Ten simple rules for conducting umbrella reviews

Paolo Fusar-Poli,[1,2,3] Joaquim Radua[1,4,5]

[1]Early Psychosis: Interventions and Clinical-detection (EPIC) Lab, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; [2]OASIS Service, South London and Maudsley NHS Foundation Trust, London, UK; [3]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy; [4]FIDMAG Germanes Hospitalaries, CIBERSAM, Barcelona, Spain; [5]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden

Correspondence to Dr Paolo Fusar-Poli, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, London SE5 8AF, UK; paolo.fusar-poli@kcl.ac.uk

ABSTRACT

Objective Evidence syntheses such as systematic reviews and meta-analyses provide a rigorous and transparent knowledge base for translating clinical research into decisions, and thus they represent the basic unit of knowledge in medicine. Umbrella reviews are reviews of previously published systematic reviews or meta-analyses. Therefore, they represent one of the highest levels of evidence synthesis currently available, and are becoming increasingly influential in biomedical literature. However, practical guidance on how to conduct umbrella reviews is relatively limited.

Methods We present a critical educational review of published umbrella reviews, focusing on the essential practical steps required to produce robust umbrella reviews in the medical field.

Results The current manuscript discusses 10 key points to consider for conducting robust umbrella reviews. The points are: ensure that the umbrella review is really needed, prespecify the protocol, clearly define the variables of interest, estimate a common effect size, report the heterogeneity and potential biases, perform a stratification of the evidence, conduct sensitivity analyses, report transparent results, use appropriate software and acknowledge the limitations. We illustrate these points through recent examples from umbrella reviews and suggest specific practical recommendations.

Conclusions The current manuscript provides a practical guidance for conducting umbrella reviews in medical areas. Researchers, clinicians and policy makers might use the key points illustrated here to inform the planning, conduction and reporting of umbrella reviews in medicine.

Statistics in practice

INTRODUCTION

Medical knowledge traditionally differs from other domains of human culture by its progressive nature, with clear standards or criteria for identifying improvements and advances. Evidence-based synthesis methods are traditionally thought to meet these standards. They can be thought of as the basic unit of knowledge in medicine, and allow making sense of several and often contrasting findings, which is crucial to advance clinical knowledge. In fact, clinicians accessing international databases such as PubMed to find the best evidence on a determinate topic may soon feel overwhelmed with too many findings, often contradictory and not replicating each other. Some authors have argued that biomedical science[1] suffers from a serious replication crisis,[2] to the point that scientifically, replication becomes equally as—or even more—important than discovery.[3] For example, extensive research has investigated the factors that may be associated with an increased (risk factors) or decreased (protective factors) likelihood of developing serious mental disorders such as psychosis. Despite several decades of research, results have been inconclusive because published findings have been conflicting and affected by several types of biases.[4] Systematic reviews and meta-analyses aim to synthesise the findings and investigate the biases. However, as the number of reviews of meta-analyses also increased, clinicians may also feel overwhelmed with too many of them.

Umbrella reviews have been developed to overcome such a gap of knowledge. They are reviews of previously published systematic reviews or meta-analyses, and consist in the repetition of the meta-analyses following a uniform approach for all factors to allow their comparison.[5] Therefore, they represent one of the highest levels of evidence synthesis currently available (figure 1). Not surprisingly, umbrella reviews are becoming increasingly influential in biomedical literature. This is empirically confirmed by the proliferation of this type of studies over the recent years. In fact, by searching 'umbrella review' in the titles of articles published on Web of Knowledge (up to 1 April 2018), we found a substantial increase in the number of umbrella reviews published over the past decade, as detailed in figure 2. The umbrella reviews identified through our literature search were investigating a wide portion of medical branches (figure 3). Furthermore, several protocols of upcoming umbrella reviews have been recently published, confirming the exponential trend.[6–12]

However, guidance on how to conduct or report umbrella reviews is relatively limited.[5] The current manuscript addresses this area by providing practical tips for conducting good umbrella reviews in medical areas. Rather than being an exhaustive primer on the methodological underpinning of umbrella reviews, we only highlight 10 key points that to our opinion are essential for conducting robust umbrella reviews. As reference example, we will use an umbrella review on risk and protective factors for psychotic disorders recently completed by our group.[13] However, we generalise the considerations presented in this manuscript and the relative recommendations to any other area of medical knowledge.

