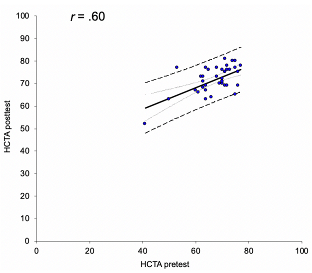
Reporting results from CORRELATION DESIGN

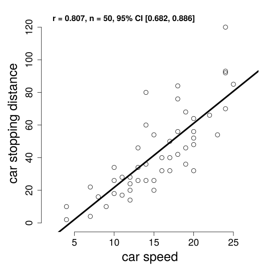
General guidance: (adapted from Cumming & Calin-Jageman, 2017)
When reporting correlations, you should usually include:

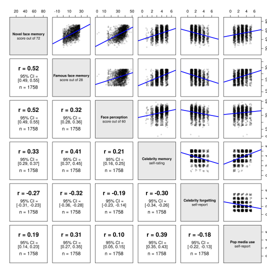
- Whether examining the correlation is planned or exploratory, unless this is already clear — be sure your research plan is more thought out than just "examine all possible correlations" or you'll likely be seeing faces in the clouds.
- Basic descriptive statistics (usually mean and SD) for both variables being correlated.
- A scatterplot or scatterplot matrix.
- The value of r and its CI. Remember that calculating a CI requires assuming the data are a sample from a bivariate normal population. If that seems problematic, make a comment and consider not reporting a CI.
- The sample size for calculating r (which, due to missing data, may not be the same as the sample size collected).
- If desired, the p value — be sure to state the null hypothesis, which is usually but not always that $\rho = 0$.
- An interpretation of the correlation that considers not only the point estimate but also the CI — consider what the full range of the CI means in terms of correlation strength, interpreted in the context of relevant past research or costs/ consequences of different effect sizes; if range restriction is evident make interpretation suitably tentative.

Model examples: (all from different studies)

Include scatterplot (left example is from ESCI; middle example is from [ShowMyData.org]) or scatterplot matrix (right example is from [ShowMyData.org]) along with a legend (not shown):







Report context and basic descriptive statistics in words:
In a planned analysis, we examined the relationship between [measure 1] (M = [mean], SD = [standard deviation]) and [measure 2] (M = [mean], SD = [standard deviation]).

Report correlation results in words (examples adapted from Cumming & Calin-Jageman, 2017):

1. The correlation between well-being and self-esteem was r = .35, 95% CI [.16, .53], N = 95, p=0.13 for a null hypothesis of $\rho = 0$. Relative to other correlates of well-being that have been reported, this is a fairly strong relationship. The CI, however, is somewhat long and consistent with anywhere from a weak positive to a strong positive relationship.
2. The correlation between well-being and gratitude was r = .35, 95% CI [−.11, .69], N = 20, p=0.13 for a null hypothesis of $\rho = 0$. The CI is quite long. These data are only sufficient to rule out a moderate to strong negative relationship between these variables.
3. The correlation between well-being and GPA was r = .02, 95% CI [−.18, .22], N = 95. The CI suggests that it is unlikely there is more than a weak positive negative relationship between these variables.

Default conventions for size of correlation:

Small/Weak:	r=.10	(for this course, use this range: 0.00 to 0.199)
Medium/Moderate:	r=.30	(for this course, use this range: 0.20 to 0.399)
Large/Strong:	r=.50	(for this course, use this range: 0.40 to 0.599, above that add adjectives like "very")

These conventions should generally be used *only until information becomes available* on one or both of: (a) **typical effect** in this area of research or (b) the **costs or consequences** of different effect sizes. For an example of the latter, Rosnow and Rosenthal (1989) discussed a study of the impact of daily aspirin vs. placebo on the dependent variable: heart attacks. At the point where number of heart attacks was substantially lower in the aspirin condition than the placebo condition, with a small CI on the difference, the study was discontinued because the effect was considered so large and important that it was no longer ethical to administer placebo. The effect size of this study: r = 0.034. A tiny effect, by conventional standards. But huge and important given the large costs and consequences of heart attacks, which include death and disability.

Specific guidelines from prior (2010) edition of APA Manual (bit.ly/apamanual_2010):
"Provide inferential statistics, including…"

- "effect-size estimates and confidence intervals ... when possible"
- "exact p values if null hypothesis significance testing (NHST) methods were used"

Specific guidelines from current (2020) edition of APA Manual (bit.ly/apamanual_2020):
"Because confidence intervals combine information on location and precision and can be directly used to infer significance levels, they are generally the best reporting strategy."

