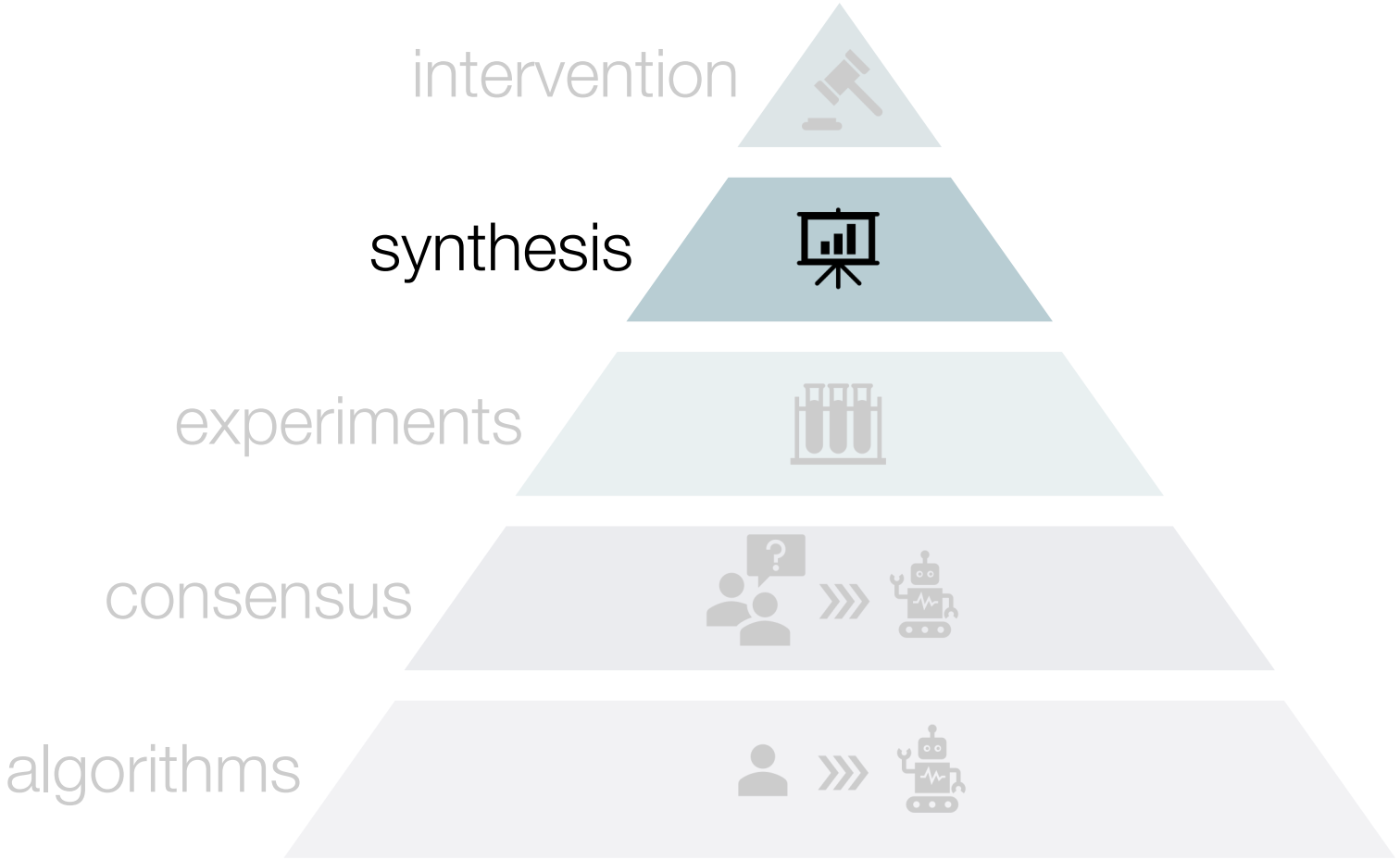
Evidence-based Decision Making
Synthesis: Limitations

Rui Mata, FS 2024

Version: February 23, 2024


intervention
synthesis
experiments
consensus
algorithms

Goals for today

- Review PRISMA guidelines
- Discuss limitations of research synthesis
- Discuss the possible automation of research synthesis

A brief history of reporting standards

1999: QUOROM: QUality Of Reporting Of Meta-analyses
see also CONSORT (Consolidated Standards of Reporting Trials)

2009: PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses

2015: PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols

2021: PRISMA 2020 statement, an updated guideline for reporting systematic reviews

Definitions

Table 1 PROSPERO and PRISMA-P

	Definition and objective
PROSPERO: International Prospective Register of Systematic Reviews	An online portal through which to register the intention to conduct a systematic review, with health-related outcomes, before it is initiated [16]. One of the main goals of PROSPERO is to make the intent of systematic reviews known before they are conducted in order to reduce the unplanned duplication of systematic reviews [15]. In addition, by requiring the documentation of *a priori* methods, the register facilitates increased transparency in the review process by allowing readers of systematic reviews to compare methods, outcomes, and analyses carried out with those planned in advance and judge whether such changes impact the results of a review.
PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols	A guideline to help authors prepare protocols for planned systematic reviews and meta-analyses that provides them with a minimum set of items to be included in the protocol. A protocol is intended to provide the rationale for the review and pre-planned methodological and analytic approach, prior to embarking on a review. Investigators should prepare a review protocol in advance of registering it in PROSPERO so that details requiring further consideration may be thought through in advance, avoiding the need for multiple amendments to registration information. PRISMA-P items have been derived largely from the PRISMA checklist and items of the PROSPERO register, in order to facilitate seamless registration.

