Appendix F Confidence interval design dependence

Confidence intervals (CIs) are defined as:

$$CI = y_0 \pm h^* \tag{F.1}$$

where y_0 is a datum and h^* is the design-dependent critical height. More specifically, h^* is given by a critical t value and simple algebraic manipulation of the design-dependent t statistic definition.

To clarify, first consider that design-dependent mean and SD definitions (Table F1) yield design-dependent t statistic definitions (Table F2). Next, given t^* one may compute the design-dependent h^* (Table F2). Last, after choosing a datum y_0, there are various acceptable null hypothesis rejection criteria (Table F3).

The main point is that hypothesis testing employs a single unambiguous criterion: $(t > t^*)$, irrespective of the particular design, making it easy to compare results across experiments. In contrast, CIs are both design- and datum-dependent.

Clearly h^* is valuable for data visualization because it represents the null hypothesis rejection criterion in the same units as the original data. However, it is also clear that h^* must be computed with careful attention to both the datum and the design, and can only be interpreted by readers if the precise datum and design are made explicit. The main manuscript therefore argues that hypothesis testing is simpler.

Table F1: Mean and standard deviation (SD) definitions for one-sample, paired and two-sample designs. For simplicity equal variance is assumed in the two-sample case.

Design	Mean	SD
One-sample	$\overline{y} = \frac{1}{J}\sum y_j$	$s_1 = \sqrt{\frac{1}{J-1}\sum(y_j - \overline{y})^2}$
Paired	$\overline{\Delta y} = \frac{1}{J}\sum(y_{Aj} - y_{Bj})$	$s_p = \sqrt{\frac{1}{J-1}\sum\left((y_{Aj} - y_{Bj}) - \overline{\Delta y}\right)^2}$
Two-sample	$\Delta\overline{y} = \overline{y}_A - \overline{y}_B$	$s_2 = \sqrt{\frac{(J_A-1)s_A^2 + (J_B-1)s_B^2}{J_A + J_B - 2}}$

Table F2: Design-dependence of the CI's critical height h^* .

Design	t	Mean	h^*
One-sample	$t_1 = \dfrac{\overline{y}}{s_1/\sqrt{J}}$	$\overline{y} = t\dfrac{s_1}{\sqrt{J}}$	$h_1^* = t^*\dfrac{s_1}{\sqrt{J}}$
Paired	$t_p = \dfrac{\overline{\Delta y}}{s_p/\sqrt{J}}$	$\overline{\Delta y} = t\dfrac{s_p}{\sqrt{J}}$	$h_p^* = t^*\dfrac{s_p}{\sqrt{J}}$
Two-sample	$t_2 = \dfrac{\Delta\overline{y}}{s_2\sqrt{\frac{1}{J_A}+\frac{1}{J_B}}}$	$\Delta\overline{y} = t_2 s_2\sqrt{\frac{1}{J_A}+\frac{1}{J_B}}$	$h_2^* = t^* s_2\sqrt{\frac{1}{J_A}+\frac{1}{J_B}}$

Table F3: Design- and datum-dependence of h^*-based null hypothesis rejection criteria. All criteria assume $\overline{y}_A \geq \overline{y}_B$.

Design	Datum (y_0)	Criterion: zero	Criterion: mean	Criterion: tail
One-sample	$\overline{y}$	$\overline{y} - h_1^* > 0$	—	—
Paired	$\overline{\Delta y}$	$\overline{\Delta y} - h_p^* > 0$	—	—
	$\overline{y}_A$	—	$\overline{y}_A - h_p^* > \overline{y}_B$	$\overline{y}_A - \dfrac{h_p^*}{2} > \overline{y}_B + \dfrac{h_p^*}{2}$
Two-sample	$\Delta\overline{y}$	$\Delta\overline{y} - \dfrac{h_2^*}{2} > 0$	—	—
	$\overline{y}_A$	—	$\overline{y}_A - h_2^* > \overline{y}_B$	$\overline{y}_A - \dfrac{h_2^*}{2} > \overline{y}_B + \dfrac{h_2^*}{2}$