Opinion

Temporal instability of evidence base: a threat to policy making?

Julia Koricheva[1] and Elena Kulinskaya[2]

[1]School of Biological Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK

[2]School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Corresponding author: Julia Koricheva (julia.koricheva@rhul.ac.uk)

Keywords: decline effect, evidence-based conservation, meta-analysis, publication bias

Abstract

A shift towards evidence-based conservation and environmental management over the last two decades has resulted in an increased use of systematic reviews and meta-analyses as tools to combine the existing scientific evidence. However, to guide policy making decisions in conservation and management the conclusions of meta-analyses need to remain stable for at least some years. Alarmingly, numerous recent studies indicate that the magnitude, statistical significance and even the sign of the effects reported in the literature might change over relatively short time periods. We argue that such rapid temporal changes in cumulative evidence represent a real threat to policy making in conservation and environmental management and call for systematic monitoring of temporal changes in evidence and exploration of their causes.

Temporal changes in cumulative evidence

In their seminal paper published in *Trends in Ecology and Evolution* 15 years ago, Sutherland et al. [1] called for conservation and environmental management to become evidence-based and proposed that support for decision making in conservation could benefit from the production of **systematic reviews** (see Glossary) including **meta-analyses** of published evidence of effectiveness of interventions [2]. Guidelines for systematic review in conservation and environmental management have been developed soon after [3] and over 600 meta-analyses on conservation topics were published to date providing assessment of the effectiveness of different conservation and management strategies [4-6]. However, the conduct of systematic review and meta-analysis provides just a snapshot of the available evidence at a more or less arbitrary point in time whereas scientific evidence is not static and tends to change over time as more research on the topic accumulates [7]. New studies may either strengthen or challenge the conclusions of previous reports. If the above changes in cumulative evidence over time are rapid and of considerable magnitude, the conclusions of meta-analysis will strongly depend on when the review was conducted and the policy-relevant recommendations derived from these reviews will quickly go out of date.

Worryingly, a growing number of studies demonstrates that substantial changes in the magnitude, statistical significance or even sign of the reported effects over time are common in ecology and evolutionary biology [8-13] as well as other disciplines [14-17]. In most cases decreases in the magnitude of the estimated effect are reported over time, a phenomenon which has been dubbed ‘a **decline effect**’ in some fields [18]. As a result, the conclusions of systematic reviews and meta-analyses may go out of date very rapidly as well. For instance, a survey of 100 meta-analyses in medicine showed that clinically important evidence that alters review conclusions about the effectiveness and harms of treatments can accumulate within relatively short time frames, i.e. 2-5 years [19]. While no similar surveys have been conducted in ecology and evolution, meta-analyses in these fields are often performed on topics where results of studies are contradictory, sample sizes are low, and the expected magnitudes of the effects are relatively small [20]. This makes temporal changes in cumulative evidence more likely. The failure of later studies to reproduce the results of the earlier studies exemplifies a broader concern about the reproducibility in science [21].

Despite its obvious scientific and practical importance, temporal changes in evidence base for conservation and environmental management have received little attention so far [7]. In this Opinion piece we review possible causes of such temporal trends, draw attention towards their potential implications for policy making and **evidence-based conservation**, and discuss the methods of detection of temporal changes. We argue that rapid temporal changes in cumulative evidence represent a real threat to policy making in conservation and environmental management and call for systematic exploration of their extent and causes in applied ecology.

Causes of temporal instability of the evidence base

Temporal changes in reported effects may occur for three main reasons. First, temporal trends may reflect true changes in the magnitude or direction of a biological effect, e.g. due to shifts in the strength and relative importance of the drivers of biodiversity loss [22-24] and to rapid adaptive evolution [25]. A well-known example in medicine is the development of antibiotic resistance which might decrease treatment efficacy over time [26]. Similar adaptive responses may occur in ecological and evolutionary studies as a result of selection pressure imposed by humans directly or indirectly. Examples of such changes include reductions in body size in animals as a result of warming temperatures [27-29] and shifting song frequencies in birds in response to anthropogenic noise [30]. As the above selection pressures increase over time, it is likely that studies published few decades ago would report smaller effects compared to the more recent studies.

Second, temporal trends in estimated **effect sizes** may occur even when the true effect size remains the same, but the proportion of studies with particular characteristics which influence the magnitude and direction of the effect (known as **moderators** in meta-analysis) changes over time. An example of such **evidence reversal** is discussed in Box 1. If there is significant **heterogeneity** in effect sizes (i.e. not all studies share the same effect) and effects are smaller or larger under particular conditions, any changes in frequency of studies on the above condition over time relative to other conditions may result in corresponding temporal changes in the magnitude of the overall estimated effect (Box 1, [11, 31]). Changes in prevalence of particular research or statistical methods over time may also result in similar effects if such methods differ in the magnitude of the estimated effects that they produce [32, 33]. It is therefore crucial to examine the amount of heterogeneity and its causes in a meta-analysis, particularly as high heterogeneity should be expected in ecological and evolutionary studies [34].