METHODS

Educational and critical (non-systematic) review of the literature focusing on key practical issues that are necessary for conducting and reporting robust umbrella reviews. The authors selected illustrative umbrella reviews to highlight key methodological findings. In the results, we present 10 simple key points that the authors of umbrella reviews should carefully address when planning and conducting umbrella reviews in the medical field.

RESULTS

Ensure that the umbrella review is really needed

The decision to develop a new umbrella review in medical areas of knowledge should be stimulated by several factors, e.g. the topic of interest may be highly controversial or it may be affected by potential biases that have not been investigated systematically. The authors can explore these issues in the existing literature. For example, they may want to survey and identify a few examples of meta-analyses on the same topic that present contrasting findings. Second, a clear link between the need to address uncertainty and advancing clinical knowledge should be identified a priori, and acknowledged as the strong rationale for conducting an umbrella review. For example, in our previous work we speculated that by clarifying the evidence for an association between risk or protective
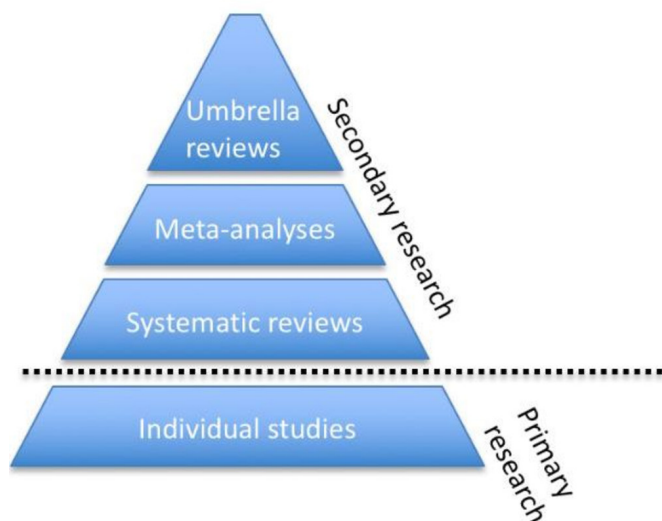
Figure 1 Hierarchy of evidence synthesis methods.

factors and psychotic disorders we could improve our ability to identify those individuals at risk of developing psychosis.[13] Clearly, improving the detection of individuals at risk is the first step towards the implementation of preventive approaches, which are becoming a cornerstone of medicine.[14–16] Third, provided that the two points are satisfactory, it is essential to check whether there are enough meta-analyses that address a determinate topic.[17] Larger databases can increase the statistical power and therefore improve accurateness of the estimates and interpretability of the results. Furthermore, they are also likely to reflect a topic of wider interest and impact for clinical practice. These considerations are of particular relevance when considering the mass production of useless evidence synthesis studies that are redundant, not necessary and addressing clinically irrelevant outcomes.[18]

Prespecify the protocol

As for any other evidence synthesis approach, it is essential to prepare a study protocol ahead of initiating the work and upload it to international databases such as PROSPERO (https://www.crd.york.ac.uk/PROSPERO/). The authors may also publish the protocol in an open-access journal, as it is common for randomised controlled trials. The protocol should clearly define the methods for reviewing the literature and extracting data and the statistical analysis plan. Importantly, specific inclusion and exclusion
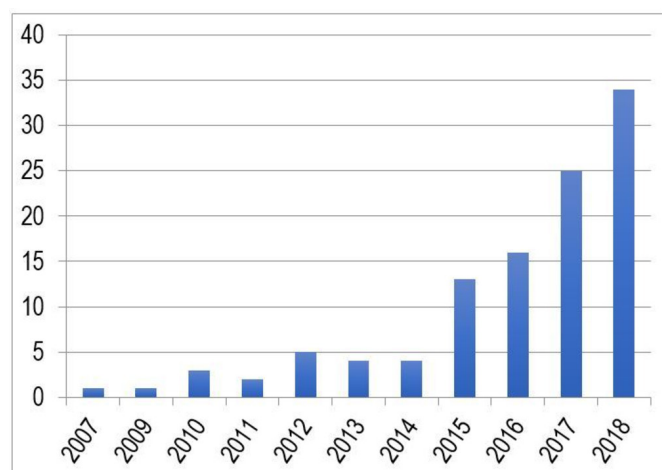

Figure 2 Web of Knowledge records containing 'umbrella review' in their title up to April 2018.