https://www.crd.york.ac.uk/PROSPERO/

PRISMA-P Group, Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4(*1), e1000326–9. http://doi.org/10.1186/2046-4053-4-1

PRISMA and PRISMA-P Guidelines: Benefits

Table 4 Proposed stakeholders, actions, and potential benefits for supporting adherence to PRISMA-P

Stakeholder	Proposed action	Potential benefits
Funders	Promote or mandate adherence to PRISMA-P or use PRISMA-P as a template for systematic review proposals for grant applications	Improved quality, completeness, and consistency of systematic review proposal submissions
		Standardized protocol content will improve peer review efficiency and investigator understanding of requirements
Systematic review authors/ groups/organizations	Use/adhere to PRISMA-P during protocol development	Improved quality, completeness, and consistency of protocol content
		Enables reviewers to anticipate and avoid future changes to review methods (i.e., outcomes)
		Increased awareness of minimum content for protocol reporting
		Improved completeness of reporting of completed reviews
PROSPERO (and other review registries)	Encourage the development of PRISMA-P-based protocols	Improved quality of registry entries
		Improved consistency across registry entries, protocols, and systematic reviews
Practice guideline developers	Use PRISMA-P to gauge the completeness of protocols and facilitate detection of selective reporting when considering reviews for guideline inclusion	Enables easy comparison across protocols, registry entries, and completed systematic reviews
Policymakers	Advocate use of PRISMA-P by those funding and carrying out systematic reviews	May yield better quality, more complete, and more consistent reviews to inform decision-making
Journal editors	Encourage compliance to PRISMA-P for authors submitting protocols for publication	Improved quality, completeness, and consistency of protocols over those published in journals not endorsing PRISMA-P
	Offer PRISMA-P as a template to assist in protocol writing for publication	Increased efficiency in protocol peer and author understanding of journal requirements
		Improved transparency and interpretation of reviews by readers
Educators	Use PRISMA-P as a training tool	Simplified teaching and grading of protocols
	Encourage adherence in students submitting protocols for coursework	Improved quality, completeness, and consistency of protocol content
Students	Develop protocols for coursework or research using PRISMA-P	Improved understanding of the minimum protocol content
		Well-trained systematic reviewer going into the workforce

PRISMA and PRISMA-P Guidelines

Box 2 |: Helping to develop the research question(s): the PICOS approach

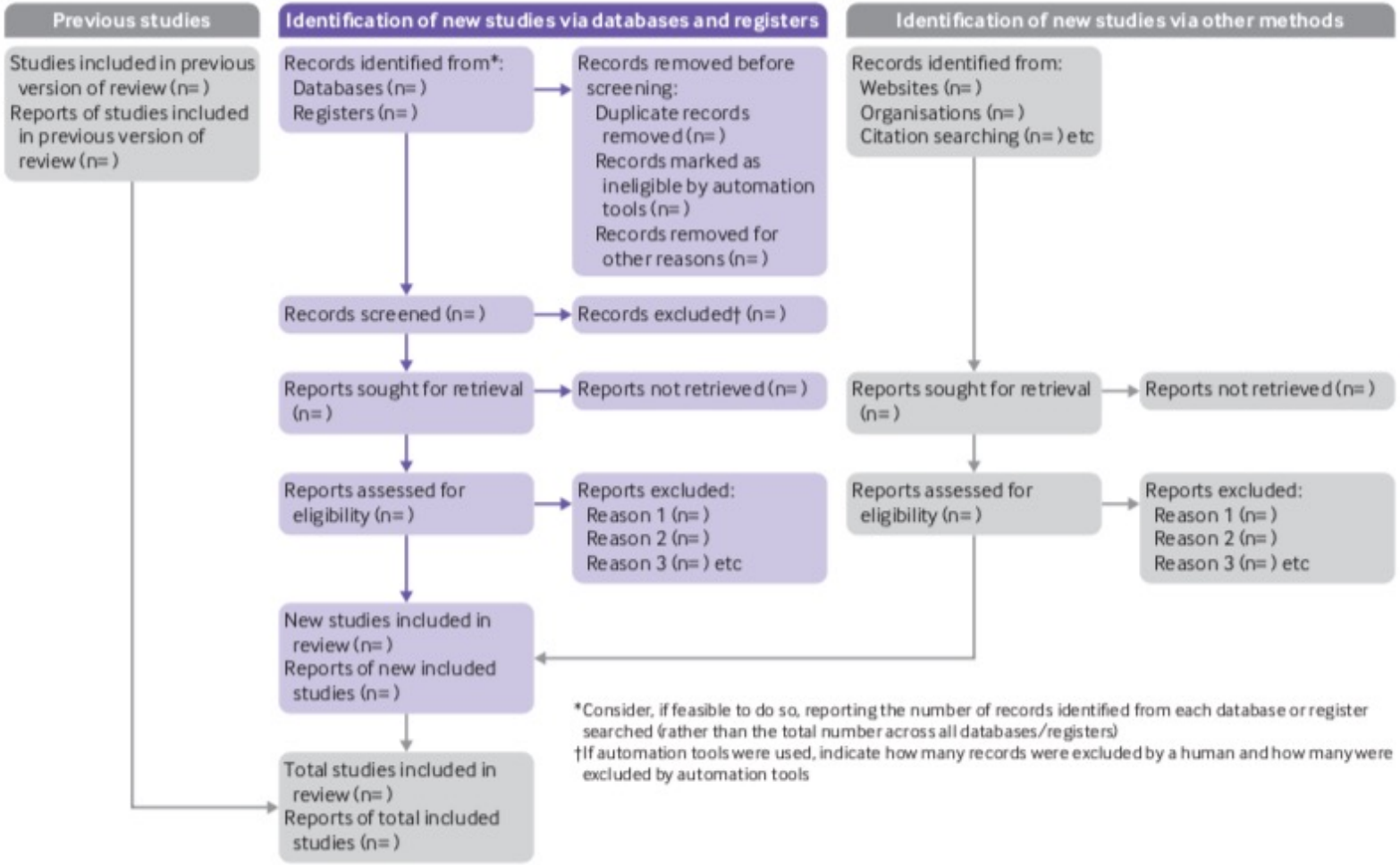
Formulating relevant and precise questions that can be answered in a systematic review can be complex and time consuming. A structured approach for framing questions that uses five components may help facilitate the process. This approach is commonly known by the acronym "PICOS" where each letter refers to a component: the patient population or the disease being addressed (P), the interventions or exposure (I), the comparator group (C), the outcome or endpoint (O), and the study design chosen (S).[186] Issues relating to PICOS affect several PRISMA items (items 6, 8, 9, 10, 11, and 18).