Third, changes in magnitude and significance of the effect size estimates over time may be due to biases. Here, again, the true magnitude of the effect size might not change with time, but the estimate of the effect does. For instance, time lag in the publication of studies with non-significant results may lead to decrease in the cumulative effect over time as the number of studies with weak and non-significant effects increases. Jennions and Møller [9] suggested that such **time-lag bias** against non-significant results is the most probable cause of the observed decrease in estimated effect sizes with time in ecological and evolutionary meta-analyses. However, no studies so far have explored the relative importance of different causes of temporal trends in reported effect sizes in ecology and evolution. On the other hand, **publication bias** may also lead to overestimation of the overall effect. Nuijten et al. [35] showed that if both the original study and its conceptual replication are subject to publication bias, combining the two studies to obtain an overall effect size will result in an overestimation of the population effect size. Biases may also prevent the cumulative effects from reaching statistical significance. For instance, the attractiveness of contradictory findings to researchers and editors may lead to publication of the succession of extreme positive and negative effects, hence hindering the stabilization of the cumulative effect size over time [36, 37]. Heleno [37] argued that the consequences of the “editorial love of controversy” may be particularly severe in conservation-led decisions and might contribute to an underestimation of the impacts of human pressure on the environment. Other biases which may lead to temporal changes in cumulative evidence include bias in choice of study organisms [12] and paradigm shifts [38].

It is important to distinguish between the above causes of temporal changes in reported effects because they determine whether the current conservation or management policy needs to be modified. If true biological effects are changing over time, then actions might need to be taken to re-evaluate conservation status and conservation strategy for the given species or environmental management options might need to be reconsidered. On the other hand, if temporal changes in estimated effect sizes are due to heterogeneity among studies, the sources of this heterogeneity have to be identified to find out under what conditions the proposed management and conservation strategies are effective. Examination of temporal trends in effect sizes is thus a good diagnostic tool for detection of sources of heterogeneity. Finally, testing for presence of biases in a meta-analysis is absolutely essential, although it might be sometimes difficult to distinguish them from true heterogeneity [39].

Potential implications of temporal changes in estimated effect sizes

The magnitude and direction of the mean effect size and the breadth of its confidence interval largely determine the conclusions drawn from a meta-analysis [4]. If the magnitude, statistical significance or the direction of the estimated effect changes over time, any policy recommendations derived from a meta-analysis are likely to change as well. In Box 1 we show how two meta-analyses on the same topic conducted several years apart reached opposite conclusions on effectiveness of the same conservation measure. Such reversals in conclusions of meta-analyses represent an example of evidence reversal, a phenomenon that has only recently became a topic of formal exploration [40]. Reversals of evidence can have significant impacts on evidence-based conservation and environmental management and might necessitate revision of already implemented policies based on recommendations from the previous meta-analysis.

Moreover, evidence reversals may affect not only the effectiveness of the currently implemented policies and measures, but also the society’s and researcher’s faith in the approach to assessment of scientific evidence base. For instance, differences in the conclusions between several meta-analyses on the same topic have sometimes led to questioning whether meta-analyses constitute repeatable science [41]. While the results of two meta-analyses can differ for many other reasons (e.g. different inclusion criteria, different statistical models and moderators tested), at present we do not know what proportion of ecological meta-analyses on the same topic arrived to different conclusions because of temporal changes in the estimated effect sizes.

Conversely, a lack of temporal changes in the estimated effect sizes may also convey important policy information when changes in effectiveness over time are expected. For instance, agri-environment schemes (AES) in Europe have been used for ca 25 years and are the biggest conservation expenditure in Europe [42]. National AES programs are revised every 7 years allowing countries to use novel scientific insights and modify their agri-environmental programs to increase their efficiency. However, a meta-analysis by Batáry et al. [42] showed that effectiveness of AES has not changed as a result of the revision of the EU's agri-environmental programmes in 2007. The authors point out that this lack of increase in effectiveness over time is worrying in view of forthcoming reductions in AES budget as it is unlikely that increased effectiveness of the scheme will compensate for the future budget cuts.

