APPLICATION

`orchaRd 2.0`: An R package for visualizing meta-analyses with orchard plots

---

Shinichi Nakagawa[1], Malgorzata Lagisz[1], Rose E. O'Dea[2], Patrice Pottier[1], Joanna Rutkowska[3], Alistair M. Senior[4], Yefeng Yang[1], Daniel W. A. Noble[5]

[1] Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia

[2] Wissenschaftskolleg zu Berlin, Wallotstraße 19, 14193, Berlin, Germany

[3] Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland

[4] Charles Perkins Centre and School of Life and Environmental Sciences, University of Sydney, Camperdown, NSW 2006, Australia

[5] Division of Ecology and Evolution, Research School of Biology, Australian National University, Canberra, ACT, Australia

* Correspondence: S. Nakagawa & D. W. A. Noble

e-mail: s.nakagawa@unsw.edu.au

email: daniel.noble@anu.edu.au

Short title: The Orchard Plot

Abstract

1. Although meta-analysis has become an essential tool in ecology and evolution, reporting of meta-analytic results can still be much improved. To aid this, we have introduced the orchard plot, which presents not only overall estimates and their confidence intervals but also shows corresponding heterogeneity (as prediction intervals) and individual effect sizes.

2. Here, we have added significant enhancements by integrating many new functionalities as `orchaRd 2.0`. This updated version allows the visualisation of heteroscedasticity (different variances across levels of a categorical moderator), marginal estimates (e.g., marginalising out effects other than the one visualized), conditional estimates (i.e., estimates of different groups conditioned upon specific values of a continuous variable), and visualizations of all types of interactions between two categorical/continuous moderators.

3. `orchaRd 2.0` has additional functions which calculate key statistics from multilevel meta-analytic models such as I^2 and R^2. Importantly, `orchaRd 2.0` contributes to better reporting by complying with PRISMA-EcoEvo (preferred reporting items for systematic reviews and meta-analyses in ecology and evolution). Taken together, `orchaRd 2.0` can improve the presentation of meta-analytic results and facilitate the exploration of previously neglected patterns.

4. In addition, as a part of a literature survey, we found that graphical packages are rarely cited (~3%). We plea that researchers credit developers and maintainers of graphical packages, e.g., by citations in a figure legend, acknowledging the use of relevant packages.

KEYWORDS

caterpillar plot, evidence synthesis, graphical tools, credible interval, credibility interval, summary forest plot, meta-regression

1 | INTRODUCTION

Meta-analysis has become an essential synthesis tool across the medical, social, and biological sciences (Gurevitch *et al.* 2018; Cooper, Hedges & Valentine 2019; Higgins *et al.* 2019; Schmid, Stijnen & White 2021). In fields such as medicine, meta-analytic results are typically shown in a forest plot that presents effect sizes and their 95% confidence intervals (CIs) from each study in the meta-analysis. However, in ecology and evolution, the use of forest plots is rare because meta-analyses in this field often include > 100 effect sizes, making a traditional forest plot impractical (Senior *et al.* 2016; Gurevitch *et al.* 2018). Instead, researchers use a “forest-like plot” with the overall mean effect size estimate and their 95% CIs for different levels of a categorical moderator (predictor variable). Such estimates are derived from a meta-regression model or from subset/sub-group analyses. For example, such a plot could show estimates from five different taxa, six different geographical areas, or three different methods. A recent survey found 72 out of 102 ecological and evolutionary meta-analyses presented forest-like plots (Nakagawa *et al.* 2021). Contributing to the popularity of forest-like plots is the fact that meta-analytic moderators are often categorical rather than continuous variables. Despite their popularity, forest-like plots in ecology and evolution often lack important information such as individual effect sizes and estimates of heterogeneity among effect sizes (Schild & Voracek 2015; Nakagawa *et al.* 2021).

Nakagawa et al. (2021) introduced an information-rich version of a forest-like plot, named the ‘orchard’ plot. Orchard plots provide: 1) point estimates (i.e., regression coefficients); 2) CIs, 3) prediction intervals, PIs (which show heterogeneity among effect sizes); and 4) individual effect sizes scaled by their precision (the inverse of the square root of the sampling variance). Nakagawa et al. (2021) implemented the orchard plot using functions that use the most popular and comprehensive meta-analysis *R* package, `metafor` (Viechtbauer 2010), and `ggplot2` graphics (Wickham 2009). However, the original implementation (`orchaRd 1.0`) was limited to relatively simple meta-regression (models with moderators), as the package only allowed the user to draw orchard plots from a model with one categorical moderator. In addition, it was only possible to visualise a meta-regression model that assumed homoscedasticity across levels of the single categorical moderator (i.e., all levels have the same variance, which may be unrealistic, e.g., Zajitschek *et al.* 2020; Wilson *et al.* 2022).