criteria should be prespecified. For example, inclusion criteria from our umbrella review[13] were: (a) systematic reviews or meta-analyses of individual observational studies (case-control, cohort, cross-sectional and ecological studies) that examined the association between risk or protective factors and psychotic disorders; (b) studies considering any established diagnosis of non-organic psychotic disorders defined by the International Classification of Disease (ICD) or the Diagnostic and Statistical Manual of Mental Disorders (DSM); (c) inclusion of a comparison group of non-psychotic healthy controls, as defined by each study and (d) studies reporting enough data to perform the analyses. Similarly, the reporting of the literature search should adhere to the Preferred Reporting Items for Systematic Reviews and Meta-analyses recommendations[19] and additional specific guidelines depending on the nature of the studies included (eg, in case of observational studies, the Meta-analysis of Observational Studies in Epidemiology guidelines[20]). Quality assessment of the included studies is traditionally required in evidence synthesis studies. In the absence of specific guidelines for quality assessment in umbrella reviews, A Measurement Tool to Assess Systematic Reviews, a validated instrument[21 22] can be used.

Clearly define the variables of interest

Umbrella reviews are traditionally conducted to measure the association between certain factors and a determinate clinical outcome. The first relevant point to conducting a good umbrella review is therefore to define consistent and reliable factors and outcomes to be analysed.

The definition of the type of factor (eg, risk factor or biomarker) of interest may be particularly challenging. For example, in our review we found that childhood trauma was considered as a common risk factor for psychosis,[13] but available literature lacked standard operationalisation. Our pragmatic approach was to define the factors as each meta-analysis or systematic review had defined them. Another issue relates to whether and how analysts should group similar factors. For example, in our umbrella review[13] we wondered whether to merge the first-generation immigrant and second-generation immigrant risk factors for psychosis in a unique category of 'immigrants'. However, this would have introduced newly defined categories of risk factors that were not available in the underlying literature. Our solution was not to combine similar factors if the meta-analysis or systematic review had considered and analysed them separately. Similarly, it may be important not to split categories into subgroups (eg, childhood sexual abuse, emotional neglect, physical abuse) if the meta-analysis or systematic review had considered them as a whole (eg, childhood trauma). Restricting the analyses to only the factors that each individual meta-analysis or systematic review had originally introduced may mitigate the risk of introducing newly defined factors not originally present in the literature. Such an approach is also advantageous to minimise the risk of artificially inflating the sample size by creating large and unpublished factors, therefore biassing the hierarchical classification of the evidence. The additional problem may be that a meta-analysis or a systematic review could report both results, that is, pooled across categories and divided according to specific subgroups. In this case, it is important to define a priori what kind of results is to be used. Pooled results may be preferred since they usually include larger sample sizes. Finally, there may be two meta-analyses or systematic reviews that address the same factor or that include individual studies that are overlapping. In our previous umbrella review, we selected the meta-analysis or systematic review with the largest database and the most recent one.[13]

A collateral challenge in this domain may relate to the type of factors that analysts should exclude. For example, in our previous umbrella review[13] we decided to focus on risk and protective factors for psychosis only, and not on biomarkers. However, in the lack of clear etiopathogenic mechanisms for the onset of psychosis, the boundaries between biomarkers collected before the onset of the disorder and risk and
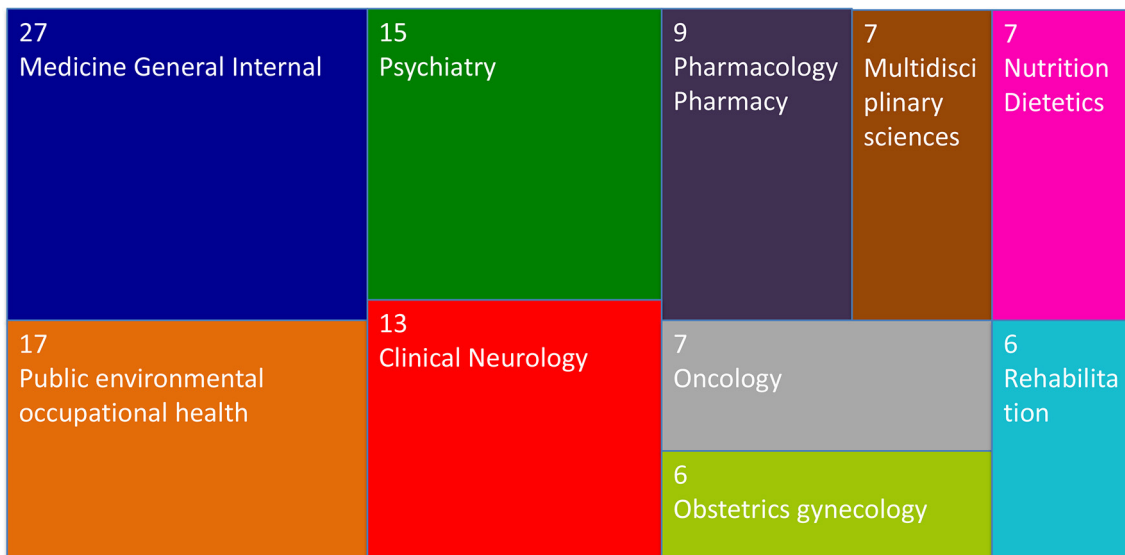
| 27 Medicine General Internal | 15 Psychiatry | 9 Pharmacology Pharmacy | 7 Multidisciplinary sciences | 7 Nutrition Dietetics |
| 17 Public environmental occupational health | 13 Clinical Neurology | 7 Oncology / 6 Obstetrics gynecology | | 6 Rehabilitation |

