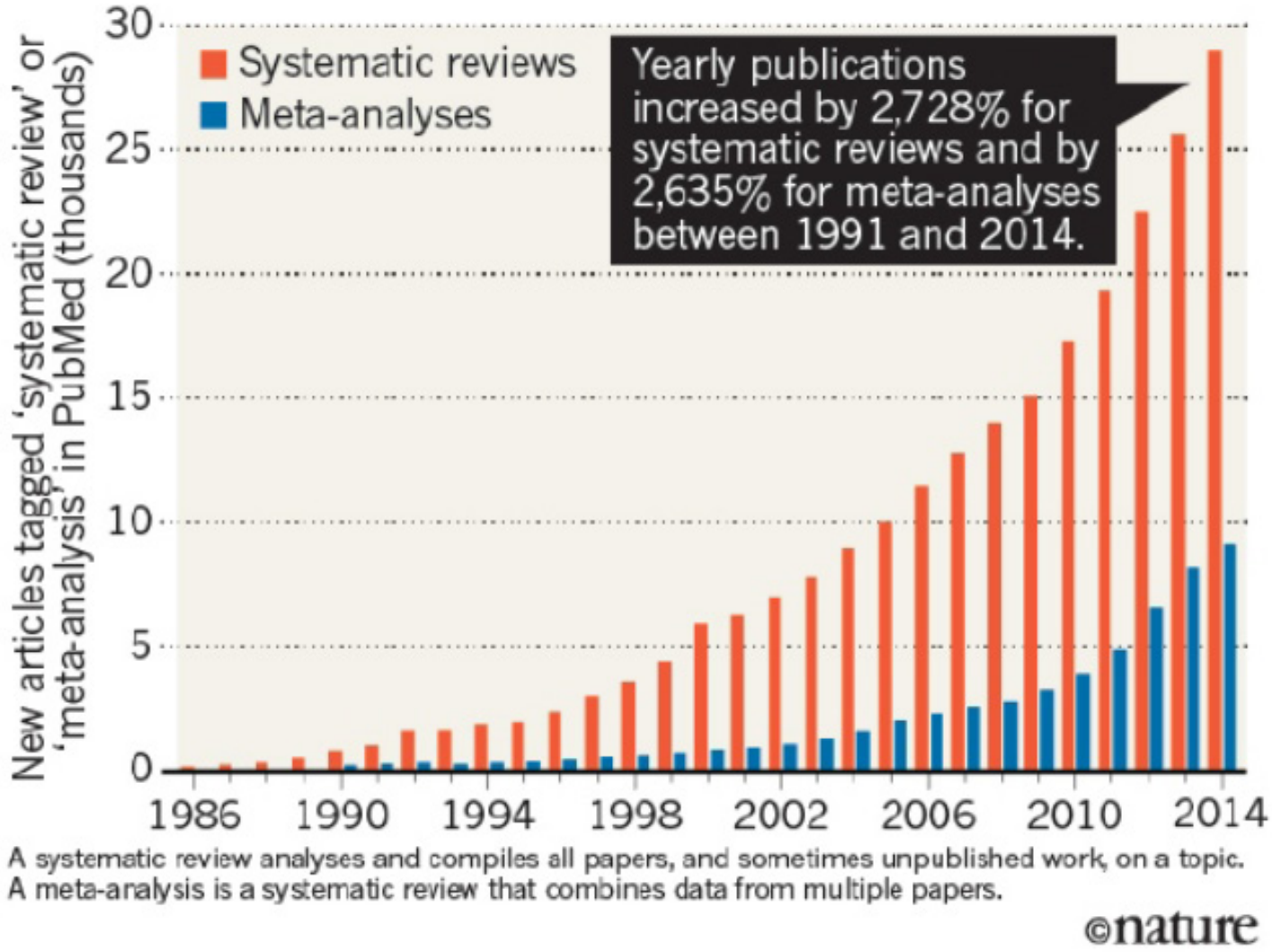
Green S, et al: Chapter 1 -Introduction. In *Cochrane handbook for systematic reviews of interventions*. Edited by: Higgins J, Green S. West Sussex, England: The Cochrane Collaboration. 2008.

Systematic review: Necessary elements

- Clearly stated, explicit objectives (including predefined research questions, eligibility criteria for studies, and a systematic search strategy)
- Transparent and reproducible methodology
- Systematic presentation of the characteristics and findings of included studies
- Assessment of the methodologic quality of included studies and overall quality of evidence

Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions*. Chichester, UK: Wiley Online Library; 2011.

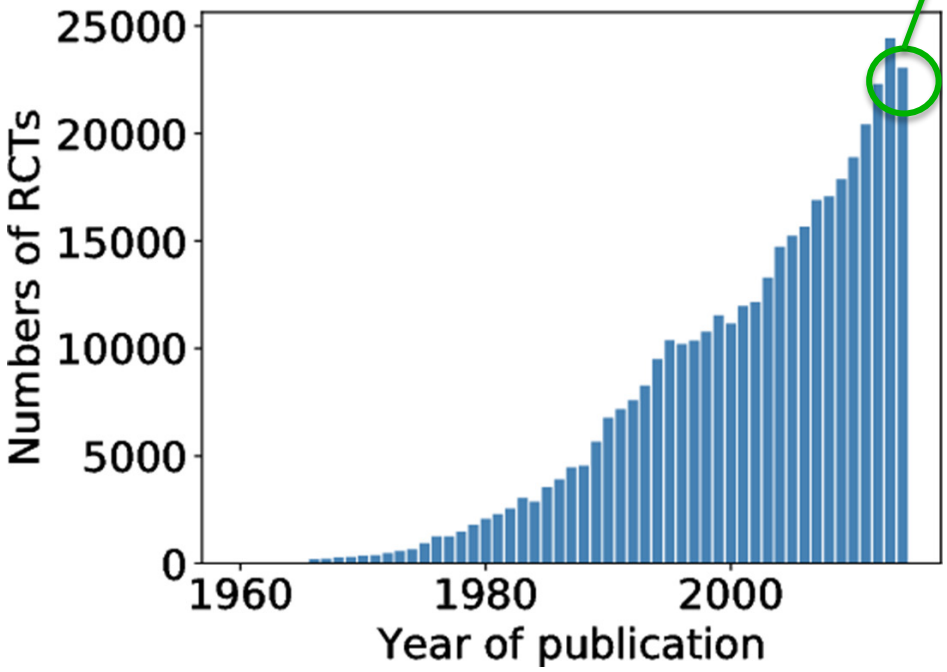
Systematic reviews and meta-analyses published in PubMed, annually, 1986-2014