Testing for temporal trends and updating the results of systematic reviews and meta-analyses

Several relatively simple and straightforward statistical approaches which allow testing for temporal trends in estimated effect sizes are available (reviewed in [7, 43, 44], and Box 2), but are unfortunately seldom used by ecologists. For instance, only 5% of 322 meta-analyses in plant ecology published between 1996 and 2013 have tested for temporal changes in estimated effects [45]. We argue that such tests have to become a routine part of ecological meta-analyses and one of the important criteria for review quality control evaluation [46]. Temporal trends in estimated effects can be detected in a meta-analysis by including publication year as a moderator into **meta-regression** [13, 16] (Figure IIA). A **cumulative meta-analysis** (CMA) in which studies are entered into the analysis in chronological order provides another useful tool for detection of changes in cumulative evidence over time [47]. As all visual tools, CMA plots might be subject to misinterpretation and should be supplemented by formal statistical methods which should take into account multiple testing inherent in CMA [44]. Therefore, we recommend the use of cumulative meta-analysis in combination with control plots [44](Box 2), which can be plotted using R package qcc [48].

Another class of methods has been developed for sequential clinical trials in medicine where the accumulated evidence is periodically reviewed as the trial progresses with a view of stopping the trial early if required. Applications of these techniques to meta-analysis exist [49-52], but we do not recommend their use (see critique of these approaches in [53, 54]). Furthermore, some ecological meta-analyses assess temporal changes in effect sizes by subdividing studies into groups based on the publication year (e.g. by decades or published before and after year X) and comparing mean effect sizes between the groups [42, 55]. This relatively crude approach ignores likely gradual character of temporal changes and their possible occurrence within as well as between the studied groups, therefore we do not recommend it.

Use of tests for temporal changes in estimated effect sizes within individual meta-analyses may prove particularly effective if such changes occur mainly early on. For instance, Fanelli et al. [56] have recently shown that declines in magnitude of the effect sizes with publication year in meta-analyses are not linear and there is a strong “first-year” effect, in which the earliest studies are more likely to overestimate the overall effect than all later ones. This effect might occur if early studies are statistically underpowered [57]. As a result, first meta-analyses on the topic based on the first few early primary studies available are particularly likely to overestimate the effect and results of such meta-analyses need to be treated with caution.

In addition to testing for temporal trends within meta-analyses, updating existing meta-analyses can also be an effective tool in early detection of evidence reversal. Useful guidelines on when and how to update systematic reviews have been recently published by the Cochrane panel [58]. In order to enable such an update, the transparency of methods used in the published ecological meta-analyses needs to improve. For instance, the database on which previous meta-analysis has been based need to be available as well as the detailed literature search strategy. Unfortunately, the majority of published ecological meta-analyses do not fulfil these criteria [45]. Another problem is that publication of meta-analyses and any subsequent updates can take many months, which means that by the time of publication these reviews are already out of date. Shojania et al. [19] proposed that when the process of submission and rejection from other journals has resulted in the passage of more than one year from the date of the previous search, authors should update the search before resubmission. Another approach to narrowing the time gap between evidence and practice and to reducing the evidence reversal impact is to conduct living systematic reviews, online summaries updated as new research becomes available [59]. This approach, however, similarly to cumulative meta-analysis, might inflate the rate of false-positive findings due to repeated testing. Therefore, previously discussed methods or the Bayesian approach discussed in Elliott et al. [52] should be be used for monitoring accumulating evidence while reducing the probability of false positives.

Concluding Remarks and Future Perspectives

We believe that more widespread application of methods for monitoring of temporal changes in reported effects (Box 2) and for updating meta-analyses will facilitate conclusions on sufficiency of evidence for policy making and timely detection of evidence reversal. Moreover, analysis of causes of temporal changes in cumulative evidence will reveal whether these changes require adjustment in previously accepted management policies. Ultimately this will allow saving of time and resources in the development of management strategies thus making conservation action more effective.

Box 1. An example of evidence reversal in conservation biology

Two meta-analyses on effects of predator removal on bird population provide a good example of how heterogeneity in effect size can lead to evidence reversal and change the conclusions and practical recommendations. The first meta-analysis by Coté and Sutherland [60] showed that predator removal significantly increases postbreeding population sizes (i.e. autumn densities) of the target bird species, but does not significantly affect breeding population sizes (Fig. I). Coté and Sutherland concluded therefore that predator removal fulfils the goal of game management (enhancing harvestable postbreeding populations) but is of less use for conservation management (increasing bird breeding population sizes). However, a more recent meta-analysis on the same topic by Smith et al. [61] arrived at the opposite conclusion, showing that the predator removal effect on breeding population numbers is statistically significant, but the effect of predator removal on postbreeding populations is no longer significant (Fig. I). Smith et al. concluded therefore that predator removal is an effective strategy for the conservation of bird populations, but not for game management. Hence, two meta-analyses on the same topic conducted 13 years apart reached opposite conclusions on the effectiveness of the assessed conservation measures. In this particular case the difference in the results of the two meta-analyses was not due to changes in true biological effects but due to heterogeneity. Smith et al. have revealed that predator removal was effective in increasing postbreeding bird populations on mainland, but not on islands. Since the proportion of studies conducted on islands increased with time and was higher in meta-analysis by Smith et al. than in the earlier meta-analysis on the same topic by Coté and Sutherland, the magnitude of the overall effect estimate of predator removal on postbreeding populations was much smaller in the former meta-analysis. This example shows the importance of updating the results of previous meta-analyses as new studies on the topic are published as well as the importance of examining the sources of variation in effect sizes and drawing inference from studies conducted under similar ecological conditions.