In this article, we enhance the visualization capabilities of the `orchaRd` package by integrating the functionalities of the R package `emmeans` (Lenth *et al.* 2018) in four ways. The first three extend orchard plots by allowing visualisation of: i) heteroscedasticity (different variances across levels of a categorical moderator); ii) marginal estimates (e.g., marginalising all other moderators apart from the one visualized); and iii) conditional estimates (i.e., estimates of different groups/levels of a categorical variable, conditioned upon specific values of a continuous variable). The fourth capability allows for 'bubble' plots to be created of: i) a continuous variable; and ii) interactions between a continuous and categorical variable from multi-moderator models with and without heteroscedasticity. In addition, we add helper functions to calculate key statistics from multilevel meta-analytic models such as I^2 (Cheung 2014) and R^2 (Aloe, Becker & Pigott 2010; Nakagawa & Schielzeth 2013) along with their CIs. These new functionalities not only better visualize meta-analytic results in ecology and evolution but also facilitate the exploration of previously neglected patterns, such as heteroscedasticity in meta-analytic data. Further, `orchaRd 2.0` improves reporting transparency in a meta-analysis by following the 'Preferred Reporting Items for Systematic reviews and Meta-Analyses in Ecology and Evolution' (PRISMA-EcoEvo; O'Dea *et al.* 2021). Importantly, our package's vignette provides detailed instructions and examples on how to use all the main functions (https://daniel1noble.github.io/orchaRd/).

2 | SURVEY

To gauge the potential usefulness of the `orchaRd` package's extensions, we surveyed 102 meta-analyses in ecology and evolution. Notably, this dataset was initially collected to quantify reporting quality of ecological and evolutionary meta-analyses to assist in creating PRISMA-EcoEvo (O'Dea *et al.* 2021). Briefly, we obtained 102 articles with meta-analyses that were published between 1

January 2010 and 25 March 2019 and part of the “Ecology” and “Evolutionary Biology” journals classified under the InCites Journal Citation Reports (Clarivate Analytics) (see more details in O'Dea *et al.* 2021). We previously explored this dataset to survey the use of forest and forest-like plots in ecology and evolution (Nakagawa *et al.* 2021).

For this study’s survey, we asked the following 10 questions.

Q1: How many papers have at least one categorical variable/moderator? (Defining a moderator as a predictor in a meta-regression analysis)

Q2: How many papers have at least one test or model for heteroscedasticity?

Q3: How many papers have at least one model with more than one categorical moderator?

Q4: How many papers have at least one model with at least one categorical moderator and one continuous moderator?

Q5: How many papers that used a multi-moderator regression have at least one forest-like plot (figure) made from the multi-moderator meta-regression?

Q6: How many papers that used a multi-moderator regression also modelled interactions?

Q7: How many papers, which use *R*, cite an *R* software package they used for meta-analysis?

Q8: How many papers, which use *R*, cite an *R* software package they used for the graphical presentation of meta-analytic results?

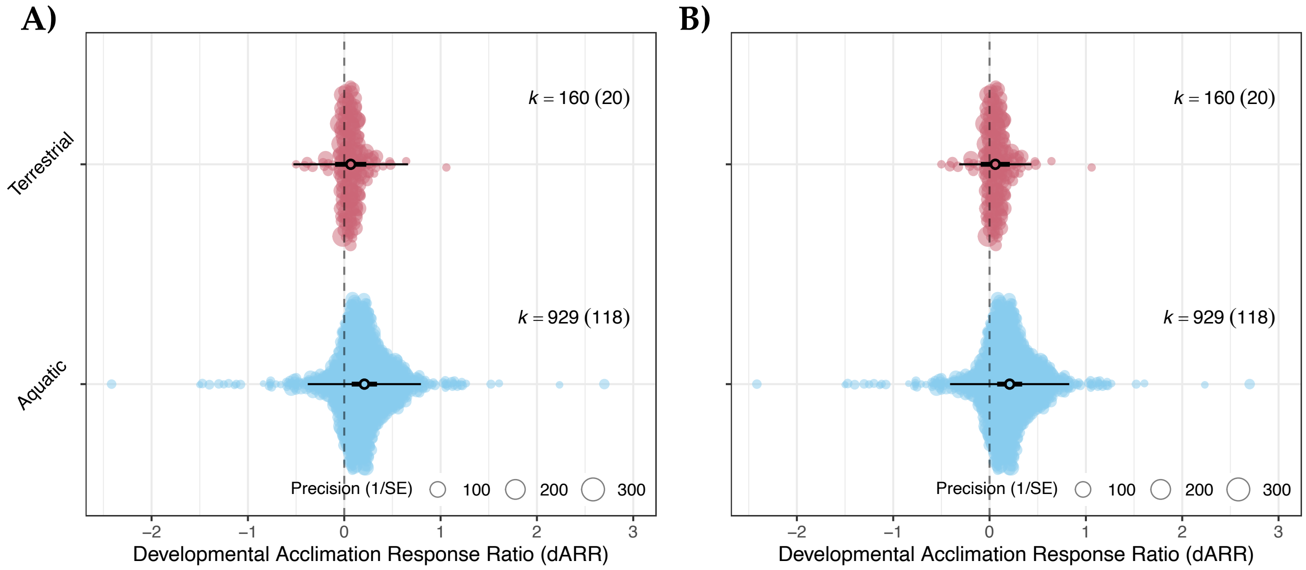
We report relevant results below, but full results of this survey can be found in the Supplementary Information.