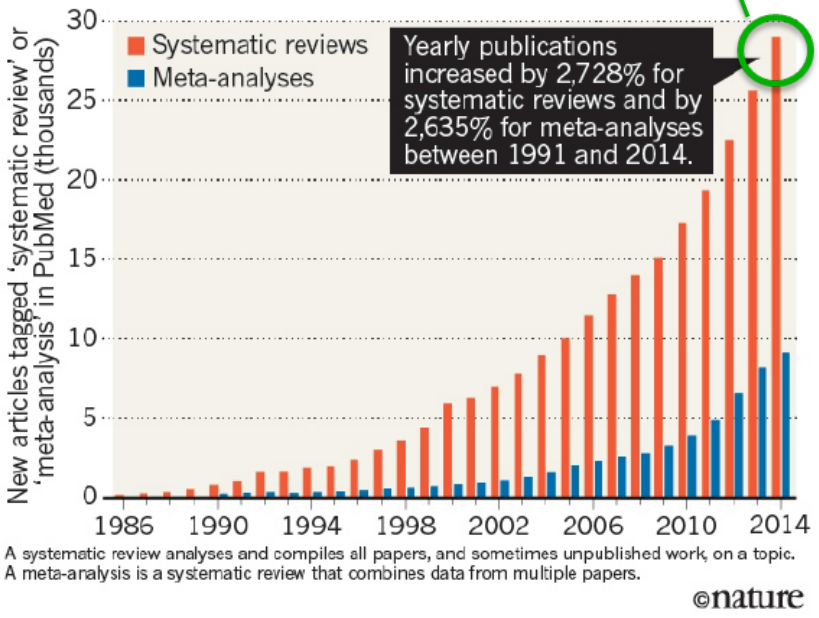
Baker M. Mass production of review articles is cause for concern. *Nature Trend Watch.* doi:10.1038/nature.2016.20617.

There seem to be some parallels…

2014: >20,000 RCTs



2014: >25,000 SRs



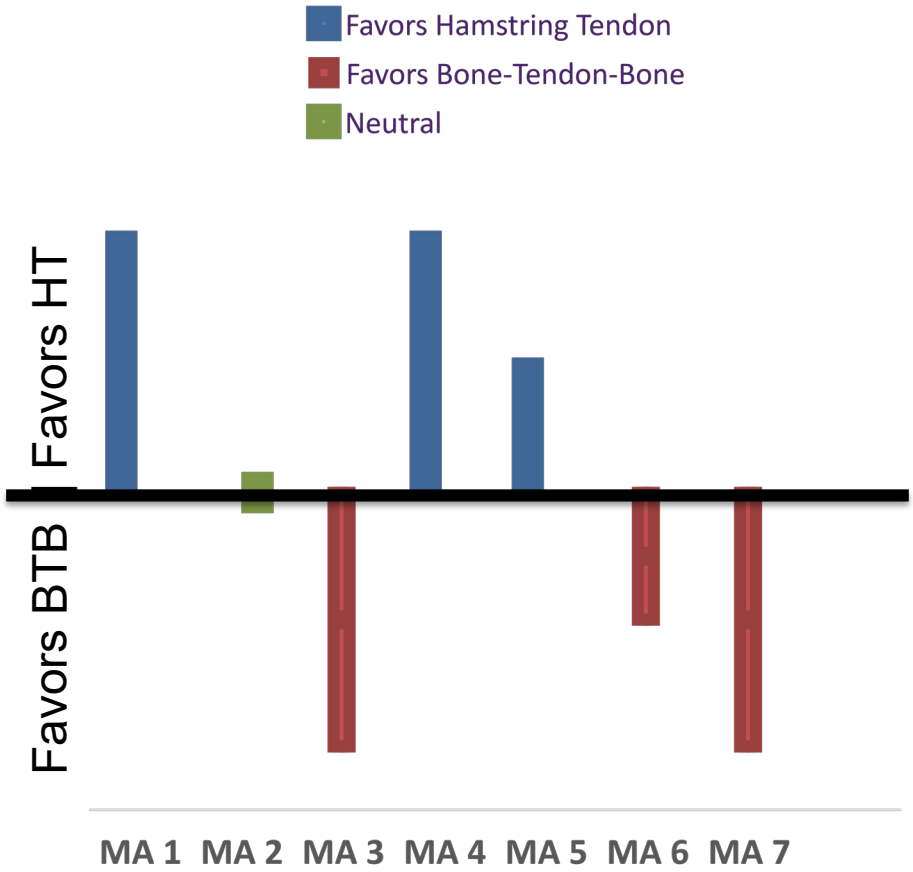
Ranking antidepressants

Meta-analysis 1	Meta-analysis 2	Meta-analysis 3
Paroxetine	Mirtazapine	Fluoxetine
Mirtazapine	Escitalopram	Sertraline
Venlafaxine	Venlafaxine	Bupropion
Nefazodone	Sertraline	Nefazodone
Sertraline	Citalopram	Trazodone
Duloxetine	Bupropion	Venlafaxine
Escitalopram	Paroxetine	Mirtazapine
Citalopram	Milnacipran	Escitalopram
Fluoxetine	Fluoxetine	Paroxetine
Bupropion	Duloxetine	Citalopram

