COMMENTARY

Umbrella reviews: what they are and why we need them

Stefania Papatheodorou[1,2]

Introduction

Evidence in clinical research and public health is hierarchical. From animal studies to observational epidemiology, randomized control trials and their synthesis, there is a pipeline of different study designs providing different nature and quality of information. Systematic reviews and meta-analyses stand on top of this hierarchy most of the times and are key components of evidence-based medicine [1]. Their production and publication has increased exponentially in the last decade across almost all disciplines with approximately 20,000 records labeled as meta-analysis in 2018 compared to 3300 in 2008. One would expect that this growth would reflect the increase in the number of primary observational and experimental studies but this is not true because all Pubmed-indexed items increased by 153% from 1991 to 2014 [2].

This abundance of published evidence synthesis cannot be considered only positive. It is not uncommon for meta-analyses on the same research question to reach different conclusions even when published within the same year. This unavoidably leads to confusion and debate in clinicians and public health policy makers on where to base their decisions on. There are several preventive measures proposed in order to minimize bias in meta-analyses. A priori publication of protocols is highly encouraged and each analytical step and every subjective judgment call should be reported [3]. Finally there are tools to appraise the reporting [4] and quality [5] of the published systematic reviews and meta-analyses.

Knowledge gap

The information that systematic reviews and meta-analyses are conveying is directly associated with the quantity, quality and comparability of the available evidence from the primary studies. In observational and experimental research, differences in interventions, metrics, outcomes, designs, participants and settings-collectively characterized as sources of heterogeneity—as well as confounding, known and unknown biases, are often inflated in a meta-analysis [6]. Because of the increase in the sample size and statistical power, the pooled estimates can have spuriously tight confidence intervals. Sometimes it is doubtful whether statistical significance is real or a function of the additive or cumulative impact of biases. In addition, meta-analyses or randomized trials, usually to ensure comparability, examine the association between one treatment option and one outcome. However, usually there are multiple treatments for the same condition and multiple clinically relevant outcomes and the meaningful question is what intervention is the best for every outcome. Single risk factors are not specific as well since many risk factors apply to several outcomes. Therefore a more comprehensive approach is needed.

An umbrella review systematically collects and evaluates information from multiple systematic reviews and meta-analyses on all clinical outcomes for which these have been performed [7]. The number of papers tagged as "umbrella review" in Pubmed from 2007 until today has increased (Fig. 1). There is a very wide range of topics covered including nutrition [8], psychiatry [9] and neurology [10], internal medicine [11] and Obstetrics and Gynecology [12]. There is also a variety of studies that can be included in the umbrella review depending on the research question: meta-analyses of observational studies examining risk factors for disease [13], interventions [14] or incorporating evidence from Mendelian Randomization studies [15].

Not all published umbrella reviews follow a standardized methodology despite the clear description of this study design several years ago [7]. Such a broad synthesis and analysis of data from many systematic reviews and

✉ Stefania Papatheodorou
spapathe@hsph.harvard.edu

1   Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

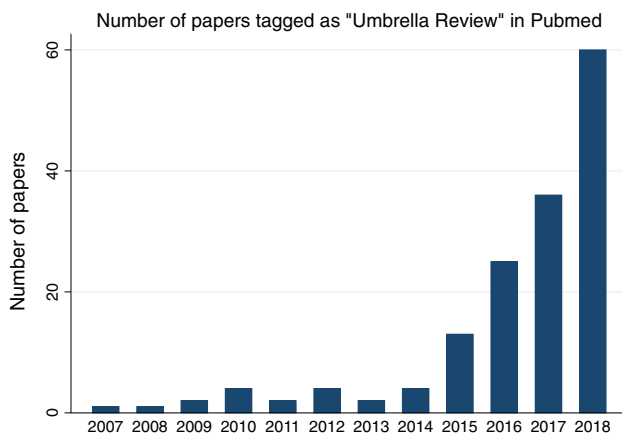2   Cyprus University of Technology, Limassol, Cyprus

Number of papers tagged as "Umbrella Review" in Pubmed

Fig. 1 Number of papers tagged as "umbrella reviews" in Pubmed

multiple meta-analyses is not simple and it requires both subject-matter experts and experienced methodologists. The basic steps on performing an umbrella review are presented below.

