**Meta-analytic approaches and new effect sizes to account for treatment differences across studies in comparative physiology**

Daniel W.A. Noble, Patrice Potter, Sammy Burke, Szymek Drobniak, Rose O’Dea, Shinichi Nakagawa

**Abstract**:

Meta-analysis is a tool that provides comparative physiologists with powerful, unbiased and quantitatively informed answers to topical research questions by synthesising effect sizes across disparate studies. Distilling published research results into standardised effect sizes that can be weighted by their sample size (i.e., inverse sampling variance) and compared across broad sets of research designs, study systems and species is its core objective and strength. Estimating overall effect sizes, and understanding what drives effect variability, provides opportunities to model how organisms will respond to pressing global challenges. Despite this ambition, research designs in comparative physiology can appear, at the outset, as being vastly different to each other (e.g., using different temperatures or treatment dosages). Differences in treatments across studies has led many to believe that meta-analysis is an exercise in comparing “apples with oranges”. Here, we dispel this myth by showing how standardised effects sizes can be used in conjunction with powerful multi-level meta-analytic models to both account for factors driving differences across studies and make them more comparable. In addition, we derive new effect size measures that provide comparative physiologists with a means to directly make effect sizes comparable without the need to resort to more complex statistical models that may be harder to interpret. Our ‘new’ effect sizes and corresponding sampling variances help physiologists deal with temperature and dosage differences across studies; allowing researchers to compare both mean differences (e.g., Q10 to compare mean differences in physiological rates over 10 degrees) and associated differences in variance (e.g, SQ10 for comparing variability in in physiological rates over 10 degrees) for physiological traits. The new effect sizes we propose, in combination with existing meta-analytic models, will pave the way for comparative physiologists to explore exciting new questions by making results from large-scale data sets from the literature more accessible and widely interpretable.

**Proposed Outline:**

**Introduction**

Introduction will introduce meta-analysis as the gold standard for comparing effects across studies, discuss how it works, and the importance of having a standardised (making sure different traits / units can be compared) and comparable effect size measure along with an associated sampling variance. Sampling variances are critical because they are used to effectively weight effect sizes so that overall conclusions / evidence for a treatment effect is based on the ‘quality’ of the study (i.e., studies with high power are given greater weight). Often sampling variances are not used in meta-analyses in comparative physiology.

**Common effect sizes for comparing treatment effects**

This section will introduce different standardised effect size measures currently commonly used such as the log response ratio (lnRR), standardised mean differences (SMD), log odds ratio (logOR) and regression slopes. We’ll look at the relationship between these and how they ca be used in conjunction with currently used physiological measures (e.g., CTmax, Q10). Currently used measures are effect sizes but currently lack a measures sampling variance limiting the use of powerful meta-analytic models and approaches in comparative physiology . We will also, more briefly, introduce how meta-analysis can also analyse changes in variance in, say, physiological rates. This can be important in combination with mean differences as variances can provide insight into the capacity of system to evolve.

**Challenges when using common effect sizes to compare across studies**

Despite common effect sizes being powerful ways to standardised units across studies, and account for other study specific effects, this is not perfect. For example, treatments across studies can use wildly different temperatures or dosages. These factors are not currently standardised in the process of deriving effect sizes and need to be dealt with through others means. This only feeds the incorrect notion that meta-analysis often compares “apples with oranges”, dampening confidence in conclusions. Despite this limitation, there simple statistical solutions to better account for differences across studies.

**Statistical approaches to account for continuous differences among treatments**

We will introduce briefly multi-level meta-analytic /meta-regression models that allow researchers to control for various sources of non-independence (i.e., effects from the same study, species etc) and explore key factors hypothesized to drive effect size heterogeneity (i.e., variance). A simple way to resolve the “apples and oranges” problem is to simply extract relevant data (e.g., dosage or temperature differences) and include these in multi-level meta-regression models. When scaling (z-transforming or centring) these coefficients the overall meta-analytic mean can be interpreted as the mean for an “average” dosage or temperature across studies. We will show how to apply these models in a supplemental material file that will provide all the code and explanations. We will also briefly touch on what are often referred to as “arm-based” meta-analytic model (these use only the mean and sd for a given treatment), which can be equivalent to models outlined above. Arm-based models can also be a solution, but we caution that the result from these models need for more careful interpretation and model fitting is more complex.

**New effect sizes to account for continuous treatment differences**

Sometimes we may want an effect size that is simply interpreted on a standardised scale with respect to how treatments vary continuously. This makes modelling simpler. In addition, it can provide improve power when sample sizes are small because estimating model coefficients requires large enough sample sizes. This section will introduce new effect sizes making an equivalence between common effect size measures described above and these new ones and how simple extensions can be applied. We will feature more heavily on a commonly used and well understood effect size measure, Q10, but we will show how this can be applied in a similar fashion to control dosage effects or other effects that might vary across treatments. We have already derived Q10 effect sizes, along with variance measures. We will do the same for dosage.

**Assumption, Limitations and Uncertainties**

This section will overview some of the assumptions inherent to using these effect sizes, their limitations and discuss some new paths forward for meta-analysis in comparative physiology.

**Conclusions**

Discuss how the use of large-meta-analytic datasets that make use of data derived from empirical research will be ever more important into the future. Meta-analysis provides the means by which science can summarise quantitative evidence in support or against a key hypothesis. Such quantitative evidence has the potential to shape the field or inform on overall effects that can be applied when modelling climate or anthropogenic impacts on organisms globally.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Supplemental Files**

We plan on having a series of “how to” guides that implement the modelling approaches, show readers how to derive effect sizes and provide them with functions and code that they can adopt in their own research to answer whatever question they are interested in tackling when using meta-analytic data. This will be open source, freely accessible and beautifully formatted in R markdown. Very importantly, it will also show readers how to interpret core parameters from different models and using different effect sizes.

**Tentative Figures and Tables**

We will have a table summarise the different effect size measures and their corresponding sampling variances. This is a condensed way to provide these formulas. We will probably have a few figures similar to Nakagawa et al. 2017. BMC Biology DOI 10.1186/s12915-017-0357-7 (e.g., Figure 4c) to demonstrate multi-level modelling in a visual way. We will also have a figure that demonstrates how effect sizes are often taken from studies, to make the point of how temperature and dosage can vary.