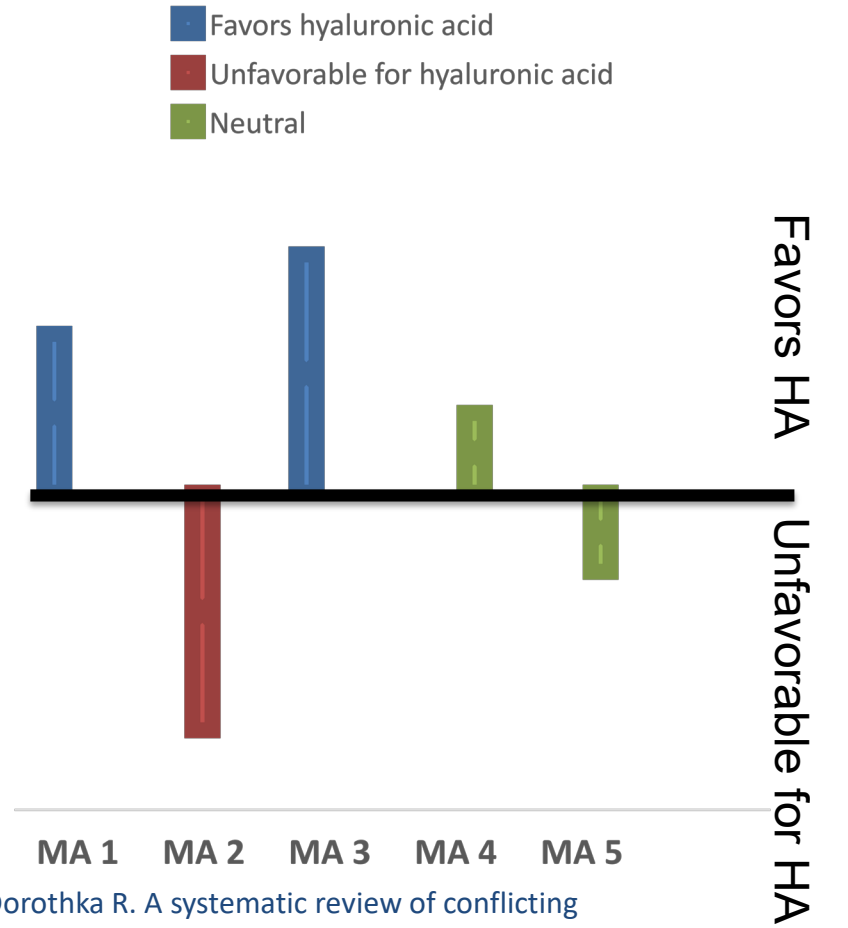
Ioannidis J. Ranking antidepressants. *Lancet*. 2009; 373: 1759-60.

Meta-analyses in orthopedic surgery

Graft choice in ACL reconstruction



Hyaluronic acid for knee OA



Vavken P, Dorothka R. A systematic review of conflicting meta-analyses in orthopaedic surgery. *Clin Orthop Relat Res*. 2009; 467: 2723-35.

The image was taken from another source and cannot be edited/modified. It may be difficult to see that there are two data points in the bar charts.

SRs of net benefit of mammography for breast cancer screening

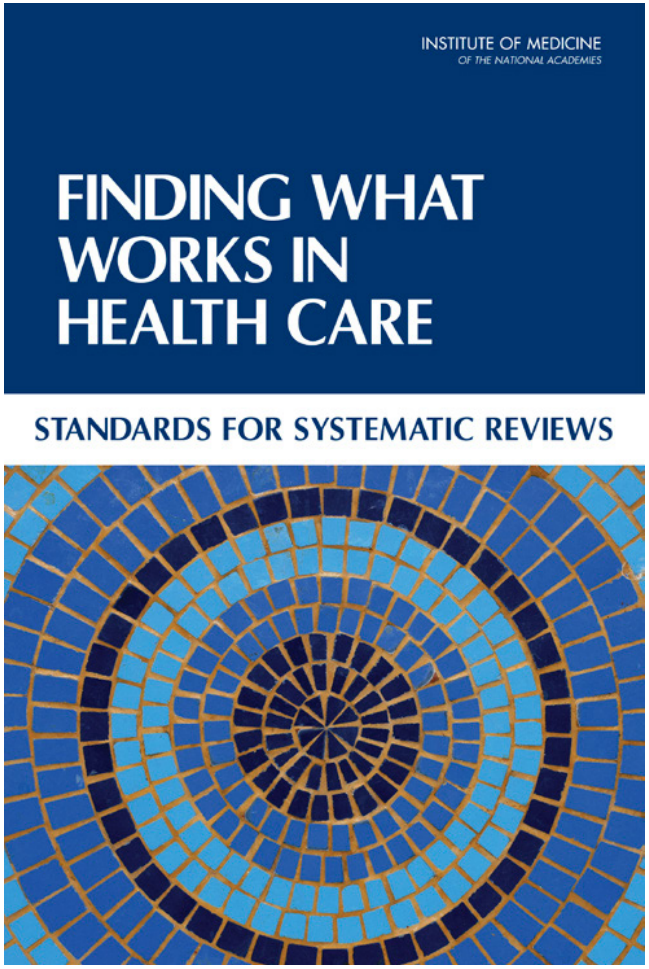
- 50 systematic reviews conducted between 2000-2015

Fig. 2 The number of systematic review conclusions by publication years during the period, including favourable conclusions (*orange*) and non-favourable conclusions (*cyan*)

- (8 RCTs have been conducted that evaluate the efficacy of conventional mammography screening on breast cancer mortality)

Raichand S, Dunn AG, Ong M-S, et al. Conclusions in systematic reviews of mammography for breast cancer screening and associations with review design and author characteristics. *Systematic Reviews*. 2017; 6:105.

What should you be looking for?: Standards for systematic reviews



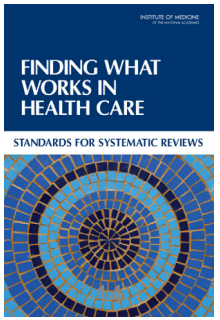
- 2011 Institute of Medicine report
- Standards cover all aspects of conducting a systematic evidence review, including:
  - **Initiating the review**
  - **Finding and assessing studies for inclusion in the report**
  - **Synthesizing the body of evidence**
  - **Reporting on the review**

Institute of Medicine of the National Academies. *Finding What Works in Health Care: Standards for Systematic Reviews*. Jill Eden, Laura Levit, Alfred Berg, et al, editors. National Academies Press. Washington, DC: 2011.