Figure 3 Focus of umbrella reviews published in Web of Knowledge (see figure 2—up to April 2018).

protective factors were not always clear. To solve this problem, we have again adopted a pragmatic approach, adopting the definitions of risk and protective factors versus biomarker as provided by each article included in the umbrella review. A further point is that if systematic reviews are included, some of them may not have performed quantitative data on specific factors.

The additional challenge would be that individual meta-analyses or systematic reviews might have similar but not identical definition of these outcomes. For example, we intended to investigate only psychotic disorders defined by standard international validated diagnostic manuals such as the ICD or the DSM. We found that some meta-analyses that were apparently investigating psychotic disorders in reality did also include studies that were measuring psychotic symptoms not officially coded in these manuals.[13] To overcome this problem, we took the decision to verify the same inclusion and exclusion criteria that were used for reviewing the literature (eg, inclusion of DSM or ICD psychotic disorders) for each individual study included in every eligible meta-analysis or systematic review.[13] Such a process is extremely time-consuming and analysts should account for it during the early planning stages to ensure sufficient resources are in place. The authors of an umbrella review may also exclusively rely on the information provided in the systematic reviews and meta-analyses, although in that case the analysts should clearly acknowledge its potential limitations in the text. Alternatively, they may rely on the systematic reviews and meta-analyses to conduct a preselection of the factors with a greater level of evidence, and then verify the data for each individual study of these (much fewer) factors.

Estimate a common effect size
The systematic reviews and meta-analyses use different measures of effect size depending on the design and analytical approach of the studies that they review. For example, meta-analyses of case-control studies may use standardised mean differences such as the Hedge's *g* to compare continuous variables, and odd ratios (ORs) to compare binary variables. Similarly, meta-analyses of cohort studies comparing incidences between exposed and non-exposed may use a ratio of incidences such as the incidence rate ratio (IRR). In addition, other measures of effect size are possible. The use of these different measures of effect size is enriching because each of them is appropriate for a type of studies, and thus we recommend also using them in the umbrella review. For example, a hazard ratio (HR) may be very appropriate for summarising a survival analysis, while it would be hard to interpret in a cross-sectional

study, ultimately preventing the readers from easily getting a glimpse of the current evidence.

However, one of the main aims of an umbrella is also to allow a comparison of the size of the effects across all factors investigated, and the use of a common effect size for all factors clearly makes this comparison straightforward. For example, in our previous umbrella review of risk and protective factors for psychosis, we found that the effect size of parental communication deviance (a vague, fragmented and contradictory intrafamilial communication) was Hedge's $g = 1.35$, whereas the effect size of heavy cannabis use was $OR = 5.17$.[13] Which of these factors had a larger effect? To allow a straightforward comparison, we converted all effect sizes to OR, and the equivalent OR of parental communication deviance was 11.55. Thus, reporting an equivalent OR for each factor, the readers can straightforwardly compare the factors and conclude that the effect size of parental communication deviance is substantially larger than the effect size of heavy cannabis use. To further facilitate the comparison of factors, the analysts may even force all equivalent OR to be greater than 1 (ie, inverting any OR < 1). For example, in our previous umbrella review, we found that the equivalent OR of self-directedness was 0.17.[13] The inversion of this OR would be 5.72, which the reader could straightforwardly compare with other equivalent ORs > 1.

