One of the most useful tools that facilitates succinct and understandable data summaries, comparison of results across studies and future power analysis are standardised effect sizes. Although statistical significance tests are often presented as the primary tests of a scientific theory, equally if not more important is the degree of difference or strength of effect that was observed. Often, the most efficient way of expressing the size of the effect will be to use a standardised effect size (e.g., Cohen’s d, partial , R2). These effect size benchmarks have become extremely tightly engrained in much of the discussion around effect sizes, especially in psychology. Many of the most commonly used benchmarks to understnaind and planning effect sizes in psychology are based on Cohen’s (1988) benchmarks. This is true despite the fact that Cohen never suggested using his benchmarks as anything but a last resort in power analysis or effect size interpretation.

E.g.; “When the above has not provided sufficient guidance, the reader has an additional recourse. For each statistical test's ES index, the author proposes, as a convention, ES values to serve as operational definitions of the qualitative adjectives" small,"" medium," and" large." This is an operation fraught with many dangers … they run a risk of being misunderstood.”, Cohen, 1988, p.12.

And despite the fact that he selected these benchmarks in a somewhat ad hoc manner e.g., “‘Small’ effect sizes must not be so small that seeking them amidst the inevitable operation of measurement and experimental bias and lack of fidelity is a bootless task, yet not so large as to make them fairly perceptible to the naked observational eye… In contrast, large effects must not be defined as so large that their quest by statistical methods is wholly a labor of supererogation.” Cohen, 1988, p.13.

This project uses a big data and text mining approach to extract effect sizes reported in standard formats in over in over two million open access articles in the PubMedCentral open access subset, and presents an online tool for accessing, visualising and downloading this dataset. This does not provide a method for estimating the expected outcome of a particular area of psychological research, just of visualising the reported effects in an area of psychology research. It can be used as heuristic guide to understanding effect sizes, but for specific estimates of the expected effect sizes from particular experiments, researchers should look to meta-analyses or other large scale replication curation projects like curatescience.org (LeBel, McCarthy, Earp, Elson, & Vanpaemel, 2018).

Use cases:

As a replacement last resort to the benchmarks seen in psychology as a whole

Allows people to quickly gather benchmarks from an area of psychology research instead of relying on Cohen’s benchmarks

As a more subtle tool to provide some idea of the effect sizes seen in particular areas of psychology

As an educational tool, allowing quick visualisation and exploration of the reported effects in different areas of psychological research

As a side effect of the relatively regular way in which these effect sizes are reported, it is possible for a text-search algorithm to be built to collect and dis

it is possible to perform relatively simple operations to extract them from previously written scientific papers, such as

Extracted values:

Partial eta squared

partial eta squared is estimated from F statistic and dfs as

It is possible to exclude or include the values as desired.

Include an “exclude all values greater with a probability of occurring < .001 given that the mean and SD is true and a normal SD” as an anomaly detection method?

Limitations:

It should be noted, this project is limited in several ways. One, although the number of articles included in this analysis is large (over 2,000,000 articles), they cannot be said to comprehensively cover all of psychology and biomedicine.

It is possible that open access jourals are in some way systematically different from the population of all journals.

The way that the text search function works means that there are definitely errors in this database. Although no false positives were identified in the validation stage of this project, it is possible that some people use one of the same sybols to represent another feature of their dataset, in which case it would appear in this database as a reported effect size. Secondly, this search mechanism does not pick up on

Previous efforts along the same lines:

In order to provide an overview of the effects reported in varied fields, serval previous efforts have attempted to extract and report on the effects that have been seen in various fields of psychology research.

Previous attempts to extract effect sizes reported in the scientific literature (e.g., Haase, Waechter, & Solomon, 1982; Quintana, 2016) have focused on particular sub-fields of psychology research and have not presented their data in a way that can easily be accessed and explored readers. Two previous studies have taken a text-analysis or big data approach to answering this question. The first, Szucs and Ioannidis (2017), focus on articles published in 18 journals frequently cited in cognitive neuroscience and psychology published between Jan 2011 to Aug 2014, not examining as large or broad a population.

Second, Bosco, Singh, Aguinis, Field, and Pierce (2015) examine the first correlation table presented in every article published in the Journal of Applied Psychology or Personnel Psychology from 1980 to 2010. Bosco et al., (2015) include the relation types (e.g., “attitudes-performance” or “behaviour-intentions”) and make their data publicly available. Bosco, Uggerslev, and Steel (2017) have continued this approach with the METABus project. The METABus project analyses the first correlation table reported in articles using a combination of text-extraction and a manual checks by a team of over 25 research assistants. Although their approach, which includes a manual check of every data-point, leads to higher-quality data than the fully automated approach of the current project, their project only assesses relations of particular type (correlations) published in a limited set of journals covering IO and management psychology.

Things to include in application:  
Search by: Author, journal, date range,

Keyword, **Optional stemmed keyword search**

Data organisation

Data is exported in a tidy data format as two files, one file including the PubMedCentral ID, DOI, Journal, Volume and Issue numbers, Publication Date, Article Title, and Authors, and the other file including the effect sizes of each type. The two files are keyed on the PMCID columns.

Table [Export file 1]

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PMCID | DOI | Journal | Issue | Volume | Publication Date | Article Title | Authors | Keywords |
| 4547492 | 10.1016/j.jbtep.2015.05.001 | Journal of Behavior Therapy and Experimental Psychiatry | NA | 49 | 2015.12.01 | Investigating the efficacy of attention bias modification in reducing high spider fear: The role of individual differences in initial bias | Elaine Fox; Konstantina Zougkou;  Chris Ashwin; Shanna Cahill | Spider phobia, Spider fear,  Attentional bias, Cognitive bias modification, Attentional training , Threat detection |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

Table [export file 2]

|  |  |  |  |
| --- | --- | --- | --- |
| PMCID | Effect size type | Effect size | Location |
| 4547492 | d | .04 | results |
| 4547492 | d | 7.5 | results |
| 4547492 | d | 3.27 | results |
| 4547492 | r | .31 | discussion |
| 4547492 | r | .06 | discussion |
| 5794850 | r | .66 | results |

Possible other sources for effect sizes:

http://olabout.wiley.com/WileyCDA/Section/id-829772.html

References

Bosco, F. A., Singh, K., Aguinis, H., Field, J. G., & Pierce, C. A. (2015). Correlational Effect Size Benchmarks. *Journal of Applied Psychology, 100*, 431-449. doi:10.1037/a0038047

Bosco, F. A., Uggerslev, K. L., & Steel, P. (2017). MetaBUS as a vehicle for facilitating meta-analysis. *Human Resource Management Review, 27*, 237-254. doi:<https://doi.org/10.1016/j.hrmr.2016.09.013>

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, New Jersey: Erlbaum.

Haase, R. F., Waechter, D. M., & Solomon, G. S. (1982). How significant is a significant difference? Average effect size of research in counseling psychology. *Journal of Counseling Psychology, 29*, 58-65. doi:10.1037/0022-0167.29.1.58

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A Unified Framework to Quantify the Credibility of Scientific Findings. *Advances In Methods and Practices in Psychological Science*, 2515245918787489. doi:10.1177/2515245918787489

Quintana, D. S. (2016). Statistical considerations for reporting and planning heart rate variability case-control studies. *Psychophysiology, 54*, 344-349. doi:10.1111/psyp.12798

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology, 15*, e2000797. doi:10.1371/journal.pbio.2000797