3 | NEW SOFTWARE CAPABILITIES

The `orchaRd 2.0` package has six main functions with three different functionalities: mod_results (creating a table or a ‘table’ function – See Fig 1 `mod_results` tables), `orchard_plot` (a figure function), `bubble_plot` (a figure function), `catepillars` (a figure function), `i2_ml` (calculating I^2 statistics or a statistics function) and `r2_ml` (statistics; each function's description is found in Table 1). Among these six functions, the core function is `orchard_plot`. This function enables users to draw orchard plots from a table created by `mod_results`, which uses `emmeans` functionality (Lenth *et al.* 2018) to process `metafor` model objects (object classes: `rma`, `rma.mv`, and `robust.rma`; Viechtbauer 2010). Below we showcase three new capabilities of `orchard_plot`. Then, we describe the other main functions. Notably, the focus of our `orchaRd` package is to visualise multilevel meta-analytic models (i.e., multiple effect sizes per study).

3.1 | Orchard plots: heteroscedasticity

Categorical variables (moderators) are extremely common in meta-analyses. In our survey, > 97% of the papers had at least one categorical variable. The categorical variable was used to subset data for sub-group analyses, where a series of meta-analyses (intercept models) were run, or to fit a meta-regression model (Q1). In many meta-analyses, researchers assumed all levels of a categorical moderator had the same variation (homoscedasticity). Our survey shows that only 5% of papers investigated heteroscedasticity while others assumed homoscedasticity (Q2). Yet, differences in variances can be as biologically insightful as differences in means among groups. For example, Pottier et al. (2022) found that not only were aquatic ectotherms more thermally plastic than their terrestrial counterparts, but their plastic responses were much more variable than those of terrestrial ectotherms (even after taking into account the sample size difference). Our `orchard_plot` now allows for visualization of modelled heteroscedasticity by depicting different prediction intervals (PIs) for different groups (Fig 1). Of importance, modelling heteroscedasticity, when it exists, reduces Type 1 error (Rubio-Aparicio *et al.* 2017; Rubio-Aparicio *et al.* 2020); and orchard plots can assist meta-analysts in finding heteroscedasticity. Incidentally, modelling heteroscedasticity for a categorical moderator becomes essential if one wants to obtain absolute group means (e.g., selection gradients; Kingsolver *et al.* 2012; Siepielski *et al.* 2017; see also Noble, Stenhouse & Schwanz 2018). Absolute estimates can be calculated assuming a 'folded' normal distribution (see Morrissey 2016; Nakagawa & Lagisz 2016), with the accuracy of mean magnitudes being dependent on within-group variances. As such, it is important that heteroscedasticity is evaluated if such an approach is taken.



Name	Mean	L 95% CI	U 95% CI	L 95% PI	U 95% PI
Aquatic	0.21	0.08	0.34	-0.38	0.80
Terrestrial	0.07	-0.09	0.23	-0.53	0.66

Name	Mean	L 95% CI	U 95% CI	L 95% PI	U 95% PI
Aquatic	0.21	0.08	0.34	-0.41	0.83
Terrestrial	0.06	-0.09	0.21	-0.32	0.43

Figure 1| Orchard plots (using `orchard_plot` function) and model result tables (using `mod_result` function) for terrestrial and aquatic ectotherm developmental acclimation response ratios (dARR). **A)** Model assuming the variance in terrestrial and aquatic ectothermic species is the same (i.e., homogeneity of variance); **B)** Model assuming the variance in terrestrial and aquatic ectothermic species is not the same (i.e., heterogeneity of variance), with the lower and upper confidence intervals (CIs) and prediction intervals (PIs) adjusted accordingly for each level of the habitat type moderator (data from Pottier *et al.* 2022).

3.2 | Orchard plots: marginal means

Many meta-analyses include multiple variables (moderators), and often they are modelled together in a single meta-regression model. In our survey, meta-analytic studies often modelled two or more categorical moderators together (Q3: 41%) and modelled at least one categorical moderator and one continuous moderator (Q4: 30%). Not all meta-analyses, which had multi-moderator models, reported marginal estimates (Q5: 27%). It is understandable because obtaining 'marginal' means becomes difficult once the number of moderators increases unless one relies on computational solutions, for example, via the `emmeans` package. Therefore, many meta-analysts have been using only estimates from uni-moderator models. We have now made it straightforward to produce marginal means from a multi-moderator meta-regression model using `orchard_plot`. It is notable that marginalisation is usually done by weighting in proportion to the frequencies in the sample (data) of different groups that are averaged over. In such a case, marginal means are often similar, if not identical, to means from a uni-moderator model. However, if 'equal' weighting is used (giving the same weights to all groups), marginalised means could be different from those from a uni-moderator model, especially when a categorical moderator is unbalanced between groups/levels (Fig 2). Equal weighting is, for example, useful when your sample is unequal in your dataset, but in the population, it should be ~50:50 %; for example, males and females in many animals (cf. Deffner, Rohrer & McElreath 2022).