An exact conversion of an effect size into an equivalent OR may not always be possible, because the measures of effect size may be inherently different and the calculations may need data that may be unavailable. For example, to convert an IRR into an OR, the analysts should first convert the IRR into a risk ratio (RR), and then the RR into an OR. However, an IRR and a RR have an important difference: the IRR accounts for the time that the researchers could follow each individual, while the RR only considers the initial sizes of the samples. In addition, even if the analysts could convert the IRR into a RR, they could not convert the RR into an OR without knowing the incidence in the non-exposed, which the papers may not report.

Fortunately, approximate conversions are relatively straightforward[23] (table 1). On the one hand, the analysts may assume that HRs, IRRs, RRs and ORs are approximately equivalent as far as the incidence is not too large. Similarly, they may also assume that Cohen's *d*, Glass'∆ and Hedge's *g* are approximately equivalent as far as the variances in patients and controls are not too different and the sample sizes are not too small. On the other hand, the analysts can convert Hedge's *g* into equivalent OR using a standard formula.[23] For other measures such as the risk difference, the ratio of means or the mean difference, the analyst will need

Table 1 Possible conversions of some effect sizes to equivalent ORs

Conversion	Justification
IRR to RR	The following formula, straightforwardly derived from the definitions of incidence rate ratio (IRR) and risk ratio (RR), converts the former into the latter: $$RR = \frac{average\left(time_{exposed}\right)}{average\left(time_{non\text{-}exposed}\right)} \times IRR$$ Fortunately, if the incidences are small enough, the average follow-up times are similar in exposed and non-exposed, the fraction in the left is approximately 1 and thus: $$RR \approx IRR$$
RD to RR	The following formula, straightforwardly derived from the definitions of risk difference (RD) and RR, converts the former into the latter: $$RR = 1 + \frac{1}{p_{non\text{-}exposed}} \times RD$$ Thus, analysts might need an estimation of the probability of developing the disease (p) in the non-exposed.
RR to OR	The following formula, straightforwardly derived from the definitions of RR and OR, converts the former into the latter: $$OR = \frac{1 - p_{non\text{-}exposed}}{1 - p_{exposed}} \times RR$$ Fortunately, if the probabilities of developing the disease (p) are small enough, the fraction in the left is approximately 1, and thus: $$OR \approx RR$$
RoM to MD	The following formula, straightforwardly derived from the definitions of ratio of means (RoM) and mean difference (MD), converts the former into the latter: $$MD = m_{controls} \times \left(RoM - 1\right)$$ Thus, analysts might need an estimation of the mean (m) in controls.
MD to Glass' Δ	The following formula, straightforwardly derived from the definitions of mean difference (MD) and Glass' Δ, converts the former into the latter: $$\Delta = \frac{1}{s_{controls}} \times MD$$ Thus, analysts might need an estimation of the SD (s) in controls.
Glass' Δ to Cohen's d	The following formula, straightforwardly derived from the definitions of Glass' Δ and Cohen's d, converts the former into the latter: $$d = \sqrt{\frac{(n_{controls}-1)\cdot s^2_{controls} + (n_{cases}-1)\cdot s^2_{controls}}{(n_{controls}-1)\cdot s^2_{controls} + (n_{cases}-1)\cdot s^2_{cases}}} \times \Delta$$ Fortunately, if the variances (s^2) in cases and controls are similar enough, the square root in the left is approximately 1, and thus: $$d \approx \Delta$$
Hedge's g to Cohen's d	The following formula, straightforwardly derived from the definitions of Hedge's g and Cohen's d, converts the former into the latter: $$d = \frac{1}{J(df)} \times g$$ Fortunately, if the sample sizes are large enough, the small-sample correction factor (J) is approximately 1, the fraction in the left is approximately 1 and thus: $$d \approx g$$
Pearson's r to Cohen's d	The following standard formula[23] converts a Pearson's r into an approximate Cohen's d: $$d \approx \frac{2 \cdot r}{\sqrt{1 - r^2}}$$

a few general estimations (table 1). In any case, such approximations are acceptable because the only aim of the equivalent OR is to provide a visual number to allow an easy comparison of the effect sizes of the different factors.