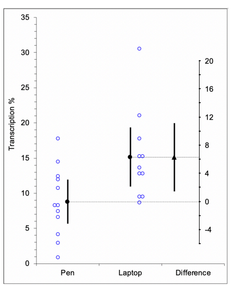
Reporting results from INDEPENDENT GROUPS DESIGN

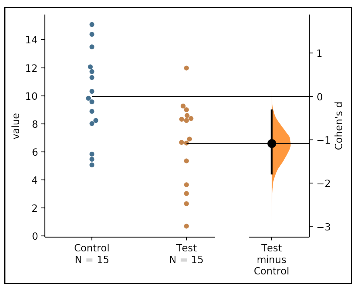
General guidance: (adapted from Cumming & Calin-Jageman, 2017)
For research questions about group differences your report will typically include:

- Whether the comparison is planned or exploratory, unless this is already clear.
- Basic descriptive statistics (usually mean and SD) for both groups.
- The group difference in raw units (M2-M1) and its CI.
- A standardized effect size estimate ($d_{unbiased}$ is best), preferably with its CI. In your Method section, make clear how your standardized effect size was calculated (which denominator was used). For independent groups design, the denominator is most typically s_{pooled}, or s_p, and this is what ESCI provides (for both d and $d_{unbiased}$). The CI that ESCI provides for $d_{unbiased}$ is also, according to Cumming (2012), the best CI to use for regular d.
- If desired, the p value — be sure to state the null hypothesis, which is usually but not always that $\mu_2 - \mu_1 = 0$.
- A figure, preferably one like ESCI produces that shows all the raw data and that emphasizes the estimated group difference and its CI. State in the figure caption that error bars represent 95% CIs. Estimationstats.com produces excellent high-resolution, publication-quality graphs of this sort, though note that it uses a slightly different method than ESCI for computing the CI and p-value, so if you use it, include its explanation of its approach.
- An interpretation of the group difference that considers not only the point estimate but also the CI. You can focus on the group difference in original units or the standardized group difference, whichever you feel will be easier for readers to understand. Be sure, though, that the language in your conclusion is consistent with the research design (causal for experimental, relational for non-experimental).
- For complex studies, writing out all this information for each dependent variable can become overwhelming. Fortunately, a table can provide a very efficient way to summarize lots of comparisons, although of course you'll still need to write your interpretation as text. Here's a link to an example you can use as a model.

Model examples: (all from different studies)

Include figure (left example from ESCI; right example from EstimationStats.com) along with a legend (not shown):





Report results in words (examples quoted from Cumming & Calin-Jageman, 2017):

1. The first planned analysis found that scores in the power condition (M = 32.6%, s = 14.8%, n = 68) were very slightly higher than in the control condition (M = 31.0%, s = 14.7%, n = 83). Relative to other findings in the field, the difference in performance was very small ($M_{power} - M_{control}$) = 1.6%, 95% CI [-3.2, 5.4], $d_{unbiased}$ = 0.11, 95% CI [-0.21, 0.43], p=__ for a null hypothesis of $\mu_2 - \mu_1 = 0$). The CI is fairly short, indicating that the effect of power on performance is, most likely, no more than moderate in size.
2. The participants in this sample had a much wider range of ages than in the study we replicated. Although this was not a planned analysis, we also assessed the effect of power on motor skill when restricting the analysis to only those participants younger than 30. In this restricted sample, scores in the power condition (M = 37.4%, s = 14.1%, n = 29) were somewhat higher than in the control condition (M = 34.5%, s = 14.5%, n = 42). The difference in performance was moderate ($M_{power} - M_{control}$) = 3.0%, 95% CI [-4.0, 9.8], $d_{unbiased}$ = 0.20, 95% CI [-0.27, 0.68], p=__ for a null hypothesis of $\mu_2 - \mu_1 = 0$). The CI is long, and is consistent with power producing a small detriment, no change, up to a large benefit. This exploratory analysis, then, does not provide clear evidence of a benefit of power in younger participants.
3. Motor skill for men (M = 33.8%, s = 12.8%, n = 62) was a little higher than for women (M = 29.7%, s = 15.8%, n = 89). The difference in performance may seem small in terms of raw scores ($M_{men} - M_{women}$ = 4.0%, 95% CI [-0.7, 8.8]), but the standardized effect size was moderate ($d_{unbiased}$ = 0.28, 95% CI [-0.05, 0.61], p=__ for a null hypothesis of $\mu_2 - \mu_1 = 0$) relative to the existing literature. However, both CIs are quite long, and are consistent with anywhere from no advantage up to a large advantage for men. More data are required to estimate more precisely the degree to which gender may to motor skill.

Default conventions for size of Cohen's d:

Small/Weak:	d=.20	(for this course, use this range: 0.00 to 0.349)
Medium/Moderate:	d=.50	(for this course, use this range: 0.35 to 0.649)
Large/Strong:	d=.80	(for this course, use this range: 0.65 to 1.00, above that add adjectives like "very")

These conventions should generally be used *only until information becomes available* on one or both of: (a) **typical effect** in this area of research or (b) the **costs or consequences** of different effect sizes. See more under Correlation Design.