- *P*—Providing information about the population requires a precise definition of a group of participants (often patients), such as men over the age of 65 years, their defining characteristics of interest (often disease), and possibly the setting of care considered, such as an acute care hospital.
- *I*—The interventions (exposures) under consideration in the systematic review need to be transparently reported. For example, if the reviewers answer a question regarding the association between a woman's prenatal exposure to folic acid and subsequent offspring's neural tube defects, reporting the dose, frequency, and duration of folic acid used in different studies is likely to be important for readers to interpret the review's results and conclusions. Other interventions (exposures) might include diagnostic, preventive, or therapeutic treatments; arrangements of specific processes of care; lifestyle changes; psychosocial or educational interventions; or risk factors.
- *C*—Clearly reporting the comparator (control) group intervention(s)—such as usual care, drug, or placebo—is essential for readers to fully understand the selection criteria of primary studies included in the systematic review, and might be a source of heterogeneity investigators have to deal with. Comparators are often poorly described. Clearly reporting what the intervention is compared with is important and may sometimes have implications for the inclusion of studies in a review—many reviews compare with "standard care," which is otherwise undefined; this should be properly addressed by authors.
- *O*—The outcomes of the intervention being assessed—such as mortality, morbidity, symptoms, or quality of life improvements—should be clearly specified as they are required to interpret the validity and generalisability of the systematic review's results.
- *S*—Finally, the type of study design(s) included in the review should be reported. Some reviews include only reports of randomised trials, whereas others have broader design criteria and include randomised trials and certain types of observational studies. Still other reviews, such as those specifically answering questions related to harms, may include a wide variety of designs ranging from cohort studies to case reports. Whatever study designs are included in the review, these should be reported.

Independently from how difficult it is to identify the components of the research question, the important point is that a structured approach is preferable, and this extends beyond systematic reviews of effectiveness. Ideally the PICOS criteria should be formulated a priori, in the systematic review's protocol, although some revisions might be required because of the iterative nature of the review process. Authors are encouraged to report their PICOS criteria and whether any modifications were made during the review process. A useful example in this realm is the appendix of the "systematic reviews of water fluoridation" undertaken by the Centre for Reviews and Dissemination.[187]

PRISMA-P Group, Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4(*1), e1000326–9. http://doi.org/10.1186/2046-4053-4-1

PRISMA 2020



Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Bmj*, *372*, n71. http://doi.org/10.1136/bmj.n71

Noteworthy changes

Box 2: Noteworthy changes to the PRISMA 2009 statement

- Inclusion of the abstract reporting checklist within PRISMA 2020 (see item #2 and table 2).
- Movement of the 'Protocol and registration' item from the start of the Methods section of the checklist to a new Other section, with addition of a sub-item recommending authors describe amendments to information provided at registration or in the protocol (see item #24a-24c).
- Modification of the 'Search' item to recommend authors present full search strategies for *all* databases, registers and websites searched, not just at least one database (see item #7).
- Modification of the 'Study selection' item in the Methods section to emphasise the reporting of how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process (see item #8).
- Addition of a sub-item to the 'Data items' item recommending authors report how outcomes were defined, which results were sought, and methods for selecting a subset of results from included studies (see item #10a).
- Splitting of the 'Synthesis of results' item in the Methods section into six sub-items recommending authors describe: the processes used to decide which studies were eligible for each synthesis; any methods required to prepare the data for synthesis; any methods used to tabulate or visually display results of individual studies and syntheses; any methods used to synthesise results; any methods used to explore possible causes of heterogeneity among study results (such as subgroup analysis, meta-regression); and any sensitivity analyses used to assess robustness of the synthesised results (see item #13a-13f).
- Addition of a sub-item to the 'Study selection' item in the Results section recommending authors cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded (see item #16b).
- Splitting of the 'Synthesis of results' item in the Results section into four sub-items recommending authors: briefly summarise the characteristics and risk of bias among studies contributing to the synthesis; present results of all statistical syntheses conducted; present results of any investigations of possible causes of heterogeneity among study results; and present results of any sensitivity analyses (see item #20a-20d).
- Addition of new items recommending authors report methods for and results of an assessment of certainty (or confidence) in the body of evidence for an outcome (see items #15 and #22).
- Addition of a new item recommending authors declare any competing interests (see item #26).
- Addition of a new item recommending authors indicate whether data, analytic code and other materials used in the review are publicly available and if so, where they can be found (see item #27).

