Readme for Codes and Datasets used in Purgar et al. 2022: Quantifying research waste in ecology

This file contains five data tables and two R codes

Data tables include:

1) **Dataset\_starting** = all the studies and extracted effect sizes as obtained after the initial screening and data extraction

2) **Dataset\_pooled** = derived from ‘Dataset\_starting’ by either removing some effect sizes (those that refer only to power or pooling effect sizes from the same study. Pooled and removed effect sizes are marked in the column ‘remove/pool’ of the ‘Dataset\_starting’ data table

3) **Dataset\_MA\_final** = derived from Dataset\_pooled by removing estimates that refer to the same substage or stage of the research life-cycle. Here, the worst estimates are kept as they are limiting the value of research. All the removed effect sizes are marked in the column ‘selecting the worst estimate’ of the Dataset\_pooled data table. This table is used in the meta-analyses (thus, has to be uploaded when running ‘Meta\_analysis\_waste’ code

4) **Meta\_analytic\_means** = estimates of the main research waste components with 95%CI of the meta-analytic mean. Used to create Figures (‘Figures\_waste’ code)

5) **Study\_planning\_MA\_means** = estimates of the research waste components at the Study planning stage with 95%CI of the meta-analytic mean. Used to create Figures (‘Figures\_waste’ code)

R codes include:

1) **Meta\_analysis\_waste** = scripts to run meta-analyses (uses data table ‘Dataset\_MA\_final’)

2) **Figures\_waste** = scripts to create main and supplementary figures. Uses ‘Meta\_analytic\_means’ and ‘Study\_planning\_MA\_means’ data tables

Datasets 1,2 and 3 contain following entries:

|  |  |
| --- | --- |
| **Variable** | **Meaning** |
| Authors [for datasets 1 & 2] | Author(s) of a study |
| Title [for datasets 1 & 2] | The full title of the study |
| Scientific field | General scientific field stated in the study or assigned later by the review team |
| Subfield | Subfield stated in the study |
| Exact topic/coverage | The exact topic/coverage as extracted from the study |
| Degree of generality | The degree of generality is a product of the exact topic/coverage and was determined by a consensus between MP, AC, and TK. Studies that are narrower in their topic (e.g., Cassey et al. (2004), surveying facultative sex-ratio adjustment in birds) were coded as 1, while the most general studies (e.g. covering the whole area of ecology) were coded as 3. We coded 2 studies that were quite general, but not as general as covering the full field of ecology (e.g. conservation biology) |
| Type of studies | Type of studies that the focal article has used to derive the estimate of waste: experimental studies, observational studies, or any (experimental and observational) |
| Type of literature | Type of studies the focal article used to derive the estimate of waste: peer-reviewed, unpublished papers, published and unpublished doctoral thesis, or any |
| Period | A span of years from the first-published article used in a focal article, to the last published one |
| Stage | Stage of the research-cycle for which waste was estimated: Study planning, Reporting, or Publication |
| Substage | Only for the Study planning stage where substage can be: Core study design, Blinding, Analysis, depending on what exact time point the waste has happened |
| Exact effect | The exact type of estimated loss, as reported in the focal study |
| N | Size of the sample used in a meta-study for calculating the estimates |
| Units sampled | What was the main data point in a meta-study |
| Estimate | A quantitative assessment of the various types of loss, expressed as a percentage. |
| Notes [datasets 1 & 2] | Any additional notes about the effect size or a study. |

Additional columns specific for each dataset include:

1) ‘Dataset\_starting’ includes columns:

- ‘remove/pool’. This column denotes whether an entry was removed or pooled with other entries to create ‘Dataset\_pooled’ table.

- ‘Retrieved in’. This column codes for whether the study was identified in the ‘primary search’ using search string; in 1st, 2nd, 3d, or 4th round of backward and forward search; or sourced externally.

2) ‘Dataset\_pooled’ includes column

- ‘selecting the worst estimate’ that codes for whether and entry was removed from the Dataset\_MA\_final (because several estimates of the same stage of research waste exist for a study)

3) ‘Dataset\_MA\_final’ includes columns

- ‘Ref.n’ (see references list below)

- ‘DOG2’ same as DOG but the two broader categories (i.e. "2" and "3") are merged (now coded as DOG2)

- ‘N\_f’ codes for the number of Units sampled that were wasted;

- ‘Remove from the global MA’ codes for whether the effect size was used in the global MA for calculating meta-analytic mean of Study planning stage or not. Effect sizes that were unused in the global MA were used in the MA to calculate the meta-analytic means of substages of Study planning.