Report the heterogeneity and potential biases
As with single meta-analyses, an umbrella review should study and report the heterogeneity across the studies included in each meta-analysis and the potential biases in the studies to show a more complete picture of the evidence. Independently of the effect size and the p value, the level of evidence of an effect (eg, a risk factor) is lower when there is large heterogeneity, as well as when there is potential reporting or excess significance bias. The presence of a large between-study heterogeneity may indicate, for example, that there are two groups of studies investigating two different groups of patients, and the results of a single meta-analysis for the two groups may not represent either of the groups. The presence of potential reporting bias, on the other hand, might mean that studies are only published timely in indexed journals if they find one type of results, for example, if they find that a given psychotherapy works. Of course, if the meta-analysis only includes these studies, the results will be that the psychotherapy works, even if it does not. Analysts can explore the reporting bias that affects the smallest studies with a number of tools, such as the funnel plot, Egger and similar tests and trim and fill methods.[4] Finally, the presence of potential excess significance bias would mean that the number of studies with statistically significant results is suspiciously high, and this may be related to reporting bias and to other issues such as data dredging.[4]

Perform a stratification of evidence
A more detailed analysis of the umbrella reviews identified in our literature search revealed that some of them, pertaining to several clinical medical areas (neurology, oncology, nutrition medicine, internal medicine, psychiatry, paediatrics, dermatology and neurosurgery) additionally stratified the evidence using a classification method. This classification was obtained through strict criteria, equal or similar to the one listed below[24–26]:

► convincing (class I) when number of cases >1000, $p<10^{-6}$, $I^2<50\%$, 95% prediction interval excluding the null, no small-study effects and no excess significance bias;
► highly suggestive (class II) when number of cases >1000, $p<10^{-6}$, largest study with a statistically significant effect and class I criteria not met;
► suggestive (class III) when number of cases >1000, $p<10^{-3}$ and class I–II criteria not met;
► weak (class IV) when $p<0.05$ and class I–III criteria not met;
► non-significant when $p>0.05$.

We strongly recommend the use of these or similar criteria because they allow an objective, standardised classification of the level of evidence. However, the analysts should not forget that the variables used in these criteria are continuous and the set of cut-off points are only cut-off points. For example, the difference between a factor that includes 1000 patients and a factor that includes 1001 patients is negligible, but according to the criteria, the former can only be class IV, whereas the latter could be class I.

Table 2	Recommended elements in the summary tables or plots
Description	Name of the factor (eg, 'parental communication deviance'), optionally with a very few details such as a brief literally description of the factor, the number of studies, the number patients and controls, the time of follow-up, whether the studies are prospective, etc.
Effect size	The measure of effect size (eg, 'Hedge's g'), the outcome and its CI (eg, '1.35 (95% CI 0.97 to 1.73)') and the equivalent OR (eg, '11.55').
Features used for classification of evidence	These statistics may vary depending on the criteria used, and could be: the p value, the fraction of variance that is due to heterogeneity (I^2), the prediction interval, the CI of the largest study, the results of an Egger test (or another method to assess potential small study reporting bias) and the results of an excess significance test. To simplify the table, the authors of the umbrella review may choose to report only the bits that are most important, eg, the results of an Egger test could be simply 'yes' or 'no'.
Evidence class	Based on the features above.
Other statistics	Some statistics not used for the classification of evidence but that are of great interest, although we suggest that the amount of information presented is limited to ensure that the table is simple enough.

Conduct (study-level) sensitivity analyses

Depending on the type of umbrella review (eg, risk or protective factors, biomarkers etc), a few sensitivity analyses may enrich the final picture. For instance, in an umbrella review of potential risk and protective factors, establishing the temporality of the association is critical in order to minimise reverse causation. This may be seen in scenarios similar to the following example: many smokers quit smoking after developing lung cancer, and thus a cross-sectional study could report that the prevalence of lung cancer is higher in ex-smokers than in smokers, and erroneously conclude that quitting smoking causes lung cancer. To avoid this reverse causation, studies must address the temporality, that is, observing that patients first developed lung cancer and after quit smoking, rather than the other way round. In an umbrella review, the analysts may address temporality with a sensitivity analysis that includes only prospective studies. Our recent umbrella review provides an example of sensitivity analyses investigating temporality of association.[13]

Report transparent results

An umbrella review generates a wealth of interesting data, but the analysts should present them adequately in order to achieve one of the main aims: to summarise clearly the evidence. This is not always straightforward. They may design tables or plots that report all information of interest in a simplified way. One approach could be, for example, including a table with the effect size (and its CI), the equivalent OR, the features used for the classification of the level of evidence and the resulting evidence. Parts of this table could be graphical, for example, the analysts may choose to present the equivalent OR as a forest plot. In any case, the readers should be able to know easily the effect size and the degree of evidence of the factors from the tables and plots. Table 2 shows a summary of the key statistics that we suggest to report in any umbrella review.