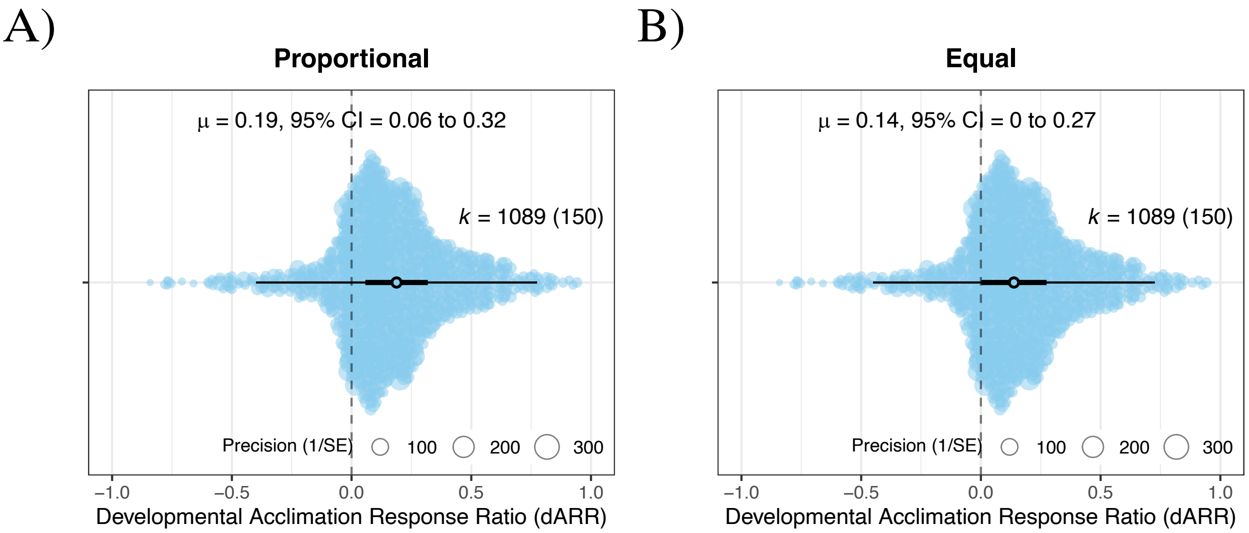


Figure 2| Orchard plots of overall meta-analytic mean developmental acclimation response ratios (dARR). **A)** Marginalised mean estimate assuming aquatic and terrestrial ectotherms are weighted proportionally to their representation in the sample of data (see Fig. 1 for sample sizes for each group); **B)** Marginalised mean estimate assuming aquatic and terrestrial ectotherms are weighted equally. Comparing the mean and 95% confidence intervals shows how estimates affect the mean and the inferences.

3.3 | Orchard plots: conditional means

As mentioned above, our survey showed that it was not uncommon to have a study with a continuous moderator and a categorical moderator (Q3: 30%). For such a combination one can estimate group-level means (and overall means) conditioned upon specific values of a continuous moderator (Fig 3). For example, O'Dea et al. (2019) estimated how thermal environments during development affect phenotypic mean and variance. They found that increasing temperature did not change phonotypic means, while phenotypic variance increased as developmental temperature increased. Examining 'conditional' means is illuminating and important for statistical inference because the statistical significance of conditional estimates can change along the gradient of a continuous moderator. Yet, none of the 32 papers with a model containing at least one categorical and continuous moderator presented conditional estimates, as for example are depicted in Fig 3B (see also Vendl *et al.* 2022).