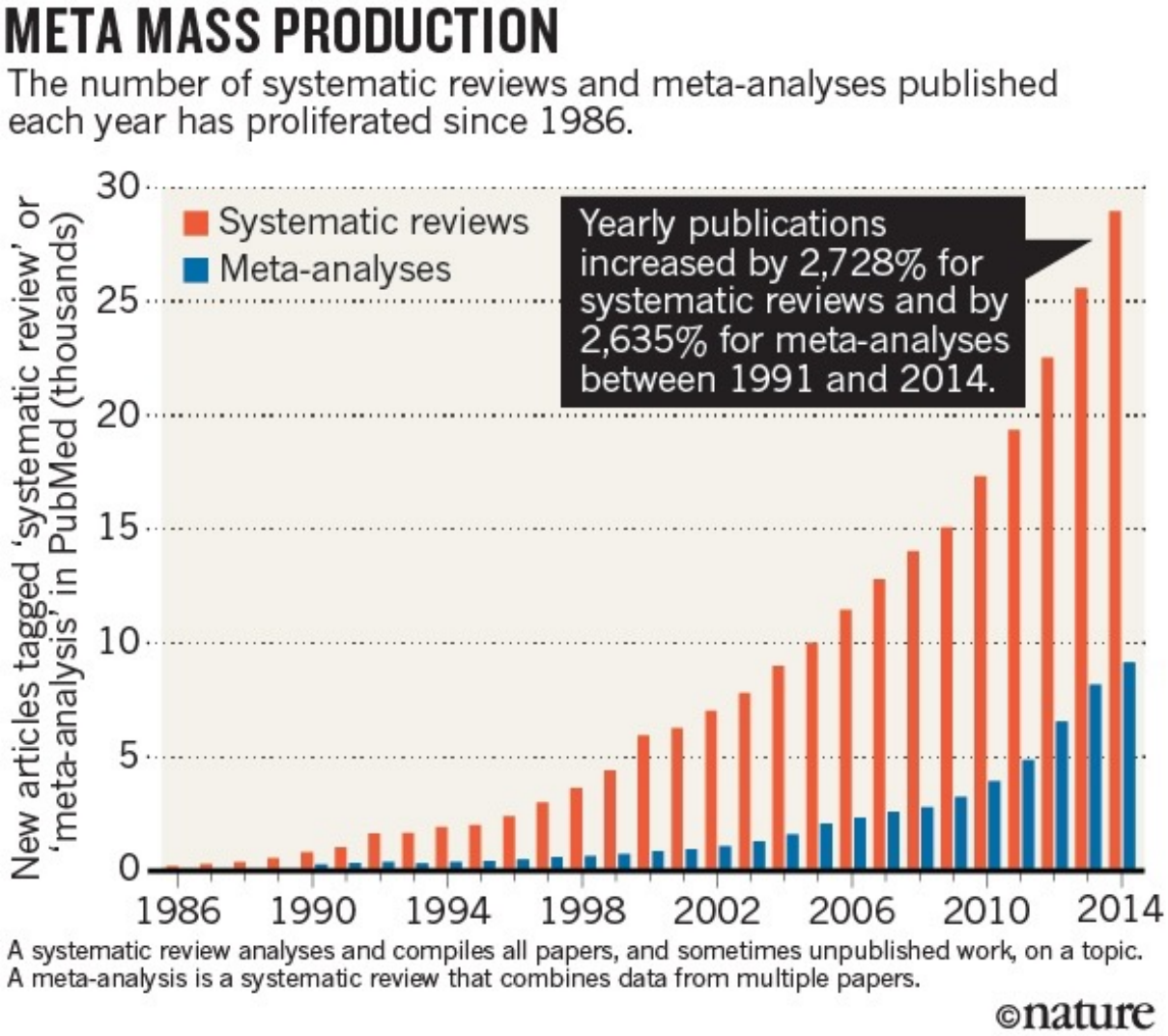
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Bmj*, *372*, n71. http://doi.org/10.1136/bmj.n71

A long history of critique of research synthesis

- Garbage-in, garbage out ("mega-silliness", Eysenck, 1978)

- "Instead of promoting evidence-based medicine and health care, these instruments often serve mostly as easily produced publishable units or marketing tools" (Ioannidis, 2016)

A long history of critique of research synthesis



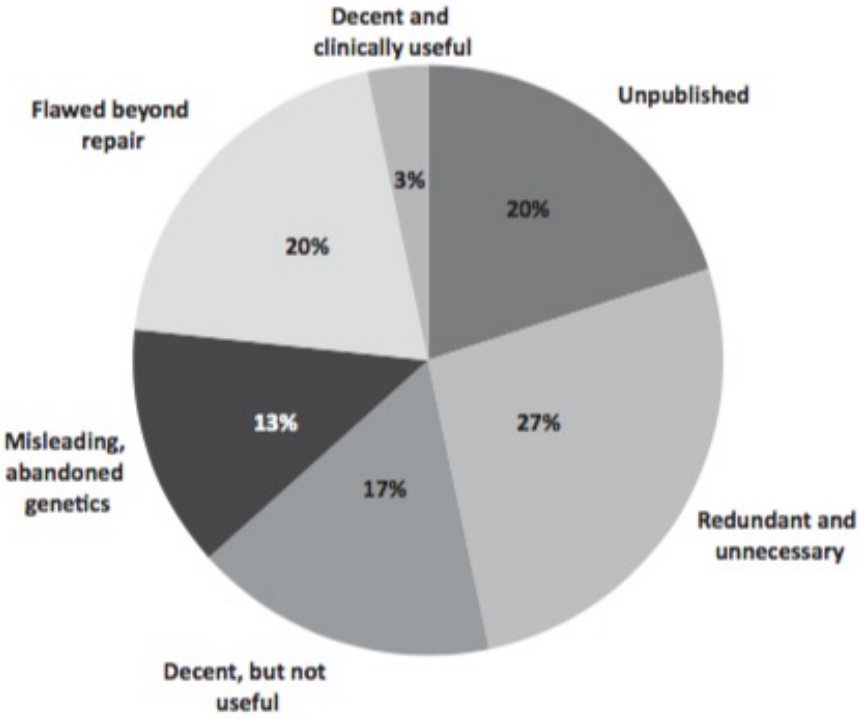
https://www.nature.com/news/mass-production-of-review-articles-is-cause-for-concern-1.20617