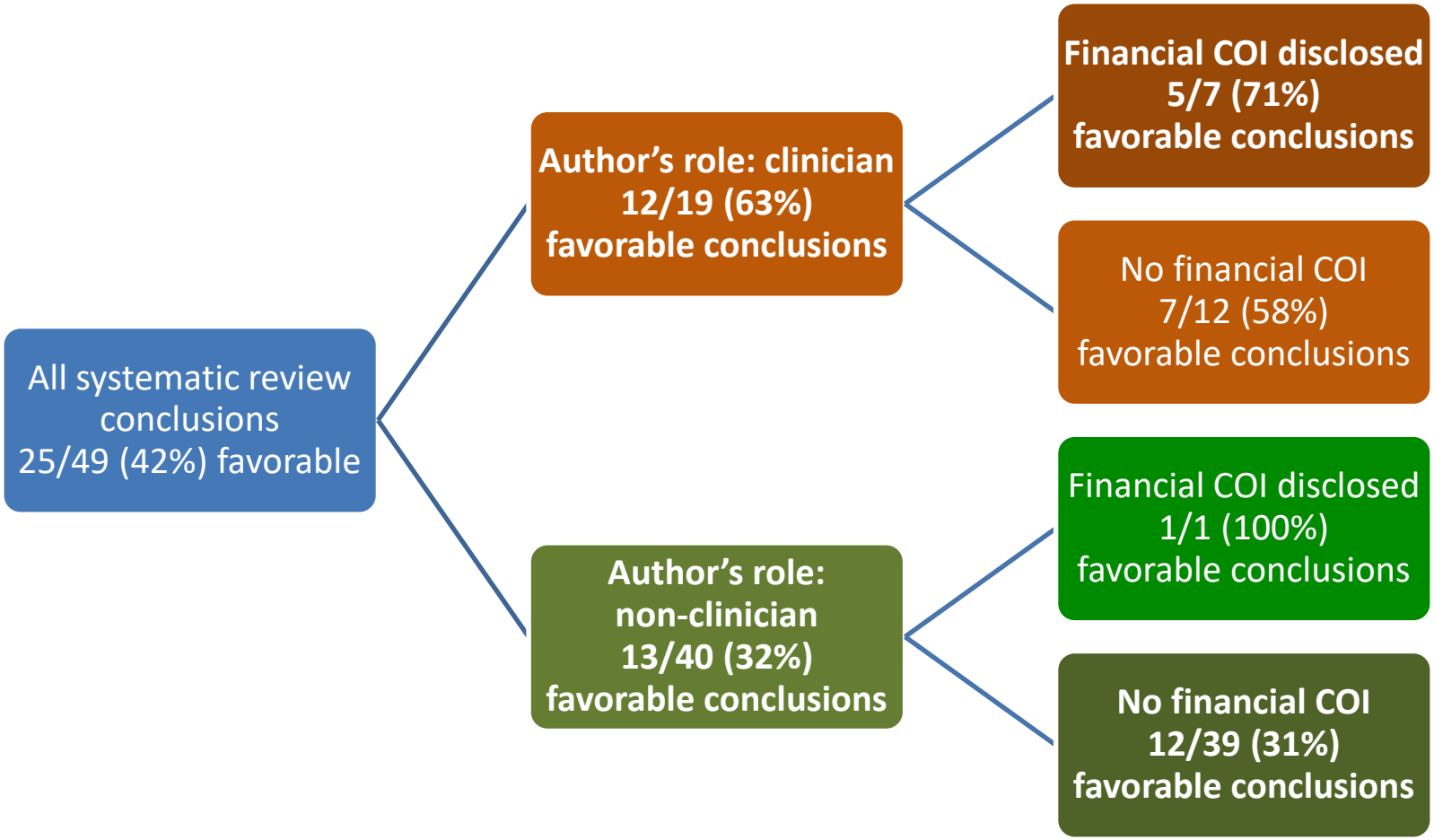
Standards for initiating a systematic review (highlights)

- Manage bias and conflict of interest of the team conducting and for individuals providing input into the systematic review
  - Require each team member and contributing individual to disclose potential COI and professional or intellectual bias
  - Exclude individuals with a clear financial conflict from being part of the review team
  - Exclude individuals from being part of the review team or from providing input whose professional or intellectual bias would diminish the credibility of the review in the eyes of the intended users

- Ensure user and stakeholder input as the review is designed and conducted
  - Protect the independence of the review team to make the final decisions about the design, analysis, and reporting of the review



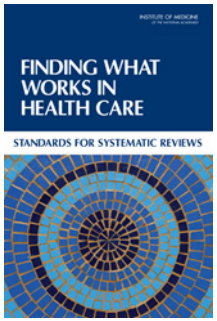
Association of financial and intellectual COI with systematic review outcomes: Example



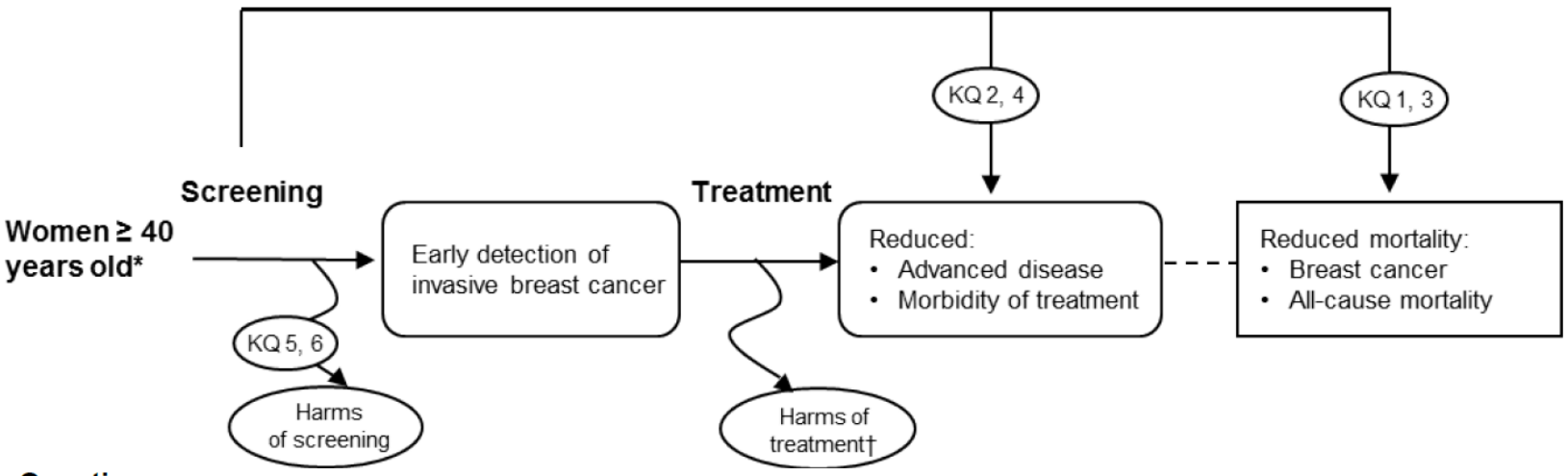
Raichand S, Dunn AG, Ong M-S, et al. Conclusions in systematic reviews of mammography for breast cancer screening and associations with review design and author characteristics. *Systematic Reviews*. 2017; 6:105.

