Reporting Guidelines

Laurel Brehm

last updated October 2019

Here are some best-practice guidelines for how to report models. For any project, there are multiple ways to analyse data. In your write-up, you should make it clear what choices you made and why you did what you did.

It's good to report the following:

In "Participants" section:

- Power simulations that led you to select sample size.

In "Data analysis" section:

- What model/test you ran (e.g., a t-test, a linear mixed effect model...)
- For a regression model with categorical predictors: how you contrast coded your data (effects coded, dummy coded, sum coded)
- What software you used and which version (e.g., R 3.6.1), and any packages– with the version (lme4, effects...)
- To get quickly, use function citation in R: citation('lme4') [for citation for lme4 package and version]; citation('base') [for the 'base' verion of R you have]

In "Results" section:

- Descriptives of data (in text: means, SDs, CIs)
- Results of your statistical tests/models (in text using e.g. APA format, or in table)

In "References" section:

- add those citations to packages in!

0.1 Example

Here's an example from some of my recent work on memory for language production:

Zormpa, Meyer, & Brehm (2019): "Slow naming of pictures facilitates memory for their names" `https://link.springer.com/article/10.3758%2Fs13423-019-01620-x`

Participants:

Sixty individuals (18 male, mean age 22.62 years; range: 18-30 years) participated in this experiment. Eight participants were excluded–two due to technical problems during study and six for not completing the test phase. One additional participant was excluded because of substantially lower hit rates (approx. 25%) than the other participants. This left data from 51 individuals. Participants were recruited from the Max Planck Institute participant database and received 8 euros. All were native Dutch speakers with normal or corrected-to-normal vision; none reported speech or language problems. A power analysis using an effect size of 3% from Zormpa et al. (2019) showed that 48 participants would provide sufficient power to answer our main research question. Ethical approval was given by the Ethics Board of the Social Sciences Faculty of the Radboud University.

From the method section:

Analyses were run using the lme4 package (Version 1.1-18-1; Bates, Maechler, Bolker, & Walker, 2015) in R (Version 3.5.0; R Core Team, 2018) with the optimizer BOBYQA (Powell, 2009). Initially, the maximal models were fit and then reduced to overcome convergence problems or overfitting (correlations exceeding .95). Reported p values were obtained from maximum likelihood tests comparing a full model with one without the effect of interest. Reported 95% confidence intervals were calculated using the profile method of the confint function.

From results:

[Text describing the first analysis]:

The first analysis examined the effect of probe type (target vs. foil) on memory performance, as measured by yes responses in the memory task, to assess overall accuracy and response bias. This analysis was run separately from the prime condition analysis, as the design of this study was not fully crossed–that is, foils did not appear in a prime condition. Probe type was sum-to-zero contrast-coded (targets = .5, foils = -.5). The random effects structure included by-participant and by-item intercepts and by-participant and by-item random slopes for probe type.

Results and a visualization appear in Table 1 and Fig. 1. The significant negative intercept reflects a no bias. The significant effect of probe type reflects that participants were more likely to say yes to targets than foils– that is, they were highly accurate in differentiating between old and new items.

[Text describing second analysis:]

We then examined the effect of prime condition on memory performance. Prime condition was Helmert coded and split into two contrasts. The first contrast tested the effect of generation by comparing the identity condition (contrast = -0.5) to the average of the backward and unrelated conditions (contrast for both = 0.25), while the second contrast tested the effect of processing time (as a result of competition) by comparing the backward (contrast = -0.5) to the unrelated condition (contrast = 0.5). The random effects structure included by-participant and by-item intercepts and by-participant and by-item slopes for the generation contrast.

Results appear in Table 2 and Fig. 1. The significant positive intercept term reflects a yes bias to targets, indicating high accuracy. Hit rates were significantly higher in the backward and unrelated conditions than in the identity condition, showing a memory benefit for generated words. In contrast, hit rates did not differ significantly between the backward and unrelated conditions.