A long history of critique of research synthesis

Policy Points:

- Currently, there is massive production of unnecessary, misleading, and conflicted systematic reviews and meta-analyses. Instead of promoting evidence-based medicine and health care, these instruments often serve mostly as easily produced publishable units or marketing tools.
- Suboptimal systematic reviews and meta-analyses can be harmful given the major prestige and influence these types of studies have acquired.
- The publication of systematic reviews and meta-analyses should be realigned to remove biases and vested interests and to integrate them better with the primary production of evidence.

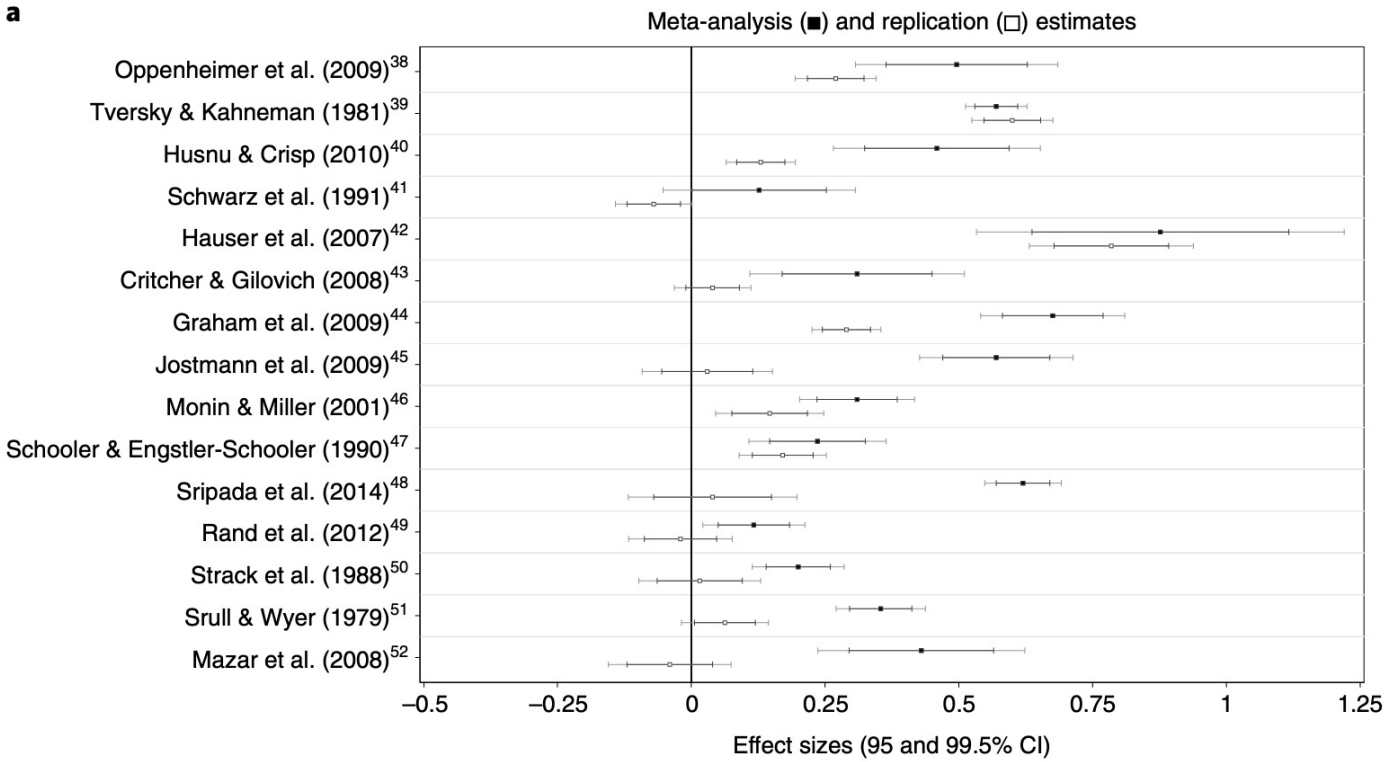
Figure 4. A Summary Overview of Currently Produced Meta-analyses



Ioannidis, J. P. A. (2016). The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly, 94*(3), 485–514. http://doi.org/10.1111/1468-0009.12210

Limitations of Research Synthesis

Results of meta-analyses and replication studies



"We find that meta-analytic effect sizes are significantly different from replication effect sizes for 12 out of the 15 meta-replication pairs. These differences are systematic and, on average, meta-analytic effect sizes are almost three times as large as replication effect sizes. We also implement three methods of correcting meta-analysis for bias, but these methods do not substantively improve the meta-analytic results." - these findings suggest that meta-analysis may not be able to adjust inflated effect sizes that arise from publication bias/selective reporting.