Standards for initiating a systematic review (highlights)

- Formulate the topic for the systematic review:
  - Develop an analytic framework
  - Use a standard format to articulate each clinical question of interest
- Develop a systematic review protocol:
  - Describe the study inclusion/exclusion criteria
  - Describe which outcome measures, time points, interventions, and comparison groups will be addressed
  - Describe the search strategy for identifying relevant evidence, procedures for study selection, and the data extraction strategy
  - Describe the approach to critically appraising individual studies
  - Describe the method for evaluating the body of evidence, including qualitative and quantitative synthesis strategies
  - Describe any planned subgroup analyses
- Make the final protocol publicly available (e.g., publish on PROSPERO)



Analytic Framework



Key Questions:

In the target population of women age 40 years and older*:

1. What is the effectiveness of routine mammography screening in reducing breast cancer–specific and all-cause mortality, and how does it differ by age, risk factor‡, and screening interval?
2. What is the effectiveness of routine mammography screening in reducing the incidence of advanced breast cancer and treatment-related morbidity§, and how does it differ by age, risk factor‡, and screening interval?
3. How does the effectiveness of routine breast cancer screening in reducing breast cancer–specific and all-cause mortality vary by different screening modality‖?
4. How does the effectiveness of routine breast cancer screening in reducing the incidence of advanced breast cancer and treatment-related morbidity§ vary by different screening modality‖?
5. What are the harms¶ of routine mammography screening, and how do they differ by age, risk factor‡, and screening interval?
6. How do the harms¶ of routine breast cancer screening vary by different screening modality‖?

Nelson H, Cantor A, Humphrey L, et al. *Screening for breast cancer: A systematic review to update the 2009 U.S. Preventive Services Task Force Recommendation*. AHRQ Publication No. 14-05201-EF01. January 2016.

PICOTS criteria (Population, Interventions, Comparisons, Outcomes, Timing, Setting)

Criteria	Include	Exclude
Population	Women age ≥ 40 years.	Men; women age <40 years, women with pre-existing breast cancer; women with familial breast cancer syndromes; women with high-risk breast lesions; or women with previous large doses of chest radiation therapy (≥20 Gy) before age 30 years.
Interventions	KQs 1,5: Screening mammography, all methods.	KQs 1, 5: Mammography for diagnosis or surveillance.
Comparisons	KQs 1, 5: Mammography in women ages 40–49 vs. 50–59 vs. 60–69 vs. 70–79 years (or other age comparisons); annual mammography vs. biennial vs. triennial vs. alternate screening intervals vs. none.	KQs 1, 5: Data not provided by age, screening interval, or risk factor.
Outcomes	Benefits KQs 1: Reduced breast cancer mortality and all-cause mortality. Harms KQs 5: False-positive findings; anxiety; adverse effects on quality of life; false-positive biopsies; false-negative findings; false reassurance; overdiagnosis; overtreatment; radiation exposure.	Outcomes not listed as included.
Timing	Immediate, short-term, and long-term outcomes; duration of follow-up.	No follow-up.
Setting	Settings and populations of women applicable to U.S. primary care practice.	Settings not applicable to U.S. primary care practice.

Critical appraisal of individual studies

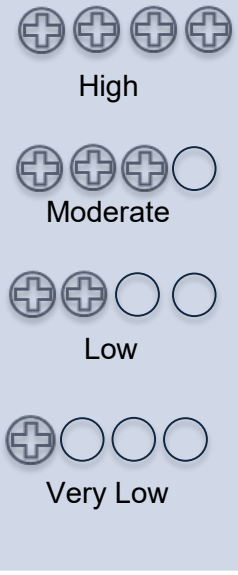
Example: Cochrane Collaboration tool for assessing risk of bias (summarizes the quality of the evidence for *single studies*)		
Bias domain	Source of bias	Authors' judgment
Selection bias	Random sequence generation	High; Low; Unclear risk of bias
	Allocation concealment	High; Low; Unclear risk of bias
Performance bias	Blinding of participants and personnel	High; Low; Unclear risk of bias
Detection bias	Blinding of outcome assessment	High; Low; Unclear risk of bias
Attrition bias	Incomplete outcome data	High; Low; Unclear risk of bias
Reporting bias	Selective reporting	High; Low; Unclear risk of bias
Other bias	Anything else, ideally prespecified	High; Low; Unclear risk of bias

Higgins JPT, Altman DG, Gotzsche P, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011; 343: d5928.

Critical appraisal of individual studies

Durand MA, Carpenter L, Dolan H, et al. Do interventions designed to support shared decision-making reduce health inequalities? A systematic review and meta-analysis. *PLoS One*. 2014; 9(4): e94670.

Method for evaluating a body of evidence

Example: GRADE criteria (summarizes the quality of the evidence *across an outcome*, or clinical question)		
Note outcome of interest/clinical question here		
GRADE criteria	Rating	Overall quality of the evidence (circle one)
Study design	RCT (starts as high quality) Non-RCT (starts as low quality)	
Risk of bias	No; Serious (-1); Very serious (-2)	
Inconsistency	No; Serious (-1); Very serious (-2)	
Indirectness	No; Serious (-1); Very serious (-2)	
Imprecision	No; Serious (-1); Very serious (-2)	
Publication bias	Undetected (0); Strongly suspected (-1)	
Other (upgrading factors, choose all that apply)	Large effect size (+1 to 2); dose-response (+1 to 2); No plausible confounding (+1 to 2)	

Guyatt G, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008; 336: 924.