Box 2. Methods of detection of temporal changes in reported effects

The simplest way to visualize a potential temporal trend in a meta-analytic dataset is by plotting effect sizes from individual primary studies against their publication years (Fig. IIA). In order to statistically test the above relationship, publication year can be used as a moderator in a meta-regression model [13, 16, 62]. Alternatively, cumulative meta-analysis (CMA) where studies are added to the analysis in chronological order and meta-analytic means are cumulatively calculated over the years can be used to visually detect temporal trends (Fig. IIB, [47]). Finally, methods of statistical quality control such as **Xbar charts** and **CUSUM charts** can be used to detect possible outliers and trends over time in meta-analysis [44, 63]. Xbar charts are based on detecting outlying observations under normality. The control limits on Xbar charts are usually plotted at 3 standard deviations, corresponding to a significance level of $\alpha = 0.0027$. The CUSUM charts plot the cumulative sums of the deviations of the sample values from a target value. The chart is restricted from falling below zero, and often two one-sided CUSUM charts (for positive and negative deviations) are plotted simultaneously.

We demonstrate the application of four different methods for detection of temporal trends in effect sizes on Figure II using a subset from the meta-analysis by Batáry et al. [64] on effects of agri-environment schemes on biodiversity as an example. A bubble plot (Fig. IIA) shows decrease in effect sizes with publication year, particularly between 1995 and 2005. The cumulative meta-analysis plot (Fig. IIB) demonstrates similar trend with initial increase of the effect until the fourth study was added to the analysis and the subsequent decrease in the magnitude of the effect. The cumulative effect size becomes significantly different from 0 at study 6, and even more so at study 7, but then the effect declines as more studies are added to the analysis. In this example, the effect size reached at study 7 (d= 1.165) is monitored over time. The Xbar chart (Fig. IIC) shows one high outlier (study 4), two low outliers (studies 11 and 14) and one significant run rule violation (a series of more than 7 negative deviations from the target value), suggesting a shift in the process mean. CUSUM chart (Fig. IID) shows that while the cumulative effects were significantly above 1.165 at studies 4 and 5, the cumulative results are significantly below this value for the last 4 studies, indicating a decrease in the mean effect size.

Figures:

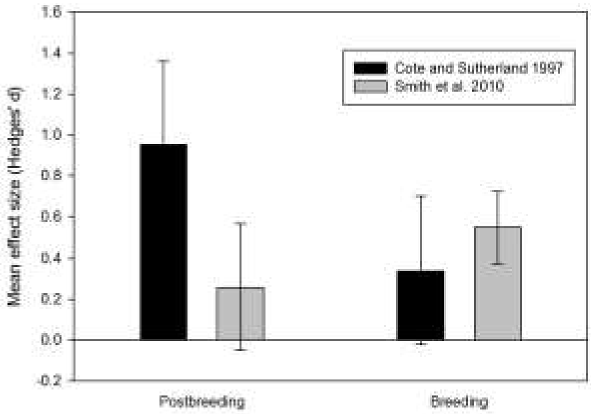
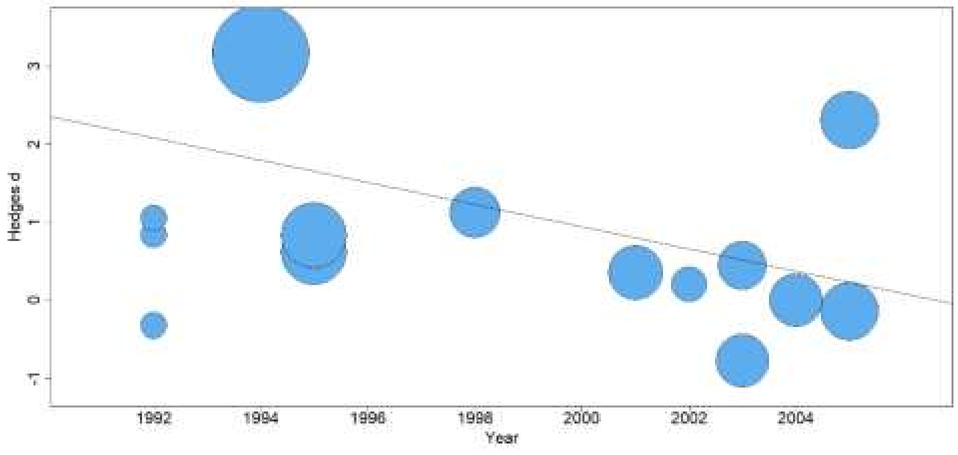