Use appropriate software

The analysts can conduct a large part of the calculations of an umbrella review with usual meta-analytical software, such as 'meta', 'metafor' or 'metansue' packages for R.[27–29] Some software includes better estimation methods of the between-study heterogeneity than others,[30–32] but this may probably represent a minor difference. That said, we recommend that the meta-analytical software is complete enough to fit random-effects models, assess between-study heterogeneity, estimate prediction intervals and assess potential reporting bias.

However, even if using good meta-analytic software, the analysts will still have to write the code for some parts of the umbrella review. On the one hand, some specific computations may not be available in standard software, such as the estimation of the statistical power in some studies, required to evaluate excess significance bias. On the other hand, meta-analytic software aims to conduct and show the results of one meta-analysis, whereas an umbrella review may include hundreds of meta-analyses, for what the analysts will have to manage and show the results of all these meta-analyses as an integrated set. For example, to create the forest plots, the analysts may write a code that takes the results of the different meta-analyses as if they were individual studies, and then calls the forest plot function of the meta-analytic software (without displaying a pooled effect). We are developing new and free umbrella review software to minimise these burdens.

Acknowledge its limitations

To report transparently the evidence, the analysts must adequately acknowledge the limitations of the umbrella review. Some limitations may be specific of a given umbrella, and others are relatively general. Among them, probably one of the most important issues is that umbrella reviews can only report what researchers have investigated, published and systematically reviewed or meta-analysed. For example, a factor may have an amazingly strong effect, but if few studies have investigated the factor, it will probably be classified as only class IV evidence because of involving <1000 patients. Indeed, if the factor was not part of any systematic review or meta-analysis, it would not be even included in the umbrella review. Fortunately, given the mass production of evidence synthesis studies it is also unlikely that a relevant area of medical knowledge is not addressed by any published systematic review or meta-analysis.[18] On the other hand, an umbrella could include all studies published, beyond those included in published reviews, but this would require updating the literature search at the level of each subdomain included in the umbrella review. This extra work would highly increase the already very high working time needed to conduct an umbrella review, to the level that most umbrella reviews could become unfeasible. Furthermore, it would probably involve the definition of new subgroups or factors that the systematic review or meta-analysis had not originally reported, making the interpretation of the final findings more difficult. Another issue is that the use of a systematic approach analysis would not allow conducting the rigorous assessment of several types of biases. Finally, a similar limitation is that the umbrella review will have most of the limitations of the included studies. For instance, if the latter assess association but not causation, the umbrella review will assess association but not causation.

CONCLUSIONS

Umbrella reviews are becoming widely used as a means to provide one of the highest levels of evidence in medical knowledge. Key points to be considered to conducting robust umbrella reviews are to ensure that they are really needed, prespecify the protocol, clearly define the variables of interest, estimate a common effect size, report the heterogeneity and potential biases, perform a stratification of the evidence, conduct (study-level) sensitivity analyses, report transparent results, use appropriate software and acknowledge the limitations.