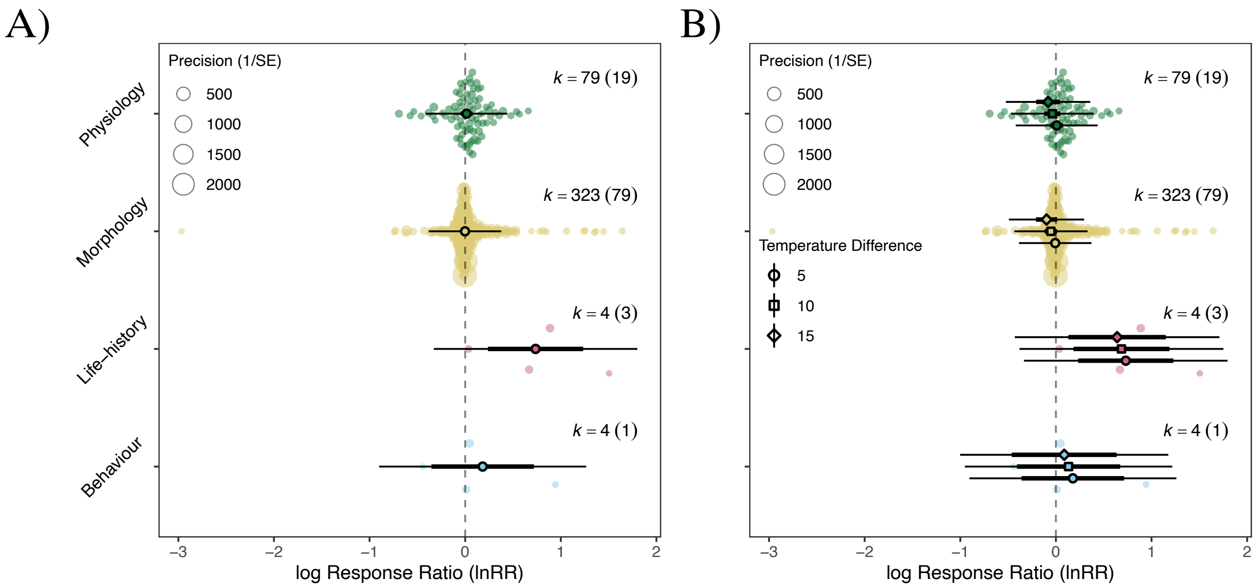


Figure 3| Orchard plots of mean log response ratios across four major trait categories (Physiology, Morphology, Life-history and Behaviour) in fish. **A)** Overall meta-analytic mean log response ratio for each trait category; **B)** Predicted overall meta-analytic mean for each trait category for three levels of temperature difference, 5, 10, and 15 degrees (data from O'Dea *et al.* 2019).

3.4 | Interactions: orchard, bubbles and bubbleless

In our survey, ~30 (out of 102) meta-analyses modelled some type of interaction (Q5). Three types of interactions might manifest in a meta-analysis, those between: 1) categorical × categorical variables; 2) categorical × continuous variables; and 3) continuous × continuous variables. The first type of interaction (categorical × categorical) can be easily visualised using an orchard plot because interactions between two categorical variables can be conceptualised as one categorical variable (e.g., a categorical variable with 2 levels × one with 3 levels is equivalent to a categorical variable with 6 levels; Fig 4A). If we want to see a plot with the second type (categorical × continuous), one can use bubble plots via the `bubble_plot` function (note that `metafor` also has a function for bubble plots, called `regplot`, which provides a single-panel interaction plot, unlike our multi-panel interaction plots; Fig 4B). The third type (continuous × continuous) is the least intuitive one to visualize, but one can also use `bubble_plot` to draw ‘bubbleless’ plots, which are line plots with multiple panels (Fig 4C); they are bubbleless because often there are only a few or no corresponding data points to plot for a given point of one of the two continuous variables.

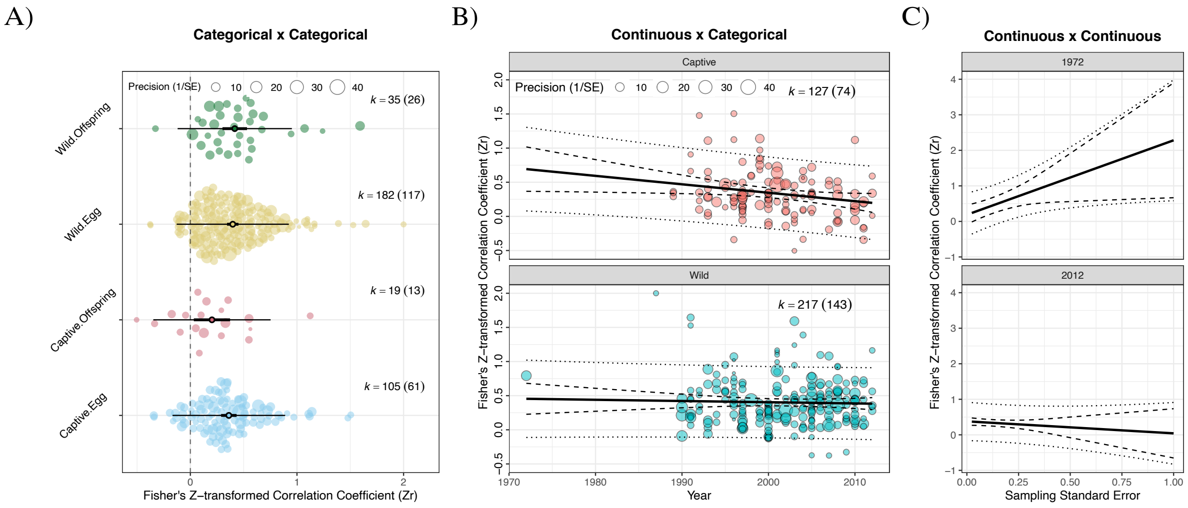


Figure 4| Orchard plots and bubble plots using Fishers’ *z*-transformed correlation coefficient (Zr). **A)** Example of two categorical moderators (Captive *vs.* Wild; Offspring *vs.* Egg) combined into a single moderator. **B)** Example of a continuous moderator (Publication Year) combined with a categorical moderator (Captive vs Wild); **C)** Two continuous moderators, Sampling Standard Error and Year, with predictions made for two years (data from Lim, Senior & Nakagawa 2014). In B and C, dashed lines represent 95% confidence intervals while dotted lines 95% prediction intervals.