Fig. I. Differences in estimates of the effects of predator removal on postbreeding and breeding population size of birds (data from meta-analyses by Côté and Sutherland [60] and Smith et al. [61]). Error bars represent 95% confidence intervals; mean effects are not significantly different from 0 if confidence intervals include 0. Number of studies included in the analysis: 13 and 51 for breeding population size estimates and 10 and 19 for postbreeding population size estimates in Côté and Sutherland and Smith et al., respectively.



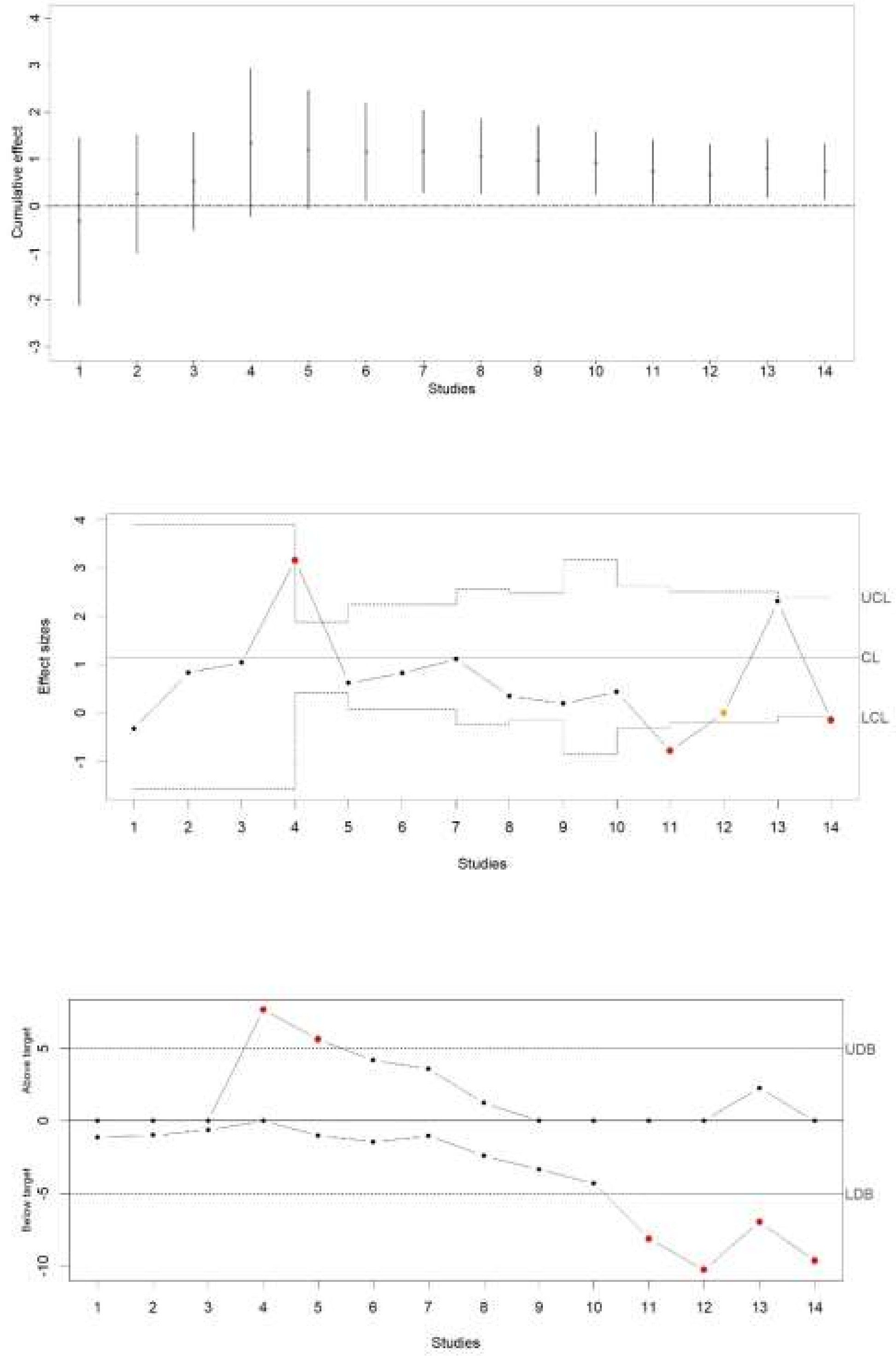


Fig. II. Illustration of four different methods of exploration of temporal trends in reported effects. We used a subset from the meta-analysis by Batáry et al. [64] representing 14 studies assessing the effects of agri-environment management on biodiversity in simple landscapes within croplands and published before 2006. Effect sizes are standardized mean differences (Hedges' d) between biodiversity measures in extensively and intensively managed fields. **A:** a bubble plot showing the results of meta-regression with publication year as a moderator. Effect sizes are weighted by their precision; larger bubbles indicate more precise estimates and smaller bubbles less precise. **B**: cumulative meta-analysis showing changes in cumulative mean effect size and the 95% confidence interval as more recent studies are added in the analysis. **C**. Xbar chart. Horizontal central line on Xbar chart corresponds to the combined effect size of the first seven studies (d= 1.165). **D**. CUSUM chart. Control limits (dashed lines) are at ±3SD, out-of-control values are in red, run test violations (a series of consecutive deviations from the expected value which are of the same sign) are in orange.

Glossary

Cumulative meta-analysis: a type of meta-analysis in which effect sizes from individual studies are entered into the analysis sequentially, one study at the time, based on some predetermined order (most commonly chronological); the mean effect size and confidence intervals are recalculated at each step.

CUSUM chart: a cumulative sum (CUSUM) chart is a type of control chart used to monitor changes in the process mean. It plots the cumulative sum of deviations of the sample values from a target value.

Decline effect: decrease in support for scientific claims over time as original studies are repeated.

Effect size: a quantitative measure of the magnitude of study outcome that puts all responses across studies in a meta-analysis on the same scale. It provides a “common currency” for comparisons of the results across studies. Metrics of effect size most commonly used in ecology include standardized mean differences, response ratios and correlation coefficients.