<div style="writing-mode: vertical">**Statistics in practice**</div>

REFERENCES

1. **Ioannidis JP,** Khoury MJ, Km J. Improving validation practices in "omics" research. *Science* 2011;**334**:1230–2.
2. **Szucs D,** Ioannidis JP, Ijp A. Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biol* 2017;**15**:e2000797.
3. **Ioannidis JP.** Evolution and translation of research findings: from bench to where? *PLoS Clin Trials* 2006;**1**:e36.
4. **Ioannidis JP,** Munafò MR, Fusar-Poli P, *et al*. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. *Trends Cogn Sci* 2014;**18**:235–41.
5. **Ioannidis JP.** Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *CMAJ* 2009;**181**:488–93.
6. **Elliott J,** Kelly SE, Bai Z, *et al*. Optimal duration of dual antiplatelet therapy following percutaneous coronary intervention: protocol for an umbrella review. *BMJ Open* 2017;**7**:e015421.
7. **Catalá-López F,** Hutton B, Driver JA, *et al*. Cancer and central nervous system disorders: protocol for an umbrella review of systematic reviews and updated meta-analyses of observational studies. *Syst Rev* 2017;**6**:69.
8. **Dinsdale S,** Azevedo LB, Shucksmith J, *et al*. Effectiveness of weight management, smoking cessation and alcohol reduction interventions in changing behaviors during pregnancy: an umbrella review protocol. *JBI Database System Rev Implement Rep* 2016;**14**:29–47.
9. **Jadczak AD,** Makwana N, Luscombe-Marsh ND, *et al*. Effectiveness of exercise interventions on physical function in community-dwelling frail older people: an umbrella review protocol. *JBI Database System Rev Implement Rep* 2016;**14**:93–102.
10. **Chai LK,** Burrows T, May C, *et al*. Effectiveness of family-based weight management interventions in childhood obesity: an umbrella review protocol. *JBI Database System Rev Implement Rep* 2016;**14**:32–9.
11. **Thomson K,** Bambra C, McNamara C, *et al*. The effects of public health policies on population health and health inequalities in European welfare states: protocol for an umbrella review. *Syst Rev* 2016;**5**:57.
12. **Krause D,** Roupas P. Dietary interventions as a neuroprotective therapy for the delay of the onset of cognitive decline in older adults: an umbrella review protocol. *JBI Database System Rev Implement Rep* 2015;**13**:74–83.
13. **Radua J,** Ramella-Cravaro V, Ioannidis JPA, *et al*. What causes psychosis? An umbrella review of risk and protective factors. *World Psychiatry* 2018;**17**:49–66.
14. **Schmidt A,** Cappucciati M, Radua J, *et al*. Improving prognostic accuracy in subjects at clinical high risk for psychosis: systematic review of predictive models and meta-analytical sequential testing simulation. *Schizophr Bull* 2017;**43**:375–88.
15. **Fusar-Poli P,** Cappucciati M, Bonoldi I, *et al*. Prognosis of brief psychotic episodes: a meta-analysis. *JAMA Psychiatry* 2016;**73**:211–20.
16. **Fusar-Poli P,** Cappucciati M, Borgwardt S, *et al*. Heterogeneity of psychosis risk within individuals at clinical high risk: a meta-analytical stratification. *JAMA Psychiatry* 2016;**73**:113–20.
17. **Siontis KC,** Hernandez-Boussard T, Ioannidis JP. Overlapping meta-analyses on the same topic: survey of published studies. *BMJ* 2013;**347**:f4501.
18. **Ioannidis JP.** The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q* 2016;**94**:485–514.
19. **Moher D,** Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;**62**:1006–12.
20. **Stroup DF,** Berlin JA, Morton SC, *et al*. Meta-analysis of observational studies in epidemiology - a proposal for reporting. *JAMA* 2000;**283**:2008–12.
21. **Shea BJ,** Grimshaw JM, Wells GA, *et al*. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;**7**:10.
22. **Shea BJ,** Bouter LM, Peterson J, *et al*. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One* 2007;**2**:5.
23. **Chinn S.** A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;**19**:3127–31.
24. **Bellou V,** Belbasis L, Tzoulaki I, *et al*. Systematic evaluation of the associations between environmental risk factors and dementia: An umbrella review of systematic reviews and meta-analyses. *Alzheimers Dement* 2017;**13**.
25. **Bellou V,** Belbasis L, Tzoulaki I, *et al*. Environmental risk factors and Parkinson's disease: an umbrella review of meta-analyses. *Parkinsonism Relat Disord* 2016;**23**:1–9.
26. **Belbasis L,** Bellou V, Evangelou E, *et al*. Environmental risk factors and multiple sclerosis: an umbrella review of systematic reviews and meta-analyses. *Lancet Neurol* 2015;**14**:263–73.
27. **Schwarzer G.** meta: an R package for meta-analysis. *R News* 2007;**7**:40–5.
28. **Viechtbauer W.** Conducting meta-analyses in R with the metafor package. *J Stat Softw* 2010;**36**:1–48.
29. **Radua J,** Schmidt A, Borgwardt S, *et al*. Ventral striatal activation during reward processing in psychosis: a neurofunctional meta-analysis. *JAMA Psychiatry* 2015;**72**:1243–51.
30. **Viechtbauer W.** Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* 2005;**30**:261–93.
31. **Veroniki AA,** Jackson D, Viechtbauer W, *et al*. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods* 2016;**7**:55–79.
32. **Petropoulou M,** Mavridis D. A comparison of 20 heterogeneity variance estimators in statistical synthesis of results from studies: a simulation study. *Stat Med* 2017;**36**:4266–80.