Evaluating a body of evidence: Example

Key question (on second-generation antidepressants)	Summary of findings	Strength of Evidence
KQ 1a. MDD: Comparative efficacy and effectiveness—Onset of action	• Consistent results from seven trials suggest that mirtazapine has a significantly faster onset of action than citalopram, fluoxetine, paroxetine, and sertraline. Whether this difference can be extrapolated to other second-generation antidepressants is unclear. • Most other trials do not indicate a faster onset of action of one second-generation antidepressant compared with another.	**Moderate**
KQ 1a. Dysthymia: Comparative efficacy	• No head-to-head evidence exists. • Results from five placebo controlled trials were insufficient to draw conclusions about comparative efficacy.	**Insufficient**
KQ 4a. Comparative risk of harms: adverse events profiles	• Adverse-events profiles, based on 92 efficacy trials and 48 studies of experimental or observational design, are similar among second generation antidepressants. • The incidence of specific adverse events differs across antidepressants.	**High**

Gartlehner G, Hansen RA, Morgan LC, et al. Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression: An Update of the 2007 Comparative Effectiveness Review. AHRQ Publication No. 12-EHC012-EF. Rockville, MD: Agency for Healthcare Research and Quality. December 2011.

Consideration of validity and quality of body of evidence in systematic reviews: Example

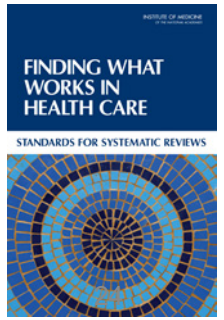
- Analysis of systematic reviews of neuraminidase inhibitors (used for prophylaxis and treatment of influenza), which have been the subject of uncertainty regarding their specific clinical benefits

Explicit consideration/discussion of important parameters of study validity and quality within systematic reviews of neuraminidase inhibitors		
Variable considered/ discussed	Reviews without financial COI (n = 19)	Reviews with financial COI (n = 7)
Publication bias	15	1
Ability to access comprehensive study data	10	0
Industry funding of available studies	8	0

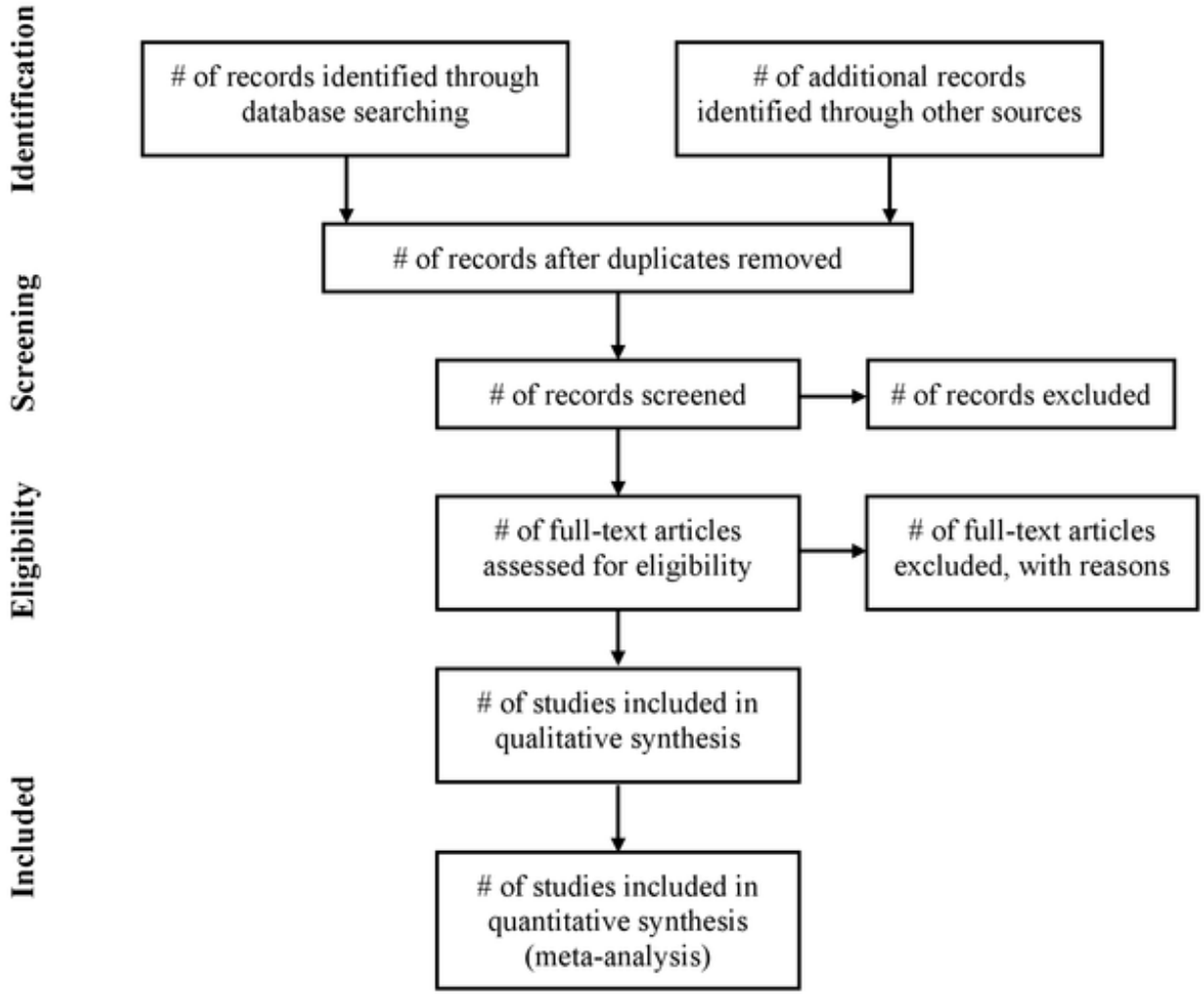
Dunn AG, Arachi D, Hudgins J, et al. Financial Conflicts of Interest and Conclusions About Neuraminidase Inhibitors for Influenza. *Ann Intern Med*. 2014; 161: 513-18.