Evidence-based conservation: conservation management actions and policy making based on systematic assessment (e.g. systematic review and meta-analysis) of existing scientific evidence of current effectiveness of different management interventions.

Evidence reversal: occurs when an existing claim is tested and the original evidence is contradicted by new evidence.

Heterogeneity: the variation in the effect size estimates among studies.

Meta-analysis: a set of statistical methods for combining magnitudes of the effects across different data sets addressing the same research question.

Meta-regression: an extension of basic meta-analysis model in which *moderators* are used to explain between-study variation in effect sizes (heterogeneity).

Moderator: a variable (continuous or categorical) which is used in meta-regression to explain between-study variation in effect sizes.

Publication bias: influence of magnitude, direction, and/or statistical significance of research findings on the probability of a study to be published.

Systematic review: the type of research synthesis on a precisely defined topic using systematic and explicit methods to identify, select, critically appraise, and analyse relevant research. Systematic review may or may not include *meta-analysis* of the data.

Time-lag bias: influence of study results on the time it takes to complete and publish a study; often refers to delayed publication of non-significant results.

Xbar ($\overline{X}$) chart: a type of control chart that is used to monitor the means of successive samples based on detecting outlying observations under normality. The control limits on Xbar charts are usually plotted at 3 standard deviations, corresponding to a significance level of α = 0.0027.

References

1. Sutherland, W.J. *et al.* (2004) The need for evidence-based conservation. *Trends Ecol. Evol.* 19, 305-308.

2. Fernandez-Duque, E. and Valeggia, C. (1994) Meta-analysis: a valuable tool in conservation research. *Conserv. Biol.* 8, 555-561.

3. Pullin, A.S. and Stewart, G.B. (2006) Guidelines for systematic review in conservation and environmental management. *Conserv. Biol.* 20, 1647-1656.

4. Gurevitch, J. *et al.* (2018) Meta-analysis and the science of research synthesis. *Nature* 555, 175-182.

5. Côté, I.M. and Stewart, G.B. (2013) Contributions of meta-analysis to conservation and management In *Handbook of meta-analysis in ecology and evolution* (Koricheva, J. *et al.* eds), pp. 420-425, Princeton University Press.

6. Haddaway, N.R. (2015) A call for better reporting of conservation research data for use in meta-analyses. *Conserv. Biol.* 29, 1242-1245.

7. Koricheva, J. *et al.* (2013) Temporal trends in effect sizes: causes, detection, and implications. In *Handbook of meta-analysis in ecology and evolution* (Koricheva, J. *et al.* eds), pp. 237-254, Princeton University Press.

