# Writing up lmer results

- Example 1
- Example 2
- Example 3
- Example 4
- Example 5

# Example 1

Laubrock, J., Engbert, R., Rolfs, M., & Kliegl, R. (2007). Microsaccades Are an Index of Covert Attention Commentary on Horowitz, Fine, Fencsik, Yurgenson, and Wolfe (2007). *Psychological Science*, *18*(4), 364-366.

> We analyzed effects on RT with linear mixed-effects models and effects on MC with generalized linear mixed models (with a logit link), using the lme4 package (Bates & Sarkar, 2006) in the R environment (R Development Core Team, 2006). In both models, subjects were specified as a random factor to control for their associated intraclass correlation (i.e., random intercept models—Pinheiro & Bates, 2000; for an application, see Oberauer & Kliegl, 2006); these kinds of models also tolerate the necessarily unequal number of validly and invalidly cued responses.

> Does MC have an effect on RT independent of the effects of spatial cue and task? In an lme model, MC (b = 6 ms, SE = 1.7 ms), cue validity (b = 81 ms, SE = 2 ms), task (b = 95 ms, SE = 2 ms), and the interaction of cue validity and task (a larger cue-validity effect in the manual task than in the saccadic task; b = 72 ms, SE = 4 ms) were significant (all ps ≤ .001). Removing MC or cue validity from the model significantly decreased the goodness of fit, as indicated by likelihood ratio tests—effect of MC: $\chi2(1) = 10.9$, p = .0009; effect of cue validity: $\chi2(2) = 1{,}442$, p < 2.2e−16. The standard deviation of average RTs across subjects was estimated as 25 ms, and the standard deviation of residuals was estimated as 42 ms, yielding a substantial intraclass correlation of .26. Although the MC effect was 13.5 times smaller than the cue-validity effect (and less consistent across subjects), there was a reliable effect of MC on RT.

> Can the later-occurring RT account for the earlier MC effect? In a generalized linear mixed model with binary MC as the dependent variable, task ($b = 1.9$, $SE = 0.3$, $z = 5.4$, $p < .001$) and the task-by-cue-validity interaction ($b = 4.0$, $SE = 0.7$, $z = 5.6$, $p <.001$) were significant; the cue-validity effect was larger in the saccadic task (.75 − .25 = .50) than in the manual task (.68 − .33 = .35). Linear and quadratic components of RT (centered separately for each subject), specified as nested within each of the four design cells, 2 were significant for all cells except for the combination of invalid cues in the manual task; the probability of a target-congruent microsaccade always decreased as reaction time increased (Fig. 1c depicts this relationship for validly cued and invalidly cued trials in the saccadic task, using a quantile representation of RT), but cue validity had a larger effect. Removing RT components or cue validity significantly decreased the goodness of fit, as indicated by likelihood ratio tests—effect of cue validity: $\chi2(2) = 220.3$, $p < 2.2e{-16}$; effect of RT: $\chi2(6) = 75.6$, $p < 2.9e{-14}$. In summary, MC indicates effects of covert attention induced by spatial cues, and it carries some information that is not contained in subsequent RT.

# Example 2

Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-specific disruption of perceptual confidence. *Psychological science*, *26*(1), 89-98.

> Effects of condition on confidence ratings and RTs were assessed via hierarchical linear mixed-effects regression using the lme4 package in R (Version 3.0.1; Bates, Maechler, Bolker, & Walker, 2014). We obtained p values for regression coefficients using the car package (Fox & Weisberg, 2011). Mixed-effects logistic regression was used to estimate effects of condition on response accuracy. All effects were taken as random at the participant level, and condition estimates and statistics reported are at the population level. Predictors were coded as follows—accuracy: error = 0, correct = 1; congruence: incongruent = 0, congruent = 1; time: preresponse = 0, postresponse = 1.

> To understand how TMS affected trial-by-trial subjective confidence, we constructed a mixed-effects regression model to explain confidence using binary predictors for congruence, time, and accuracy (Table 1)…Congruence did not affect measures of RT to either the stimulus ($\beta = -0.008$, $SE = 0.03$, $p = .78$) or the confidence rating ($\beta = 0.03$, $SE = 0.03$, $p = .33$), which makes it unlikely that effects on metacognition were mediated by effects of TMS on details of the action itself (see Fig. S1 and Tables S4 and S5 in the Supplemental Material); there were main effects of stimulation time on RT resulting from a general speeding effect of preresponse TMS—decision RT: $\beta = 0.07$, $SE = 0.02$, $p < .01$; rating RT: $\beta = 0.08$, $SE = 0.02$, $p < .001$. No effects of M1 TMS congruence on confidence were observed (Table 2; all ps > .16), despite a very similar pattern of TMS effects on RTs (Fig. S2 in the Supplemental Material). In a regression model including stimulation location as a between-groups factor, the critical three-way interaction among location, congruency, and accuracy was significant ($\beta = 0.19$, $SE = 0.09$, $p < .05$), which indicates that action-specific effects of TMS on confidence were observed at PMd but not M1.

# Example 3

Raffray, C. N., & Pickering, M. J. (2010). How do people construct logical form during language comprehension?. *Psychological science*, *21*(8), 1090-1097.

> We analyzed our data by modeling response-type likelihood using logit mixed-effect models (Jaeger, 2008). We started all analyses with a null model that included our binomial dependent variable and participants and items as random factors; we added the predictor variable or variables to this model (incrementally, if there was more than one predictor) to see whether the model was improved. Model fit was assessed using chi-square tests on the log-likelihood values to compare different models. All analyses were carried out in the R programming language and environment (R Development Core Team, 2008) using the lme4 software package (Bates, Maechler, & Dai, 2008).

# Example 4

Chetverikov, A., & Filippova, M. (2014). How to tell a wife from a hat: Affective feedback in perceptual categorization. *Acta psychologica*, *151*, 206-213.

> The liking ratings were analyzed using linear mixed-effects regression, LMER, with the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2013). In contrast to a more traditional approach with data aggregation and repeated-measures ANOVA analysis, LMER allows controlling for the variance associated with random factors without data aggregation (see Baayen, Davidson, & Bates, 2008; Judd, Westfall, & Kenny, 2012). By using random effects for subjects and stimuli, we controlled for the influence of different mean ratings associated with these variables. For the sake of brevity, we present only the F tests from the LMER results here (type III Wald F tests with Kenward–Roger degrees of freedom approximation).

> Indeed, we found a significant negative effect of response time, F(1, 486) =11.89, p b .001. However, the effect of answer correctness still was significant, F(1, 468) = 7.73, p = .006, indicating that differences between correct and incorrect answers cannot be fully explained by differences in processing fluency.

# Example 5

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature neuroscience.*

> Multilevel logistic regression. We fit multilevel logistic regression models to the choices from the decision task using the lme4 package (http://cran.r-project (http://cran.r-project). org/web/packages/lme4/index.html) in the R statistical language (http://www (http://www). r-project.org/). All coefficients (exclusive of fMRI covariates described below) were taken as random effects —that is, varying from subject to subject around a group mean.