Specific guidelines from APA: See Correlation Design

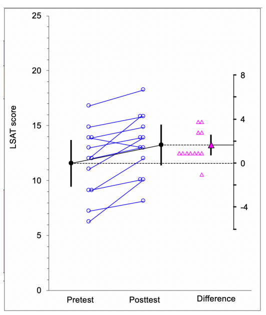
Reporting results from PAIRED DESIGN

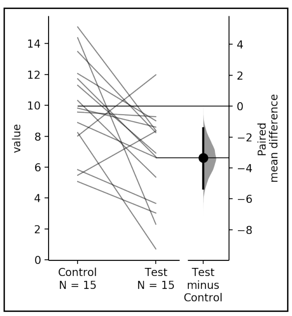
General guidance: (adapted from Cumming & Calin-Jageman, 2017)
Reporting comparisons for the paired design is pretty similar to comparing independent groups, but with a few vital additional things to remember. Typically, you need to report:

- Whether the comparison is planned or exploratory, unless this is already clear.
- Basic descriptive statistics (usually mean and SD) for both sets of measurements.
- The average difference (M_{diff}), its standard deviation (s_{diff}), and its CI.
- A standardized effect size estimate ($d_{unbiased}$ is best), preferably with its CI. In your Method section, make clear how your standardized effect size was calculated (which denominator was used). For paired design, the best denominator is most typically $s_{average}$, or s_{av}, and this is what ESCI provides (for both d and $d_{unbiased}$). The CI that ESCI provides for $d_{unbiased}$ is also, according to Cumming (2012), the best CI to use for regular d.
- Essential for future meta-analysis: the standard deviation of the differences (s_{diff}) and/or the correlation between the two measures (r); it is helpful to report both.
- A figure, if possible, preferably one like ESCI produces that shows the raw data, the paired nature of the data, and the estimated difference with its CI. State in the figure caption that error bars represent 95% CIs. It is best to join the means with a line to indicate a repeated measure, as in Figures 8.1 and 8.2, and Data paired and Summary paired in ESCI (note that EstimationStats.com, see below, has chosen not to do this).
- An interpretation of the group difference that considers not only the point estimate but also the CI. As usual, be careful to match the language you use to the research design (causal for experimental, relational for non-experimental).

Model examples: (all from different studies)

Include figure (left example from ESCI; right example from EstimationStats.com) along with a legend (not shown):





Report results in words (examples adapted from Cumming & Calin-Jageman, 2017):

1. The first planned analysis found that, when drinking the juice with a label stating “Generic”, participants rated their enjoyment fairly high (M = 6.86, s = 2.13, scale from 1 to 10). The same participants given the same drink with the label “Organic”, however, rated their enjoyment even higher (M = 7.82, s = 2.09). This difference in enjoyment was substantial, relative to the scale and previous marketing research (M_{diff} = 0.96, 95% CI [0.48, 1.44], N = 51, r = .68, s_{diff} = 2.8). In standardized terms, this is a large effect ($d_{unbiased}$ = 0.45, 95% CI [0.22, 0.69], p=.004 for a null hypothesis of $\mu_2 - \mu_1 = 0$), and the CI is consistent with at least small up to a quite large impact on enjoyment.
2. As expected, heart rate at rest was fairly low (M = 66.80 bpm, s = 10.8). During exploration, though, we noticed that when participants were asked to recall a memory of intense happiness their heart rate was somewhat higher (M = 69.50 bpm, s = 10.92). Thus, emotion was related to an increase in heart rate (M_{diff} = 2.70 bpm, 95% CI [0.14, 5.22], $d_{unbiased}$ = 0.24, 95% CI [0.01, 0.48], p=02 for a null hypothesis of $\mu_2 - \mu_1 = 0$, N = 20, r = .87, s_{diff} = 3.2). The CI indicates that happiness is likely related to an increase in heart rate, but the CI is quite long and the degree of increase could be anywhere from around zero up to fairly large. In addition, the analysis was exploratory. An independent replication is needed to estimate more precisely the degree to which happiness may be associated with an increased heart rate.
3. In a preregistered replication with a larger sample size (N = 48) we observed the same increase in heart rate during recall of a happy memory. Heart rate at rest was again fairly low (M = 63.7 bpm, s = 10.1). During the recall of a happy memory, heart rate was higher (M = 68.8 bpm, s = 12.8). This emotion-induced increase in heart rate was slightly larger than in the first study (M_{diff} = 5.1 bpm, 95% CI [3.0, 7.2], $d_{unbiased}$ = 0.44, 95% CI [0.24, 0.64], p=.001 for a null hypothesis of $\mu_2 - \mu_1 = 0$, r = .82, s_{diff} = 3.9). With the larger sample size the CI is much shorter, and indicates that happiness is related to a moderate increase in heart rate.