8. Barto, E.K. and Rillig, M.C. (2012) Dissemination biases in ecology: effect sizes matter more than quality. *Oikos* 121, 228-235.

9. Jennions, M.D. and Møller, A.P. (2002) Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. *Proc. R. Soc. Lond. B* 269, 43-48.

10. Leimu, R. and Koricheva, J. (2004) Cumulative meta-analysis: a new tool for detection of temporal trends and publication bias in ecology. *Proc. R. Soc. Lond. B* 271, 1961-1966.

11. Nykanen, H. and Koricheva, J. (2004) Damage-induced changes in woody plants and their effects on insect herbivore performance: a meta-analysis. *Oikos* 104, 247-268.

12. Saikkonen, K. *et al.* (2006) Model systems in ecology: dissecting the endophyte-grass literature. *Trends Plant Sci.* 11, 428-433.

13. Sánchez-Tójar, A. *et al.* (2018) Meta-analysis challenges a textbook example of status signalling and demonstrates publication bias. *eLife* 7, e37385.

14. Fanshawe, T.R. *et al.* (2017) A large-scale assessment of temporal trends in meta-analyses using systematic review reports from the Cochrane Library. *Res. Synth. Meth.* 8, 404-415.

15. Monsarrat, P. and Vergnes, J.-N. (2018) The intriguing evolution of effect sizes in biomedical research over time: smaller but more often statistically significant. *GigaScience* 7, 1-10.

16. Gehr, B.T. *et al.* (2006) The fading of reported effectiveness. A meta-analysis of randomized controlled trials *BMC Med. Res. Methodol.* 6, 25.

17. Johnsen, T.J. and Friborg, O. (2015) The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: a meta-analysis. *Psychol. Bull.* 141, 747-768.

18. de Bruin, A. and Della Sala, S. (2015) The decline effect: How initially strong results tend to decrease over time. *Cortex* 73, 375-377.

19. Shojania, K.G. *et al.* (2007) How quickly do systematic reviews go out of date? A survival analysis. *Ann. Intern. Med.* 147, 224-233.

20. Arnqvist, G. and Wooster, D. (1995) Meta-analysis: synthesizing research findings in ecology and evolution. *Trends Ecol. Evol.* 10, 236-240.

21. Begley, C.G. and Ioannidis, J.P.A. (2015) Reproducibility in Science. *Circ. Res.* 116, 116-126.

22. Mace, G.M. (2010) Drivers of biodiversity change. In *Trade-offs in conservation: deciding what to save* (Leader-Williams, N. *et al.* eds), pp. 349-364, Blackwell

23. Rudel, T.K. *et al.* (2009) Changing drivers of deforestation and new opportunities for conservation. *Conserv. Biol.* 23, 1396-1405.

24. Møller, A.P. *et al.* (2008) Populations of migratory bird species that did not show a phenological response to climate change are declining. *Proc. Natl. Acad. Sci. U. S. A.* 105, 16195-16200.

25. Strauss, S.Y. *et al.* (2008) Evolution in ecological field experiments: implications for effect size *Ecol. Lett.* 11, 199-207.

26. Fischbach, L.A. *et al.* (2002) Sources of variation of Helicobacter pylori treatment success in adults worldwide: a meta-analysis. *Int. J. Epidemiol.* 31, 128-139.

27. Tseng, M. *et al.* (2018) Decreases in beetle body size linked to climate change and warming temperatures. *J. Anim. Ecol.* 87, 647-659.

28. Baudron, A.R. *et al.* (2014) Warming temperatures and smaller body sizes: synchronous changes in growth of North Sea fishes. *Global Change Biol.* 20, 1023-1031.

29. Caruso, N.M. *et al.* (2014) Widespread rapid reductions in body size of adult salamanders in response to climate change. *Global Change Biol.* 20, 1751-1759.

30. Roca, I.T. *et al.* (2016) Shifting song frequencies in response to anthropogenic noise: a meta-analysis on birds and anurans. *Behav. Ecol.* 27, 1269–1274.

31. Barto, E.K. and Rillig, M.C. (2010) Does herbivory really suppress mycorrhiza? A meta-analysis. *J. Ecol.* 98, 745-753.

32. Timi, J.T. and Poulin, R. (2007) Different methods, different results: temporal trends in the study of nested subset patterns in parasite communities. *Parasitology* 135, 131-138.

33. Simmons, L.W. *et al.* (1999) Fluctuating paradigm. *Proc. R. Soc.Lond. B* 266, 593-595.

34. Senior, A.M. *et al.* (2016) Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology* 97, 3293-3299.

35. Nuijten, M.B. *et al.* (2015) The replication paradox: combining studies can decrease accuracy of effect size estimates. *Rev. Gen. Psychol.* 19, 172-182.

