Doi

Journal

Year

Issue

Last name of first author

Title

Figure ID

Data availability (yes, no)

Design (2x2, 2x3, etc be consistent, 2x3 and 3x2 are same)

Sample sizes (vector of numbers, eg. 5,5,5,4)

Analysis (factorial or flat)

t-tests (combined or separate)

Tests (t, welch, mann-whitney-wilcoxan, main effect)

Comment on test

Interaction statistic reported (yes, no)

Quote reporting

Word “Interaction” used in results (yes, no)

If yes, quote the usage

Use “synergy” to describe results (yes, no)

If yes, quote the usage

If yes, were results positive (evidence for synergy) or negative (not evidence for synergy) (pos, null)

Use “antagonism” to describe results (yes, no)

If yes, quote the usage

Used interaction statistic to support synergy/antagonism (yes/no)

If no, how supported?

“A difference is significance is not necessarily significant” (yes, no)

If yes, quote the usage

Notes

1. Each experiment in a paper gets its own line, so any one paper will have multiple lines of data
2. In “Analysis”, was analysis factorial or flat
3. T-tests, did they used the combined sigma from anova/lm or separate t-test. Combined would include something like Tukey. Bonferroni is ambiguous.
4. Tests is the kind of test used in “Analysis”
5. Is p-value from main effect in anova (let’s talk about this, it’s pretty meaningless in factorial anova).
6. Ignore interactions related to pre/post or other longitudinal. (measurements over time)

Be aware of pseudoreplication

#### 15.10.3.5 Adjustment for multiple tests

set the adjust = argument to

1. “none” – no adjustment
2. “dunnettx” – [Dunnett’s test](https://en.wikipedia.org/wiki/Dunnett%27s_test" \t "_blank) is a method used when comparing all treatments to a single control. For a factorial design, this only makes sense for a flattened analysis.
3. “tukey” – [Tukey’s HSD method](https://en.wikipedia.org/wiki/Tukey%27s_range_test) is a method used to compare all pairwise comparisons.
4. “bonferroni” – [Bonferroni](https://en.wikipedia.org/wiki/Bonferroni_correction" \t "_blank) is a general purpose method to compare any set of multiple tests. The test is conservative. A better method is “holm”
5. “holm” – [Holm-Bonferroni](https://en.wikipedia.org/wiki/Holm%E2%80%93Bonferroni_method) is a general purpose method like the Bonferroni but is more powerful.
6. “fdr” – controls the [false discovery rate](https://en.wikipedia.org/wiki/False_discovery_rate) not the Type I error rate for a family of tests. One might use this in an exploratory experiment.
7. “mvt” – based on the multivariate t distribution and using covariance structure of the variables.

I shared a list of measurements and a sample spreadsheet. My original thought was one big spreadsheet. Some thoughts after doing 1 paper

1. The sections in green are fairly easy to collect. The sections in yellow are easy with a search of the terms. The non-colored sections would be very, very time consuming. I think the return-on-effort would be low

2. Lots of redundancy in this paper. I think return on effort should focus on more papers and record only a subset of experiments from each paper. Say n=2

- count number of plots with factorial design, call this max

- use unif(2, 1, max) to get the two plots to use

- use only plots in main text and not supplement.

3. Interaction p could be given in results, in the caption to the figure, or in plot itself. Sometimes interaction p is at bottom of caption. Will search on "interaction" find all interaction p? No if authors report as "treatment x genotype"

4. fig 6e,f and 7e,f are good examples of pseudoreplication ("6 sections per mouse")

How I look. --

1. I look only at the figures and search for plots that \*might be\* factorial

2. So, something with 4 (2 x 2) or 6 (3 x 2)  or 9 (3 x 3) means (uncommon for more levels than that. It's not factorial if there are only 3 or 5 although sometimes with 5 there may be a 2 x 2 + 1 going on. which counts.

3. It's not always easy to tell say a 2 x 2 factorial from a 1 x 4 (1 factor with four levels). You just have to look at the legend.

4. Once a factorial plot is identified, then you have to find where they describe how they analyzed it. This is usually in the figure caption itself, often at the end of the caption but sometime in the description of the panel itself. I think it's good to check in the text if they report the interaction. I think the fast way to do this is to search for the figure number/letter (search "3g" for figure 3g).

5. Finally check the "statistical analysis" section of the methods, which may be in a supplement. See if they state how they analyzed in general

6. we want a big heads up for any method that says analyzed with "two-way anova" but then report a p-value based on F-ratio. The two-way means that did analyze with factorial model (probably -- they could have done additive but probably not) but if all they give is a p value then it would be impossible to know where this came from. It could come from the anova table (bad) or it could come from a post-hoc test (better). I'd love to have some sense of how often a two-way is reported with the p-value from the ANOVA table.