Methods

The need for a new umbrella review has to be identified a priori based on several factors: usually it is performed in topics that are highly controversial or when the biases that affect a certain research field have not been systematically evaluated. Fields with many meta-analyses having inconclusive evidence are clearly suitable for an umbrella review assessment, as it may shed light on the robustness of epidemiologic evidence by using a ranking approach.

Inclusion and exclusion criteria

As in all synthesis methods, the protocol must be registered in an open-access database such as PROSPERO (https://www.crd.york.ac.uk/PROSPERO/). The authors should describe clearly the type of systematic reviews and meta-analyses included in their assessment (observational studies, randomized trials, mendelian randomization studies or all of them). Certain criteria regarding the definition of the exposure/intervention and the outcome need to be determined just like in standard systematic reviews. The search strategy and the databases searched for records must be fully reported for all databases for replication purposes. In addition, the included studies need to provide the data in enough detail so as to perform the statistical analysis. The included studies

need to be evaluated in terms of their quality by using validated tools [5].

Measures of association and exposure-outcome categories

Systematic reviews and meta-analyses use different measures of association depending on the nature of the research question, the design and the analytical approach. In large-scale assessments, we can have relative and absolute measures describing the same exposure-outcome association. The use of different measures of association should not prohibit the researchers from synthesizing them in an umbrella review as long as they use the established methods for transformations [16]. That will allow a straightforward presentation and interpretation of the evidence.

In most cases, the definition of the risk factors of any nature (clinical, environmental, biomarkers and others) is very heterogeneous. One possible option is to use the definitions as presented in the primary studies without further categorization. This approach may reduce the risk of introducing newly defined factors not originally described in the literature. However, in some cases, it makes more sense to refer to categories of exposure for example in biomarkers. So instead of evaluating biomarkers one by one, we can evaluate large categories of biomarkers such as hormones, diet, inflammatory markers, IGF/insulin system [17]. The advantage of this method is that we may be able to collectively evaluate the state of the evidence in broad categories of research, which may make more sense in clinical practice than evaluating biomarkers one by one. This is not always possible and it requires very thoughtful and justified decisions from the analysts' side. The same reasoning applies in the definition and categorization of the outcomes.

Systematic reviews can be summarized by using a descriptive approach [18]. The conclusions can be categorized in the following categories: definite association, suggestive (possible) association, no association or inconclusive association (insufficient evidence).

Heterogeneity and other biases

Heterogeneity should be recorded and evaluated. In the presence of large between study heterogeneity, the results of the meta-analysis might not be applicable to any of the synthesized studies or in future studies. Publication bias and small study effects are incorporated in the grading criteria used for ranking the epidemiologic evidence. Finally, the excess-of-statistical-significance test is performed to evaluate whether there was a relative excess of formally

significant findings in the published literature for any reason as per Ioannidis and Trikalinos [19].

Grading criteria

Finally, the credibility of each proposed association is graded based on the following categories:

- Convincing evidence (Class I): Associations with a statistical significance of $P < 10-6$, more than 1000 cases included (or more than 20,000 participants for continuous outcomes), the largest component study reporting a significant result $P < 0.05$, a 95% prediction interval that excluded the null, absence of large heterogeneity $I2 < 50\%$, no evidence of small study effect $P > 0.10$, no evidence of excess significance ($P > 0.10$).
- Highly suggestive evidence (Class II): Associations with a statistical significance of $P < 10-6$, more than 1000 cases included (or more than 20,000 participants for continuous outcomes), the largest component study reporting a significant result $P < 0.05$.
- Suggestive evidence (Class III): Associations with a statistical significance of $P < 0.001$, more than 1000 cases included (or more than 20,000 participants for continuous outcomes).
- Weak evidence (Class IV): Associations with a statistical significance of $P < 0.05$.
- Not significant: Associations with $P \geq 0.05$.