Standards for findings and assessing individual studies for inclusion in the report (highlights)

- Conduct a comprehensive, systematic search for evidence
- Take action to address potentially biased reporting of research results
- Screen and select studies:
  - Include or exclude studies based on the protocol's prespecified criteria
  - Use observational studies in addition to RCTs to evaluate harms of interventions
- Document the search
- Manage data collection
- Critically appraise each study

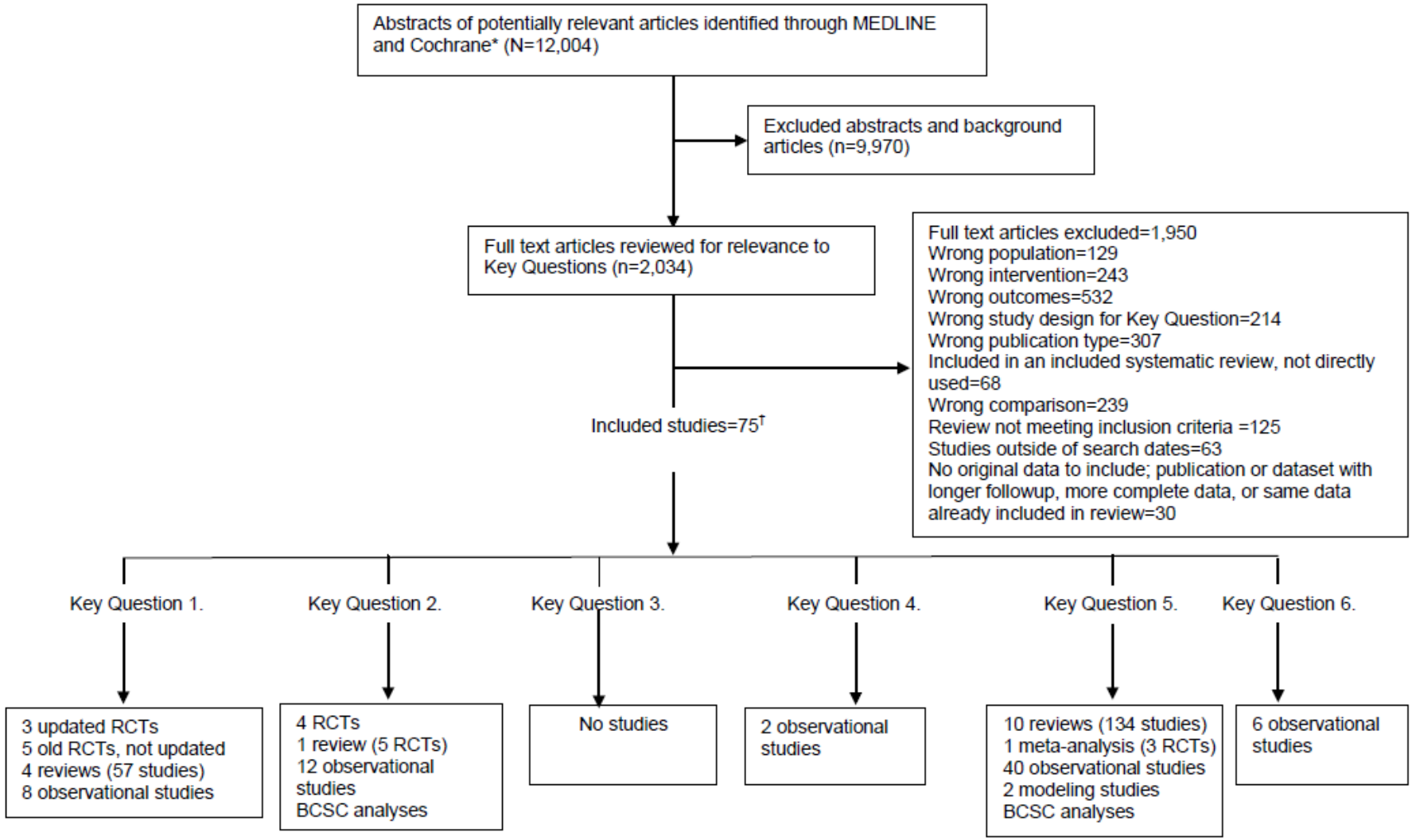


Flow of information through the different phases of a systematic review



Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*. 2009; 6(7): e1000097. https://doi.org/10.1371/journal.pmed.1000097

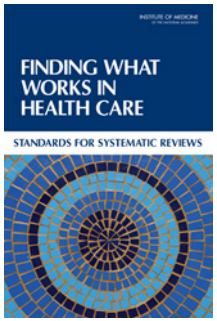
Literature flow diagram: Example



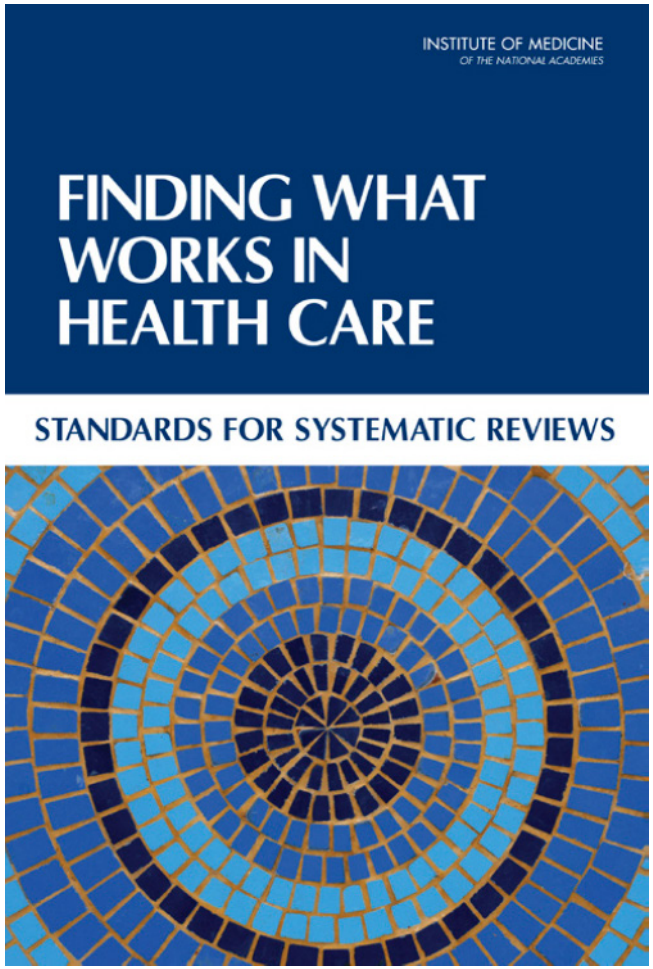
Nelson H, Cantor A, Humphrey L, et al. *Screening for breast cancer: A systematic review to update the 2009 U.S. Preventive Services Task Force Recommendation.* AHRQ Publication No. 14-05201-EF01. January 2016.