3.4 | Other functions

In addition to orchard and bubble plots, the `orchaRd` package provides ‘caterpillar’ plots (via the function `catepillars`, which is a forest plot without labels for each effect size; see our vignette – https://daniel1noble.github.io/orchaRd/). We also present two new non-plot functions to give meta-analysts convenient tools to quantify heterogeneity and variances explained by multilevel meta-analyses. The function `i2_ml` calculates I^2, which is the percentage of variation among effect sizes not driven by sampling error (much of which is due to differences in sample sizes across studies; Higgins & Thompson 2002). Our function not only calculates the original I^2 (referred to as 'total' I^2) but heterogeneity explained by each additional random effect in the model (e.g., heterogeneity due to study ID or due to species ID) (*sensu* Nakagawa & Santos 2012). Furthermore, different sets of I^2 values can be calculated for different groups (levels) for a categorical moderator model with heteroscedasticity. While I^2 is estimated from a meta-analytic (intercept-only) model, R^2 is used to quantify variance (heterogeneity) accounted by moderators. The function `r2_ml` calculates marginal R^2, proposed by Nakagawa & Schielzeth (2013) as a pseudo-R^2 for linear mixed-effects models. Notably, both `i2_ml` and `r2_ml` can provide 95% confidence interval(s), using bootstrapping.

4 | IMPROVING REPORTING

4.1 | PRISMA-EcoEvo and orchaRd

O'Dea et al. (2021) recommends information to be reported in systematic reviews and meta-analyses in ecology and evolution. Visualisations from the orchard package are completely consistent with reporting recommendations of PRISMA-EcoEvo. This is especially so with three (sub-)items, recommended for the Method section: 1) presenting the numbers of studies and effect sizes for each estimate; 2) reporting indicators of heterogeneity; and 3) including estimates and confidence intervals for moderators. The survey conducted by O'Dea et al. (2021) showed very poor reporting of these items: 57%, 52% and 59%, respectively. As one can see, our package takes care of these three items in a single orchard plot (Fig. 1-4). It is notable that now orchard plots even visualise different heterogeneities among different groups (i.e., heteroscedasticity) via prediction intervals, PIs.

4.2 | Plea and proposal

Graphical presentation can facilitate better reporting in meta-analyses. However, in our survey, only 2 papers (3.1%) out of 64 articles which used *R*, cited any graphical package(s) used for visualizing meta-analytic results (e.g., `orchaRd`; Q8). This figure starkly contrasts with 85% of the papers (55 out of 64; Q7) citing the software packages used for meta-analyses (e.g., `metafor`). This survey result marks a severe under-recognition of graphical packages. The real-world risk here is that this lack of recognition severely disincentivises developers from maintaining and further developing graphical packages.

We argue that authors should acknowledge graphical packages used for presenting meta-analyses (or any research article, for that matter), just as they do with any statistical package. We propose that graphical packages that were used to make a figure should be listed at the end of the figure legend. This standardised reporting format will mean packages do not necessarily need to be listed in the methods, but they will still be given credit. We note, however, that an *R* package can have many dependencies (i.e., other required *R* packages other than 'base' packages). For example, `orchaRd 2.0` is dependent on `emmeans`, `ggplot2` and `metafor`. We freely admit that we do not have a satisfying answer on whether dependencies should also be credited. However, for now we think it is reasonable to suggest that researchers provide the reference (in a figure legend or main text) for the immediate *R* function and package they used to make their figure.

5 | CONCLUSIONS

As the presence and influence of meta-analyses grow in the field, it is more important than ever to visualize meta-analytic results in an information-rich manner. Here, we have introduced an expanded version of `orchaRd` (version 2.0), which enables researchers to readily visualize complex as well as simple meta-analytic results, a task that was previously difficult for many. New functionalities that allow for marginal and conditional means to be plotted will improve model communication by allowing for a holistic visual interpretation of the complex numerical information generated by the analysis (see Fig 1-4). Also, we introduce functions for calculating I^2 and R^2 for multilevel meta-analytic models, which have become standard in ecological and evolutionary meta-analyses. Finally, we hope our paper also becomes a reminder of the importance of acknowledging graphical packages. Adequate attribution of credits will create a more sustainable environment for developers and maintainers of graphical packages.

ACKNOWLEDGEMENTS

We thank Wolfgang Viechtbauer for his continuous effort to maintain and develop an amazing package, `metafor`. We also, acknowledge that Russell Lenth made his wonderful package, `emmeans` compatible with `metafor`. SN and ML were supported by an ARC (Australian Research Council) Discovery Grant (DP210100812).

SN and DWAN conceived the initial idea and wrote the first draft. DWAN and SN led programming and implementations from the inputs from REO, AMS and PP. ML, JR and YY conducted the survey. All authors contributed to the design of the study and to editing and commenting on drafts.