**Reference list for papers in the data tables**

**References for studies that were used in the meta-analysis, same codes as in Data table Dataset\_MA\_final, under ‘Ref.n’ column**

1. Bennett, L. T. & Adams, M. A. Assessment of ecological effects due to forest harvesting: approaches and statistical issues. *Journal of Applied Ecology* **41**, 585–598 (2004).
2. Campbell, H. A. *et al.* Finding our way: On the sharing and reuse of animal telemetry data in Australasia. *Sci Total Environ* **534**, 79–84 (2015).
3. Cassey, P., Ewen, J. G., Blackburn, T. M. & Møller, A. P. A survey of publication bias within evolutionary ecology. *Proc Biol Sci* **271 Suppl 6**, S451-454 (2004).
4. Gillespie, B. R., Desmet, S., Kay, P., Tillotson, M. R. & Brown, L. E. A critical analysis of regulated river ecosystem responses to managed environmental flows from reservoirs. *Freshwater Biology* **60**, 410–425 (2015).
5. Jennions, M. D. & Møller, A. P. Publication bias in ecology and evolution: an empirical assessment using the ‘trim and fill’ method. *Biol Rev Camb Philos Soc* **77**, 211–222 (2002).
6. Kardish, M. R. *et al.* Blind trust in unblinded observation in Ecology, Evolution, and Behavior. *Frontiers in Ecology and Evolution* **3**, 51 (2015).
7. Møller, A. P., Thornhill, R. & Gangestad, S. W. Direct and indirect tests for publication bias: asymmetry and sexual selection. *Animal Behaviour* **70**, 497–506 (2005).
8. Parker, T. H. What do we really know about the signalling role of plumage colour in blue tits? A case study of impediments to progress in evolutionary biology. *Biol Rev Camb Philos Soc* **88**, 511–536 (2013).
9. Sánchez-Tójar, A. *et al.* Meta-analysis challenges a textbook example of status signalling and demonstrates publication bias. *eLife* **7**, e37385 (2018).
10. Zaitsev, A. S., Gongalsky, K. B., Malmström, A., Persson, T. & Bengtsson, J. Why are forest fires generally neglected in soil fauna research? A mini-review. *Applied Soil Ecology* **98**, 261–271 (2016).
11. Zvereva, E. L. & Kozlov, M. V. Biases in studies of spatial patterns in insect herbivory. *Ecological Monographs* **89**, e01361 (2019).
12. Forstmeier, W. & Schielzeth, H. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner’s curse. *Behav Ecol Sociobiol* **65**, 47–55 (2011).
13. Holman, L., Head, M. L., Lanfear, R. & Jennions, M. D. Evidence of Experimental Bias in the Life Sciences: Why We Need Blind Data Recording. *PLOS Biology* **13**, e1002190 (2015).
14. Koricheva, J. Non-significant results in ecology: a burden or a blessing in disguise? *Oikos* **102**, 397–401 (2003)
15. Yoccoz, N. G. Use, Overuse, and Misuse of Significance Tests in Evolutionary Biology and Ecology. *Bulletin of the Ecological Society of America* **72**, 106–111 (1991).
16. van Wilgenburg, E. & Elgar, M. A. Confirmation Bias in Studies of Nestmate Recognition: A Cautionary Note for Research into the Behaviour of Animals. *PLOS ONE* **8**, e53548 (2013).
17. Fidler, F., Burgman, M. A., Cumming, G., Buttrose, R. & Thomason, N. Impact of criticism of null-hypothesis significance testing on statistical reporting practices in conservation biology. *Conserv Biol* **20**, 1539–1544 (2006).
18. Haddaway, N. R., Styles, D. & Pullin, A. S. Evidence on the environmental impacts of farm land abandonment in high altitude/mountain regions: a systematic map. *Environmental Evidence* **3**, 17 (2014).
19. Mrosovsky, N. & Godfrey, M. The path from grey literature to Red Lists. *Endangered Species Research* **6**, 185–191 (2008).
20. Brlík, V. *et al.* Weak effects of geolocators on small birds: A meta-analysis controlled for phylogeny and publication bias. *Journal of Animal Ecology* **89**, 207–220 (2020).