Standards for synthesizing the body of evidence (highlights)

- Use a prespecified method to evaluate the body of evidence
- Conduct qualitative synthesis and decide if the review will include a quantitative synthesis (meta-analysis); if so:
  - Address heterogeneity among study effects
  - Accompany all estimates with measures of statistical uncertainty
  - Assess the sensitivity of conclusions to changes in the protocol, assumptions, and study selection (do sensitivity analyses!)



Standards for reporting



PRISMA

TRANSPARENT REPORTING OF SYSTEMATIC REVIEWS AND META-ANALYSES

PRISMA checklist

Section/topic	#	Checklist item
TITLE		
Title	1	Identify the report as a systematic review, meta-analysis, or both.
ABSTRACT		
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.
INTRODUCTION		
Rationale	3	Describe the rationale for the review in the context of what is already known.
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).

PRISMA checklist

Section/topic	#	Checklist item
METHODS		
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.

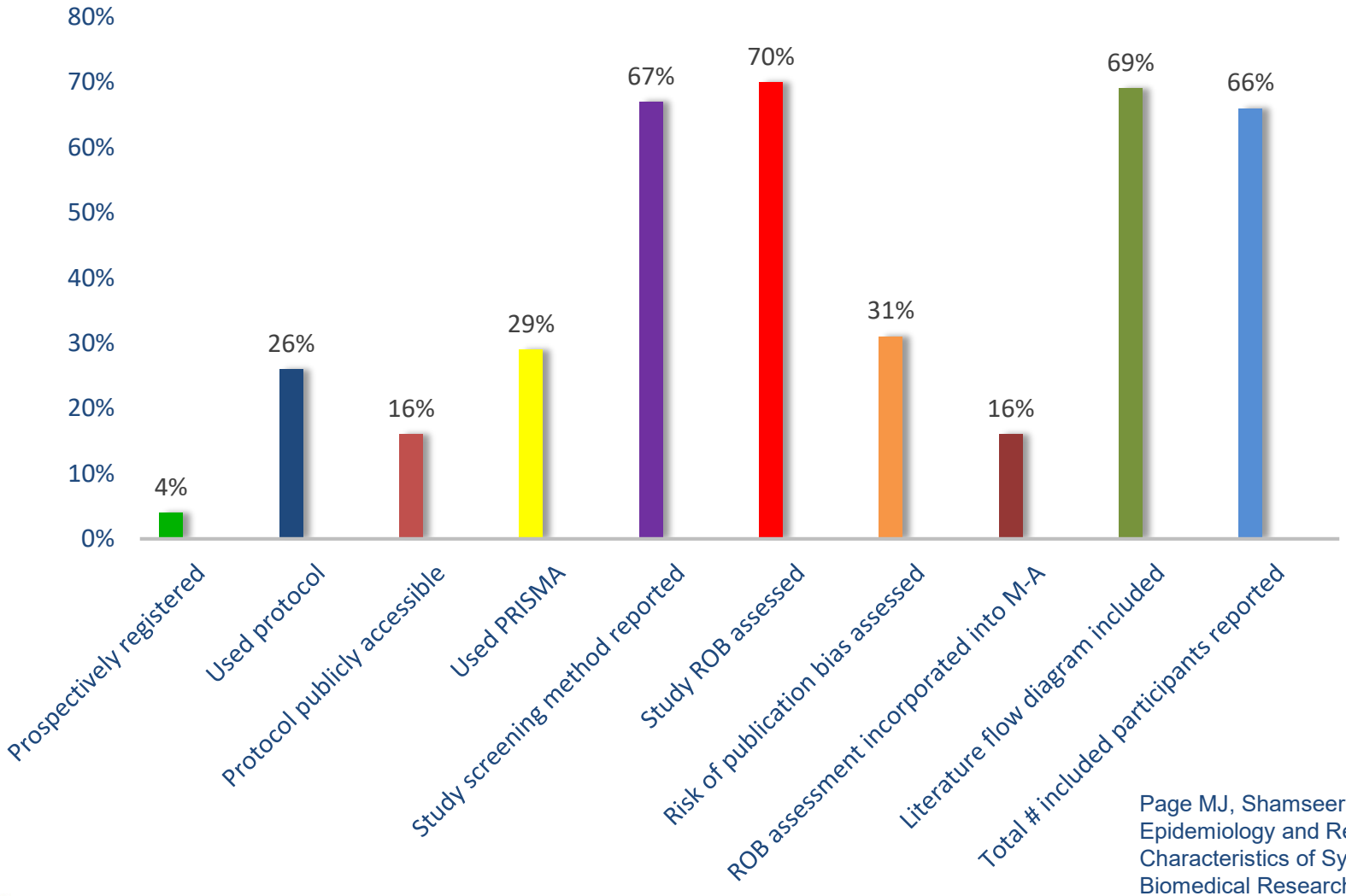
PRISMA checklist

Section/topic	#	Checklist item
RESULTS		
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).

PRISMA checklist

Section/topic	#	Checklist item
DISCUSSION		
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.
FUNDING		
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.

Characteristics of systematic reviews from convenience sample indexed in Medline, February 2014 (682 total identified)



Page MJ, Shamseer L, Altman DG, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. *PLOS Med*. 2016; 13(5): e1002028. doi:10.1371/journal.pmed.1002028.

Questions