CONFLICT OF INTEREST

The author reported no conflict of interest

ORCID

Shinichi Nakagawa: 0000-0002-7765-5182
Malgorzata Lagisz: 0000-0002-3993-6127
Rose E. O’Dea: 0000-0001-8177-5075
Patrice Pottier: 0000-0003-2106-6597
Joanna Rutkowska: 0000-0003-0396-1790

Yefeng Yang: 0000-0002-8610-4016

Alistair M. Senior: 0000-0002-7765-5182

Daniel W. A. Noble: 0000-0001-9460-8743

REFERENCES

Aloe, A.M., Becker, B.J. & Pigott, T.D. (2010) An alternative to R-2 for assessing linear models of effect size. *Research Synthesis Methods,* **1,** 272-283.

Cheung, M.W.L. (2014) Modeling Dependent Effect Sizes With Three-Level Meta-Analyses: A Structural Equation Modeling Approach. *Psychological Methods,* **19,** 211-229.

Cooper, H.M., Hedges, L.V. & Valentine, J.C. (2019) *The handbook of research synthesis and meta-analysis,* 3rd ed. edn. Russell Sage Foundation, New York.

Deffner, D., Rohrer, J.M. & McElreath, R. (2022) A Causal Framework for Cross-Cultural Generalizability. *Advances in Methods and Practices in Psychological Science,* **5**.

Gurevitch, J., Koricheva, J., Nakagawa, S. & Stewart, G. (2018) Meta-analysis and the science of research synthesis. *Nature,* **555,** 175-182.

Higgins, J.P.T. & Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine,* **21,** 1539-1558.

Higgins, J.P.T.e., Thomas, J.e., Chandler, J.e., Cumpston, M.e., Li, T.e., Page, M.J.e. & Welch, V.A.e. (2019) *Cochrane handbook for systematic reviews of interventions,* Second edition. edn. Wikey, Chichester, UK.

Kingsolver, J.G., Diamond, S.E., Siepielski, A.M. & Carlson, S.M. (2012) Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology,* **26,** 1101-1118.

Lenth, R., Singmann, H., Love, J., Buerkner, P. & Herve, M. (2018) Emmeans: Estimated marginal means, aka least-squares means. *R package version,* **1,** 3.

Lim, J.N., Senior, A.M. & Nakagawa, S. (2014) Heterogeneity in Individual Quality and Reproductive Trade-Offs within Species. *Evolution,* **68,** 2306-2318.

Morrissey, M.B. (2016) Meta-analysis of magnitudes, differences and variation in evolutionary parameters. *Journal of Evolutionary Biology,* **29,** 1882-1904.

Nakagawa, S. & Lagisz, M. (2016) Visualizing unbiased and biased unweighted meta-analyses. *Journal of Evolutionary Biology,* **29,** 1914-1916.

Nakagawa, S., Lagisz, M., O'Dea, R.E., Rutkowska, J., Yang, Y.F., Noble, D.W.A. & Senior, A.M. (2021) The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond. *Research Synthesis Methods,* **12,** 4-12.

Nakagawa, S. & Santos, E.S.A. (2012) Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology,* **26,** 1253-1274.

Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution,* **4,** 133-142.

Noble, D.W.A., Stenhouse, V. & Schwanz, L.E. (2018) Developmental temperatures and phenotypic plasticity in reptiles: a systematic review and meta-analysis. *Biological Reviews,* **93,** 72-97.

O'Dea, R.E., Lagisz, M., Hendry, A.P. & Nakagawa, S. (2019) Developmental temperature affects phenotypic means and variability: A meta-analysis of fish data. *Fish and Fisheries,* **20,** 1005-1022.

O'Dea, R.E., Lagisz, M., Jennions, M.D., Koricheva, J., Noble, D.W.A., Parker, T.H., Gurevitch, J., Page, M.J., Stewart, G., Moher, D. & Nakagawa, S. (2021) Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biol Rev Camb Philos Soc,* **96,** 1695-1722.

Pottier, P., Burke, S., Zhang, R.Y., Noble, D.W.A., Schwanz, L.E., Drobniak, S.M. & Nakagawa, S. (2022) Developmental plasticity in thermal tolerance: Ontogenetic variation, persistence, and future directions. *Ecology Letters,* **25,** 2245-2268.

Rubio-Aparicio, M., Lopez-Lopez, J.A., Viechtbauer, W., Marin-Martinez, F., Botella, J. & Sanchez-Meca, J. (2020) Testing Categorical Moderators in Mixed-Effects Meta-analysis in the Presence of Heteroscedasticity. *Journal of Experimental Education,* **88,** 288-310.

Rubio-Aparicio, M., Sanchez-Meca, J., Lopez-Lopez, J.A., Botella, J. & Marin-Martinez, F. (2017) Analysis of categorical moderators in mixed-effects meta-analysis: Consequences of using pooled versus separate estimates of the residual between-studies variances. *British Journal of Mathematical & Statistical Psychology,* **70,** 439-456.