Default conventions for size of Cohen’s d:

Small/Weak:	d=.20	(for this course, use this range: 0.00 to 0.349)
Medium/Moderate:	d=.50	(for this course, use this range: 0.35 to 0.649)
Large/Strong:	d=.80	(for this course, use this range: 0.65 to 1.00, above that add adjectives like “very")

These conventions should generally be used *only until information becomes available* on one or both of: (a) **typical effect** in this area of research or (b) the **costs or consequences** of different effect sizes. See more under Correlation Design.

Specific guidelines from APA: See Correlation Design

Reporting results from FREQUENCY DESIGN

General guidance: (adapted from Cumming & Calin-Jageman, 2017)

If you are making inferences about a single proportion, include:

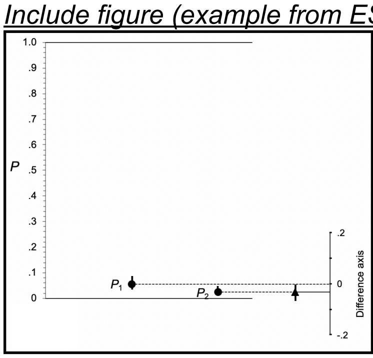
- Whether the estimate is planned or exploratory, unless this is already clear.
- The sample size that forms the denominator of the proportion, unless this is already clear.
- The proportion and its CI;
- An interpretation that considers the full range of the CI; and
- if desired, the p value for comparing to the proportion to a specified null hypothesis value.

If you are making inferences about the differences between two independent proportions, you should typically report:

- Whether the comparison is planned or exploratory, unless this is already clear.
- The sample sizes for the denominators of each proportion, unless this is already clear.
- The proportion for each group or condition.
- The difference between the proportions and the CI on the difference.
- The phi (φ) statistic and its 95% CIs. Given that phi is actually just an r value computed from two dichotomous variables, the CI on a phi is the same as the CI on the same r value and can be computed, for example, by entering the phi and its associated N (the *total* N for the two dichotomous variables being looked at) into this nice web calculator from the vassarstats website (or any of various other online calculators). Phi is also often referred to (e.g. in the crossTab function in R's userfriendlyscience package) as Cramer's V. This is because phi is a subset of Cramer's V. Cramer's V, unlike phi, can be computed for more complex frequency/proportion designs where there are more than two variables or where one or variables have more than two categories.
- Note that phi, like r, should be reported without a leading 0 because it cannot exceed 1.
- If desired, you can also report the chi-square analysis. This should include the value of chi-square, its degrees of freedom, and the p value.
- It can also be helpful to include a figure and/or table to present the frequencies and CIs underlying comparisons between proportions.

Model examples: (all from different studies)

Include figure (example from ESCI) along with a legend (not shown) and/or table (not shown):



Report results in words (examples adapted from Cumming & Calin-Jageman, 2017):

1. *[One proportion]* For the 22 attempts at communication, the proportion of hits for Pilot 1 was P = .36 [.2, .57]. This is above chance performance of .25, but the CI is fairly long and does not exclude the possibility of merely random responding (p = .20).
2. *[Difference between proportions]* Of the 303 participants in the control diet condition, the proportion who were diagnosed with cancer by the end of the study was $P_{Control}$ = .056. Of the 302 participants who ate the Mediterranean diet, the proportion who were diagnosed with cancer at the end of this study was P_{Med} = .023. Our planned comparison of these cancer rates indicates that the Mediterranean diet reduced the proportion of cancer diagnoses: ($P_{Med} - P_{Control}$) = -.033, 95% CI [−.067, −.001], χ2 = 4.30, p = .04, φ = .08, 95% CI [0.001, 0,158]. This CI suggests that the benefits of the diet are at most small by conventional standards, and potentially near-zero. Given the ease with which the diet can be adopted, however, we judge the result to be meaningful and potentially life saving.

Default conventions for size of phi (φ), which is equivalent to r computed on two dichotomous variables, and which is often referred to via the more general term Cramer's V:

Small/Weak:	φ=.10	(for this course, use this range: 0.00 to 0.199)
Medium/Moderate:	φ=.30	(for this course, use this range: 0.20 to 0.399)
Large/Strong:	φ=.50	(for this course, use this range: 0.40 to 0.599, above that add adjectives like "very")

These conventions should generally be used *only until information becomes available* on one or both of: (a) **typical effect** in this area of research or (b) the **costs or consequences** of different effect sizes. See more under Correlation Design.

Specific guidelines from APA: See Correlation Design