21. McDonald, S., Cresswell, T., Hassell, K. & Keough, M. Experimental design and statistical analysis in aquatic live animal radiotracing studies: A systematic review. *Critical Reviews in Environmental Science and Technology*, 1–30 (2021).
22. Ramage, B. S. *et al.* Pseudoreplication in tropical forests and the resulting effects on biodiversity conservation. *Conserv Biol* **27**, 364–372 (2013).
23. Kozlov, M. V. Plant studies on fluctuating asymmetry in Russia: Mythology and methodology. *Russian Journal of Ecology* **48**, 1–9 (2017).
24. O’Brien, C., van Riper, C. & Myers, D. E. Making reliable decisions in the study of wildlife diseases: using hypothesis tests, statistical power, and observed effects. *J Wildl Dis* **45**, 700–712 (2009).
25. Hurlbert, S. H. Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs* **54**, 187–211 (1984).
26. Heffner, R. A., Butler, M. J. & Reilly, C. K. Pseudoreplication Revisited. *Ecology* **77**, 2558–2562 (1996).
27. Kozlov, M. V. Pseudoreplication in ecological research: the problem overlooked by Russian scientists. *Zh Obshch Biol* **64**, 292–307 (2003).
28. Vorobeichik, E. L. & Kozlov, M. V. Impact of point polluters on terrestrial ecosystems: Methodology of research, experimental design, and typical errors. *Russ J Ecol* **43**, 89–96 (2012).
29. Hurlbert, S. H. & White, M. D. Experiments with Freshwater Invertebrate Zooplanktivores: Quality of Statistical Analyses. *Bulletin of Marine Science* **53**, 128–153 (1993).
30. Waller, B., Warmelink, L., Liebal, K., Micheletta, J. & Slocombe, K. Pseudoreplication: A widespread problem in primate communication research. *Animal Behaviour* **86**, 483–488 (2013).
31. Sallabanks, R., Arnett, E. B. & Marzluff, J. M. An Evaluation of Research on the Effects of Timber Harvest on Bird Populations. *Wildlife Society Bulletin (1973-2006)* **28**, 1144–1155 (2000).
32. Cornwall, C. E. & Hurd, C. L. Experimental design in ocean acidification research: problems and solutions. *ICES Journal of Marine Science* **73**, 572–581 (2016).
33. Johnson III, W. T. & Freeberg, T. M. Pseudoreplication in use of predator stimuli in experiments on antipredator responses. *Animal Behaviour* **119**, 161–164 (2016).
34. Christie, A. P. *et al.* Quantifying and addressing the prevalence and bias of study designs in the environmental and social sciences. *Nat Commun* **11**, 6377 (2020).

References for studies that were exluced from the meta-analyss (but present in the Dataset\_starting data table)

1. Lemoine, N. P. *et al.* Underappreciated problems of low replication in ecological field studies. *Ecology*, **97**, 2554-2561 (2016).
2. Jennions, M. D., Møller, A.P. A survey of the statistical power of research in behavioral ecology and animal behavior. *Behavioral Ecology* **14**, 438–445 (2003).
3. Smith, D., Hardy, I. C., & Gammell, M. P. Power rangers: no improvement in the statistical power of analyses published in Animal Behaviour. *Animal Behaviour*, **81**, 347-352 (2011).
4. Garamszegi, L. Z., Markó, G. & Herczeg, G. A meta-analysis of correlated behaviours with implications for behavioural syndromes: mean effect size, publication bias, phylogenetic effects and the role of mediator variables. *Evol Ecol* **26,**1213–1235 (2012).
5. Cleasby, I. R. *et al.* What is our power to detect device effects in animal tracking studies? *Methods Ecol Evol*., **12**, 1174– 1185 (2021).
6. Davidson, A. & Hewitt, C. L. How often are invasion-induced ecological impacts missed? *Biological Invasions*, **16**, 1165-1173 (2014).