Table 1 Mixed-effects logistic regression testing the effect of probe type (i.e., targets vs. foils) on memory (log-odds of *yes* responses)

Fixed effects						*Random effects*			
	Estimate	*SE*	Wald *z*	*p*	CI			Variance	*SD*
Intercept	−.96	.13	−7.24	<.001	−1.23, −.70	Participant	Intercept	.26	.51
Target vs. foil	3.92	.24	16.10	<.001	3.44, 4.42		Target vs. foil	1.28	1.13
						Item	Intercept	.76	.87
							Target vs. foil	2.43	1.56

Table 2 Mixed-effects logistic regression testing the effects of generation and processing time (as manipulated by prime condition) on memory (log-odds of *yes* responses)

Fixed effects						Random effects			
	Estimate	*SE*	Wald *z*	*p*	CI			Variance	*SD*
Intercept	1.05	.16	6.47	<.001	.73, 1.37	Participant	Intercept	1.10	1.05
Id vs. Bw & Un	.46	.16	2.87	.01	.13, .78		Id vs. Bw & Un	.69	.83
Bw vs. Un	.13	.08	1.53	.13	−.04, .29	Item	Intercept	.75	.87
							Id vs. Bw & Un	.36	.60

Note. Id = identity condition; Bw = backward condition; Un = unrelated condition

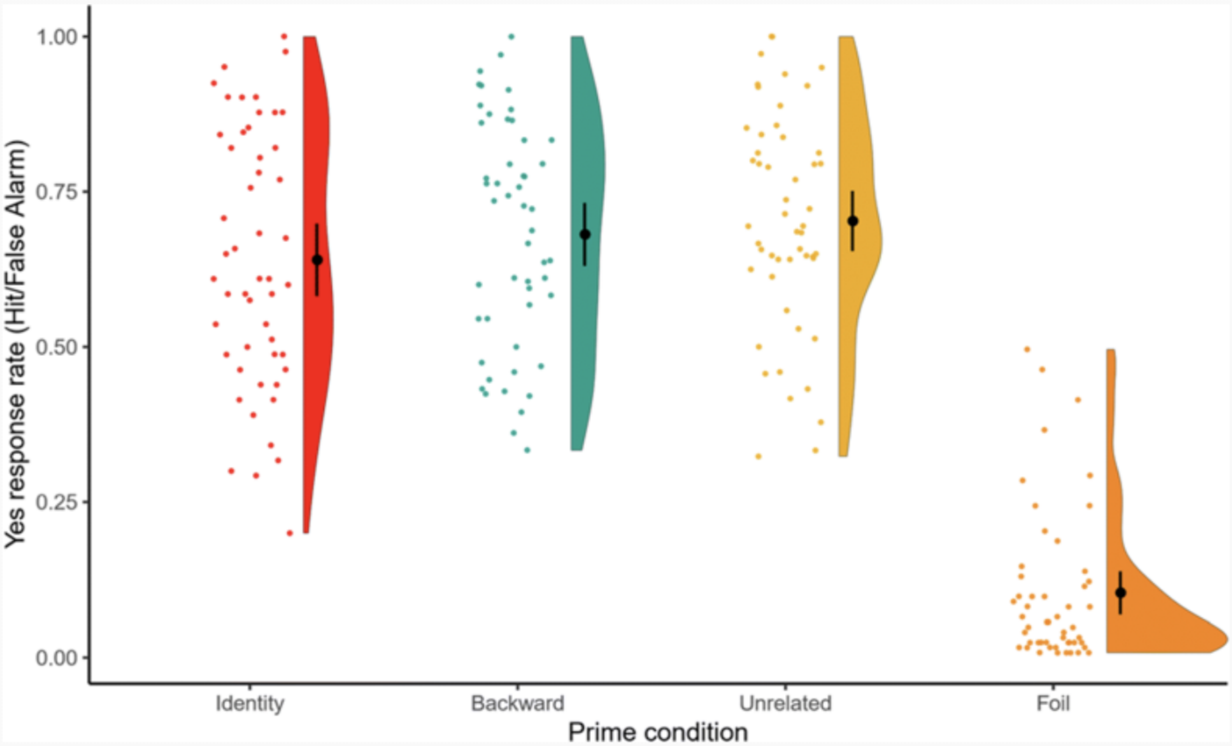

Fig. 1

Hit rates by prime condition. The dot represents the condition mean and the bars normalized within-participant 95% confidence intervals. (Color figure online)