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. ***Nature Human Behaviour***, ***4***(4), 423–434. http://doi.org/10.1038/s41562-019-0787-z

Limitations of Research Synthesis

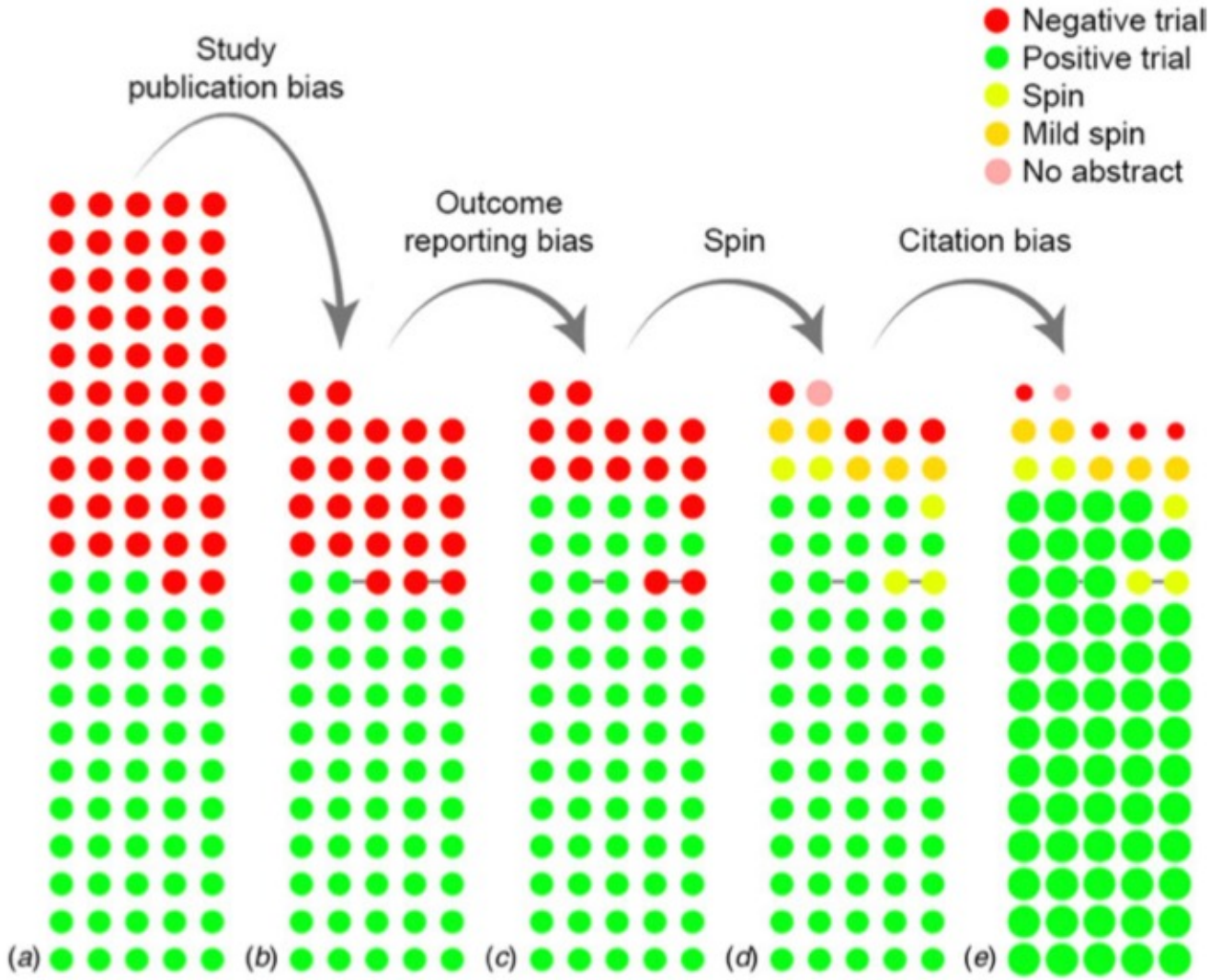
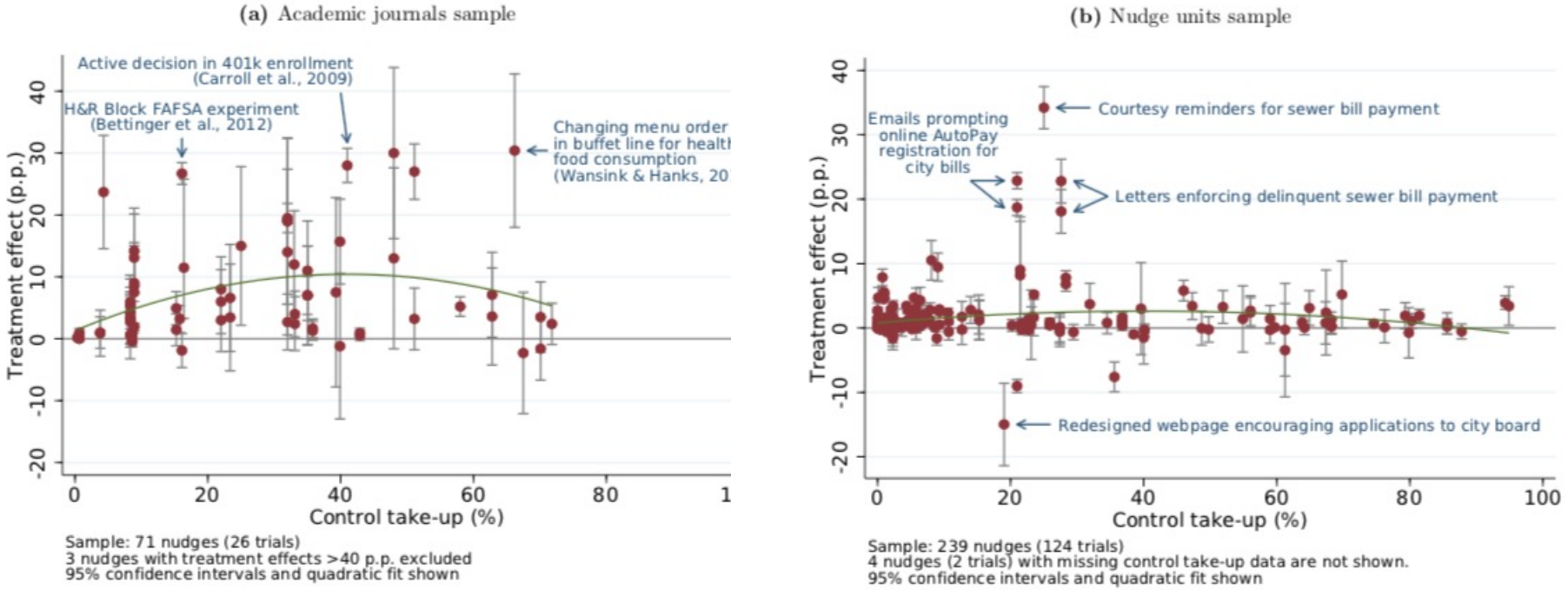