It is highly recommended that authors of umbrella reviews use these criteria because this will allow an objective, standardized classification of the level of evidence. However, since the cutoffs are continuous variables, the authors must be cautious in the interpretation because including 1000 cases is not substantially different than including 999 cases, although this is a rare occasion.

Limitations

There is a large gap in the generalizability between a single patient and a population. There is an even larger gap between a single study and a meta-analysis. An additional gap exists the synthesis of many meta-analyses of interventions or observational studies and the application of this methodology to wider domains. Given this limitation and acknowledging what it means in the interpretation of the evidence, this is a more insightful approach in order to understand the strengths and limitations of the data guiding medical decisions of individual patients.

Conclusions

Umbrella reviews have the potential to provide the highest quality of evidence, if performed and interpreted properly. With almost 200 articles characterized as umbrella reviews in Pubmed and several others registered at PROSPERO database, this is a fascinating field with a great potential for large contributions in the hierarchy of evidence.

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312(7023):71–2.
2. Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q. 2016;94(3):485–514.
3. Silagy CA, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. JAMA. 2002;287(21):2831–4.
4. Moher D, A Liberati, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.
5. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008.
6. Salanti G, Ioannidis JP. Synthesis of observational studies should consider credibility ceilings. J Clin Epidemiol. 2009;62(2):115–22.
7. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ. 2009;181(8):488–93.
8. Poole R, Kennedy OJ, Roderick P, Fallowfield JA, Hayes PC, Parkes J. Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes. BMJ. 2017;359:j5024.
9. Machado MO, Veronese N, Sanches M, Stubbs B, Koyanagi A, Thompson T, et al. The association of depression and all-cause and cause-specific mortality: an umbrella review of systematic reviews and meta-analyses. BMC Med. 2018;16(1):112.
10. Bellou V, Belbasis L, Tzoulaki I, Middleton LT, Ioannidis JPA, Evangelou E. Systematic evaluation of the associations between environmental risk factors and dementia: an umbrella review of systematic reviews and meta-analyses. Alzheimer's Dement J Alzheimer's Assoc. 2017;13(4):406–18.
11. Houze B, El-Khatib H, Arbour C. Efficacy, tolerability, and safety of non-pharmacological therapies for chronic pain: an umbrella review on various CAM approaches. Prog Neuropsychopharmacol Biol Psychiatry. 2017;79(Pt B):192–205.
12. Corso E, Hind D, Beever D, Fuller G, Wilson MJ, Wrench IJ, et al. Enhanced recovery after elective caesarean: a rapid review of clinical protocols, and an umbrella review of systematic reviews. BMC Pregnancy Childbirth. 2017;17(1):91.
13. Giannakou K, Evangelou E, Papatheodorou SI. Genetic and non-genetic risk factors for pre-eclampsia: umbrella review of systematic reviews and meta-analyses of observational studies. Ultrasound Obstet Gynecol: Off J Int Soc Ultrasound Obstet Gynecol. 2018;51(6):720–30.
14. Solmi M, Correll CU, Carvalho AF, Ioannidis JPA. The role of meta-analyses and umbrella reviews in assessing the harms of

psychotropic medications: beyond qualitative synthesis. Epidemiol Psychiatric Sci. 2018;27(6):537–42.

15. Kohler CA, Evangelou E, Stubbs B, Solmi M, Veronese N, Belbasis L, et al. Mapping risk factors for depression across the lifespan: an umbrella review of evidence from meta-analyses and Mendelian randomization studies. J Psychiatry Res. 2018;103:189–207.

16. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med. 2000;19(22):3127–31.

17. Tsilidis KK, Papatheodorou SI, Evangelou E, Ioannidis JP. Evaluation of excess statistical significance in meta-analyses of 98 biomarker associations with cancer risk. J Natl Cancer Inst. 2012;104(24):1867–78.

18. Theodoratou E, Tzoulaki I, Zgaga L, Ioannidis JP. Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ. 2014;348:g2035.

19. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin Trials. 2007;4(3):245–53.