36. Ioannidis, J.P.A. and Trikalinos, T.A. (2005) Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J. Clin. Epidemiol.* 58, 543-549.

37. Heleno, R.H. (2014) Meta-analyses and the "editorial love of controversy". *Web Ecol.* 14, 23-25.

38. Alatalo, R.V. *et al.* (1997) Heritabilities and paradigm shifts. *Nature* 385, 402-403.

39. Ioannidis, J.P.A. (2005) Differentiating biases from genuine heterogeneity: distinguishing artifactual from substantive effects. In *Publication bias in meta-analysis: prevention, assessment and adjustements* (Rothstein, H.R. *et al.* eds), pp. 287-302, Wiley.

40. Sutton, D. *et al.* (2018) Evidence reversal - when new evidence contradicts current claims: a systematic overview review of definitions and terms. *J. Clin. Epidemiol.* 94, 76-84.

41. Whittaker, R.J. (2010) Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness-productivity relationship. *Ecology* 91, 2522-2533.

42. Batáry, P. *et al.* (2015) The role of agri-environment schemes in conservation and environmental management. *Cons. Biol.* 29, 1006-1016.

43. Trikalinos, T.A. and Ioannidis, J.P.A. (2005) Assessing the evolution of effect sizes over time. In *Publication bias in meta-analysis* (Rothstein, H.R. *et al.* eds), pp. 241-259, Wiley.

44. Kulinskaya, E. and Koricheva, J. (2010) Use of quality control charts for detection of outliers and temporal trends in cumulative meta-analysis *Res. Synth. Meth.* 1, 297-307.

45. Koricheva, J. and Gurevitch, J. (2014) Uses and misuses of meta-analysis in plant ecology. *J. Ecol.* 102, 828-844.

46. Nakagawa, S. *et al.* (2017) Meta-evaluation of meta-analysis: ten appraisal questions for biologists. *BMC Biol.* 15, 18.

47. Leimu, R. and Koricheva, J. (2004) Cumulative meta-analysis: a new tool for detection of temporal trends and publication bias in ecology. *Proc. R. Soc. Lond. B* 271, 1961-1966.

48. Scrucca, L. (2004) qcc: an R package for quality control charting and statistical process control. *R News* 4/1, 11-17.

49. Pogue, J. and Yusuf, S. (1997) Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Contemp. Clin. Trials* 18, 580-593.

50. Brok, J. *et al.* (2008) Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *J. Clin. Epidemiol.* 61, 763-769.

51. Wetterslev, J. *et al.* (2008) Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J. Clin. Epidemiol.* 61, 64-75.

52. Higgins, J.P.T. *et al.* (2011) Sequential methods for random-effects meta-analysis. *Stat. Med.* 30, 903-921.

53. Kulinskaya, E. *et al.* (2016) Sequential biases in accumulating evidence. *Res. Synth. Meth.* 7, 294-305.

54. Kulinskaya, E. and Wood, J. (2014) Trial sequential methods for meta-analysis. *Res. Synth. Meth.* 5, 212-220.

55. Magdaong, E.T. *et al.* (2014) Long-term change in coral cover and the effectiveness of marine protected areas in the Philippines: a meta-analysis. *Hydrobiologia* 733, 5-17.

56. Fanelli, D. *et al.* (2017) Meta-assessment of bias in science. *Proc. Natl. Acad. Sci. U.S.A.* 114, 3714-3719.

57. Ioannidis, J.P.A. (2008) Why most discovered true associations are inflated. *Epidemiology* 19, 640-647.

58. Garner, P. *et al.* (2016) When and how to update systematic reviews: consensus and checklist. *BMJ* 354, i3507.

59. Elliott, J.H. *et al.* (2014) Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Medicine* 11, e1001603-e1001603.

60. Coté, I.M. and Sutherland, W.J. (1997) Effectiveness of removing predators to protect bird populations. *Conserv. Biol.* 11, 395-405.

61. Smith, R.K. *et al.* (2010) Effectiveness of predator removal for enhancing bird populations. *Conserv. Biol.* 24, 820-829.

62. Tuck, S.L. *et al.* (2014) Land-use intensity and the effects of organic farming on biodiversity: a hierarchical meta-analysis. *J. Appl. Ecol.* 51, 746-755.

63. Dogo, S.H. *et al.* (2017) Sequential change detection and monitoring of temporal trends in random-effects meta-analysis. *Res. Synth. Meth.* 8, 220-235.

64. Batáry, P. *et al.* (2011) Landscape-moderated biodiversity effects of agri-environmental management: a meta-analysis. *Proc. R. Soc. Lond. B.* 278, 1894-1902.