Fig. 1. The cumulative impact of reporting and citation biases on the evidence base for antidepressants. (*a*) displays the initial, complete cohort of trials, while (*b*) through (*e*) show the cumulative effect of biases. Each circle indicates a trial, while the color indicates the results or the presence of spin. Circles connected by a grey line indicate trials that were published together in a pooled publication. In (*e*), the size of the circle indicates the (relative) number of citations received by that category of studies.

De Vries, Y. A., Roest, A. M., De Jonge, P., Cuijpers, P., Munafò, M. R., & Bastiaansen, J. A. (2018). The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: The case of depression. *Psychological Medicine, 48*(15), 2453–2455. http://doi.org/10.1017/S0033291718001873

Limitations of Research Synthesis



Nudge interventions have quickly expanded from academic studies to larger implementation in so-called Nudge Units in governments. This provides an opportunity to compare interventions in research studies, versus at scale. We assemble a unique data set of 126 RCTs covering 23 million individuals, including all trials run by two of the largest Nudge Units in the United States. We compare these trials to a sample of nudge trials in academic journals from two recent meta-analyses. In the Academic Journals papers, the average impact of a nudge is very large—an 8.7 percentage point take-up effect, which is a 33.4% increase over the average control. In the Nudge Units sample, the average impact is still sizable and highly statistically significant, but smaller at 1.4 percentage points, an 8.0% increase. We document three dimensions which can account for the difference between these two estimates: (i) statistical power of the trials; (ii) characteristics of the interventions, such as topic area and behavioral channel; and (iii) selective publication. A meta-analysis model incorporating these dimensions indicates that selective publication in the Academic Journals sample, exacerbated by low statistical power, explains about 70 percent of the difference in effect sizes between the two samples. Different nudge characteristics account for most of the residual difference.

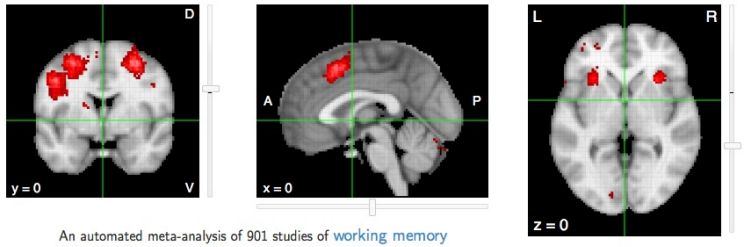
DellaVigna, S., & Linos, E. (2022). RCTs to Scale: Comprehensive Evidence from Two Nudge Units. Econometrica, 90, 81-116. https://doi.org/10.3982/ECTA18709

Automation of Research Synthesis

Neurosynth

Neurosynth is a platform for large-scale, automated synthesis of functional magnetic resonance imaging (fMRI) data.

It takes thousands of published articles reporting the results of fMRI studies, chews on them for a bit, and then spits out images that look like this:



An automated meta-analysis of 901 studies of working memory

Database Status

413429 activations reported in 11406 studies

Interactive, downloadable meta-analyses of 3107 terms

Functional connectivity and coactivation maps for over 150,000 brain locations

http://neurosynth.org

Human Behaviour-Change Project

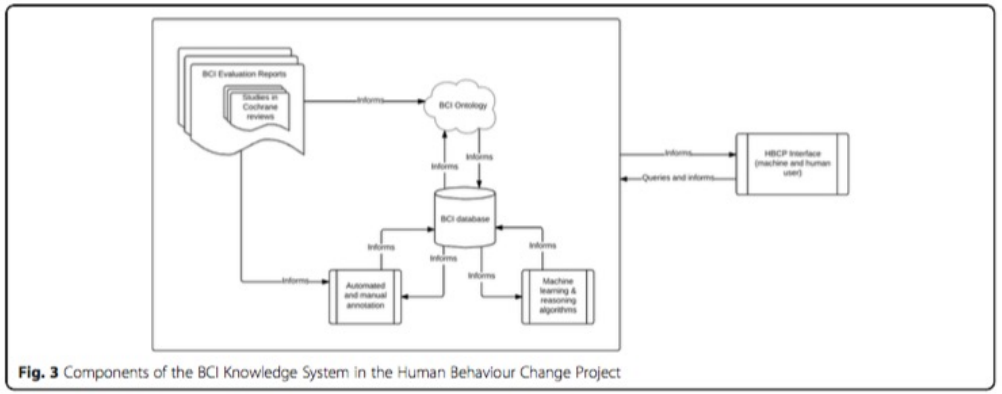


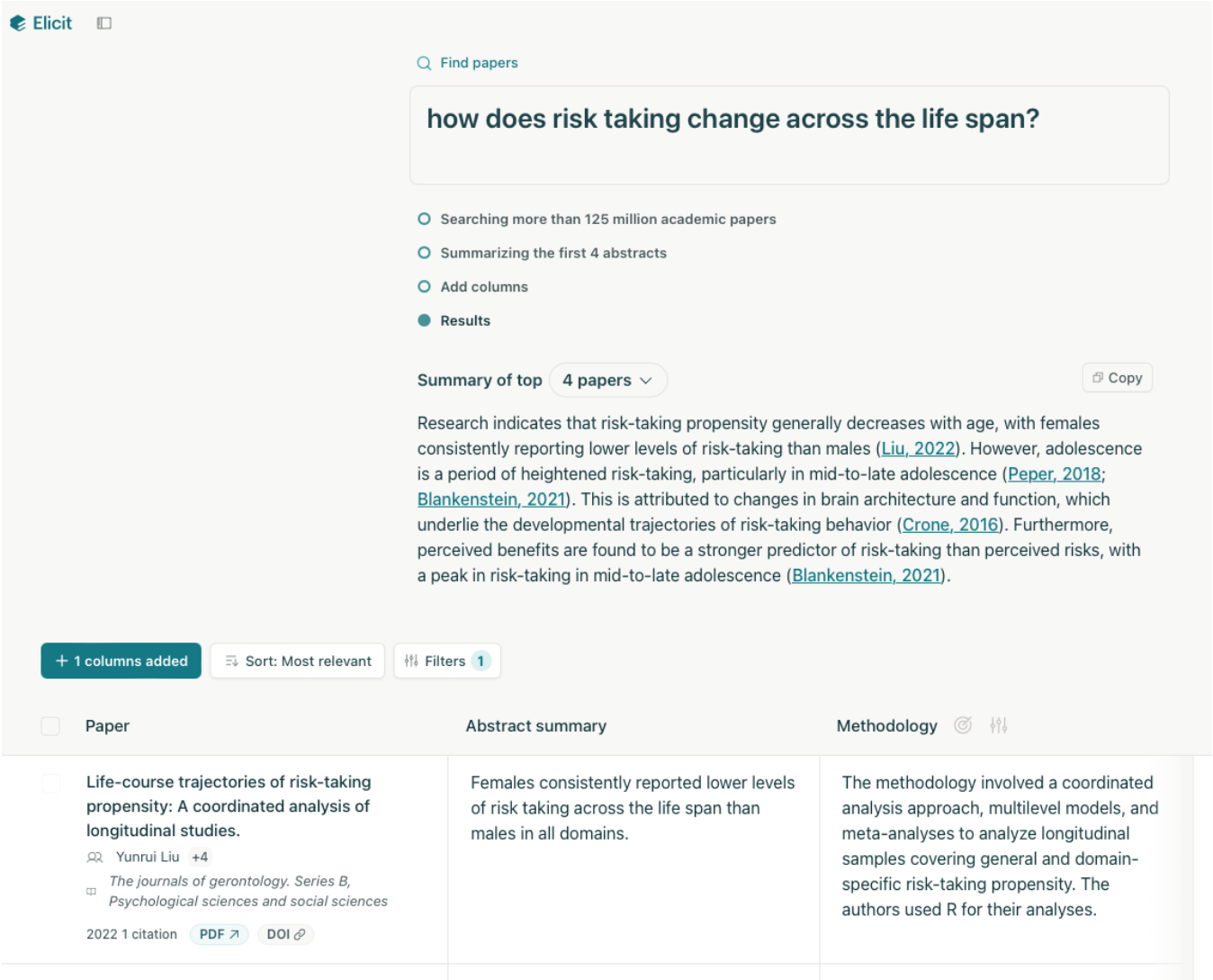
Fig. 3 Components of the BCI Knowledge System in the Human Behaviour Change Project

http://www.ucl.ac.uk/human-behaviour-change

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. ***Nature Methods,*** ***8***(8), 665–670. http://doi.org/10.1038/nmeth.1635

Michie, S., Thomas, J., Johnston, M., Mac Aonghusa, P., Shawe-Taylor, J., Kelly, M. P., et al. (2017). The Human Behaviour-Change Project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation, ***Implementation Science***, 1–12. http://doi.org/10.1186/s13012-017-0641-5

Automation of Research Synthesis



https://elicit.com

Summary

- **PRISMA:** PRISMA guidelines offer a clear set of standards for conducting research syntheses
- **Limitations**: research synthesis is not a panacea and cannot provide accurate estimates of effects in the face of large reporting biases (publication bias & file-drawer problem)
- **Automation**: automatization is possible but problems of consensus regarding terminology abound, need for manual curation and validation remains.