Schild, A.H.E. & Voracek, M. (2015) Finding your way out of the forest without a trail of bread crumbs: development and evaluation of two novel displays of forest plots. *Research Synthesis Methods,* **6,** 74-86.

Schmid, C.H.e., Stijnen, T.e. & White, I.e. (2021) *Handbook of meta-analysis,* 1st edn. CRC, Boca Ranton.

Senior, A.M., Grueber, C.E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E.S.A. & Nakagawa, S. (2016) Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology,* **97,** 3293-3299.

Siepielski, A.M., Morrissey, M.B., Buoro, M., Carlson, S.M., Caruso, C.M., Clegg, S.M., Coulson, T., DiBattista, J., Gotanda, K.M., Francis, C.D., Hereford, J., Kingsolver, J.G., Augustine, K.E., Kruuk, L.E.B., Martin, R.A., Sheldon, B.C., Sletvold, N., Svensson, E.I., Wade, M.J. & MacColl, A.D.C. (2017) Precipitation drives global variation in natural selection. *Science,* **355,** 959-+.

Vendl, C., Pottier, P., Taylor, M.D., Braunig, J., Gibson, M.J., Hesselson, D., Neely, G.G., Lagisz, M. & Nakagawa, S. (2022) Thermal processing reduces PFAS concentrations in blue food- A systematic review and meta-analysis star. *Environmental Pollution,* **304**.

Viechtbauer, W. (2010) Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software,* **36,** 1-48.

Wickham, H. (2009) *ggplot2 : elegant graphics for data analysis*. Springer, New York ; London.

Wilson, L.A.B., Zajitschek, S.R.K., Lagisz, M., Mason, J., Haselimashhadi, H. & Nakagawa, S. (2022) Sex differences in allometry for phenotypic traits in mice indicate that females are not scaled males. *Nature Communications,* **13,** 7502.

Zajitschek, S.R.K., Zajitschek, F., Bonduriansky, R., Brooks, R.C., Cornwell, W., Falster, D.S., Lagisz, M., Mason, J., Senior, A.M., Noble, D.W.A. & Nakagawa, S. (2020) Sexual dimorphism in trait variability and its eco-evolutionary and statistical implications. *Elife,* **9**.

TABLE

Table 1 Main functions in the orchard package, their general categorisation and a description of what they can be used for in combination with `metafor` meta-analytic model objects (`rma.mv`, `rma` and `robust.rma`)

Function	**Category**	**Description**
`mod_results`	Table	`mod_results` takes multi-level meta-analytic and meta-regression models (with multiple moderators – continuous or categorical) of class `rma.mv/rma/robust.rma` and calculates mean or marginalised mean meta-analytic estimates across all levels of a given moderator or overall (i.e., intercept only). The mod_results table can then be used with `orchard_plot`, `bubble_plot,` or `catepillars` to plot results graphically. If a multivariate meta-regression model (with many moderators) is provided, users can specify the '`by`' and/or '`at`' arguments to marginalise over desired levels of other moderators.
`orchard_plot`	Figure	Modified forest plot that plots the meta-analytic means, confidence intervals, prediction intervals and raw data for each level of a categorical moderator. Users can use a number of arguments for

		modifying the look of plots including the legend, colour schemes, size and weight of points and lines and angle and naming of text on the axes. Sub-setting allows the users to plot a subset of the levels for a given moderator. Additional modifications can be made by adding and modifying layers of the `ggplot` object. Plots can be made using either `mod_results` objects directly or using the `rma.mv/rma/robust.rma` model object in combination with the raw data. If a multivariate meta-regression model (many moderators) is provided directly users can specify the '`by`' and/or '`at`' arguments to marginalise over desired levels of other moderators.
`bubble_plot`	Figure	Creates a bubble plot(s) depicting the predicted mean effect size, confidence and prediction interval as a function of a continuous moderator (slope estimate) or a series of separate plots showing predictions across an additional moderator (i.e., interaction plots). Plots can be made using either `mod_results` objects directly or using the `rma.mv/rma/robust.rma` model object in combination with the raw data. Raw data is plotted, and point size is adjusted according to effect size precision.
`catepillars`	Figure	Creates a caterpillar plot from an intercept model or from mean effect size estimates for all levels of a given categorical moderator, their corresponding confidence and prediction intervals. Plots can be made using either `mod_results` objects directly or using the `rma.mv/rma/robust.rma` model object in combination with the raw data.

`i2_ml`	Statistics	Calculates heterogeneity statistics using measures of I^2 for a multilevel meta-analytic or meta-regression models. Point estimates can be calculated quicky for each level of random effect along with an estimate of total heterogeneity. Users also have the option of generating 95% confidence intervals for all I^2 estimates using the `` `boot` `` argument (percentile method). This argument will conduct parametric bootstrapping.
`r2_ml`	Statistics	Calculates marginal and conditional R^2 for multilevel meta-analytic or meta-regression models. Point estimates can be calculated quicky using a couple of different methods, but users also have the option of generating 95% confidence intervals for R^2 using the `` `boot` `` argument (percentile method). This argument will conduct parametric bootstrapping.