Best practice guidance for linear mixed-effects models in psychological science

Lotte Meteyard[1*]

Robert A.I. Davies[2]

Appendices

Appendix 1: Questionnaire

Information and consent

Project Title: Survey of use of mixed model analyses in cognitive science and related disciplines

Study Information

We are interested in the use of mixed models for data analysis in the field of psychology, cognitive science and related disciplines. To that end, we are conducting a questionnaire with researchers working in these fields. We plan to use this data to contribute to training and research on the use of mixed models in these disciplines.

You will be asked to fill out a questionnaire. All responses are anonymous. No personally identifying information will be collected. The questionnaire will ask about your experiences and opinions of mixed models. Background information on your research discipline, age, gender, seniority and place of work will also be collected. Please answer the questions honestly.

This questionnaire will be administered to groups of individuals and online. Participation is voluntary and you are free to withdraw at any time, without giving a reason.

Consent

I agree to participate in the study "Survey of use of mixed model analyses in cognitive science and related disciplines", being conducted by Dr. Lotte Meteyard and Dr. Rob Davies. I have seen and read a copy of the Study Information. I understand that no personal information will be collected. I understand that all my responses are anonymous. I understand that participation in this study is voluntary and that I can withdraw at any time without having to give an explanation.

I am happy to proceed with my participation (please check box) ☐

Questionnaire

1. Gender (circle choice): Male Female Do not want to disclose

2. Age: __________ Do not want to disclose

3. Position (circle choice)

   Undergraduate
   Postgraduate (MSc)
   Postgraduate (PhD)
   Post-doctoral researcher
   Lecturer/Assistant professor
   Reader/Senior Lecturer/Associate Professor
   Professor
   Other

4. Institution (circle choice)
   University UK
   University Other (please state country):
   Research Institute UK
   Research Institute Other (please state country)
   Other institution (please state type):

5. What is your research discipline?

6. How long have you been doing research in this discipline?

7. In your opinion, what are the main challenges to using mixed models?

8. Do you have any concerns about applying and reporting mixed models for **your own data**? If so, what are they?

9. Do you have any concerns about the use of mixed models **in your field**? If so, what are they?

10. Have you used mixed-models for data analysis?
    YES / NO

    If yes, answer question 11 (start this page).
    If no, answer questions 16 (stare page 4).

If answered YES to question 10:

11. When did you first use them (year)

12. How often do you use them (approx. % of data analyses)

13. What software do you use to fit mixed models?

14. What training / support have you used to learn mixed model analyses?
    (Please refer to specific papers, books, courses or websites, if relevant)

15. What is the typical formula for a model you employ?
    e.g. RT ~ factor1 + (1|subjectID) + (1|itemID)

16. Do you specify variance-covariance structures?

17. Do you report (circle all that apply):
    p-values
    confidence intervals
    likelihood ratio tests
    F-tests
    description of model fitting process
    results of iterated models / model testing

18. Have you compared your results following a more traditional analysis and a mixed-effects analysis?

19. Have the results been different?

20. If results were the same, what was your evaluation of the results?
    If results were different, what was your evaluation?

If answered NO to question 10:

21. Do you plan to use mixed models? If yes, please state how soon.
    YES/NO

22. What is your interest in mixed-models /
    what do you hope to use them for?

Please finish by answering the following 3 questions:

23. When mixed models are reported, would you prefer model estimates reported in text, in a table, as a partial effects plot? Why?

24. Would you share analysis code and data? Why?

25. Would you like to access analysis code and data in published reports?
    Why?

A2: Authors, papers, books and websites used to learn mixed-models

Where a resource/individual was mentioned by more than 5% of respondents, the percentage of respondents who reported it is given.

Books / documentation

Adler, J. (2010). *R in a nutshell: A desktop quick reference*. O'Reilly Media, Inc. CA, USA.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. 59%

Bates, D. M. (2010). lme4: Mixed-effects modeling with R. *URL http://lme4. r-forge. r-project. org/book*.

CRAN documentation. URL: http://cran.r-project.org/

Faraway, J. J. (2005). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press.

Fox, J. (2002). Linear mixed models. *Appendix to An R and S-PLUS Companion to Applied Regression, URL http://cran. r-project. org/doc/contrib/Fox-Companion/appendix-mixed-models. pdf*.

Fox, J., & Weisberg, S. (2010). *An R companion to applied regression*. Sage.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. 9%

Goldstein, H. (2011). *Multilevel statistical models* (Vol. 922). John Wiley & Sons.

Gries, S. T. (2013). *Statistics for linguistics with R: a practical introduction*. Walter de Gruyter.

Hox, J. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

Johnson, K. (2011). *Quantitative methods in linguistics*. John Wiley & Sons.

Kliegl, R. (2014) Linear Mixed Model: Eye Movements During Reading of Uighur. URL: http://rpubs.com/Reinhold/17313 (see also: http://rpubs.com/Reinhold/22193)

Kruschke, J. (2010). *Doing Bayesian data analysis: A tutorial introduction with R*. Academic Press.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media. 5%

Levy, R. (in progress) Probabilistic Models in the Study of Language. URL: http://idiom.ucsd.edu/~rlevy/pmsl_textbook/text.html

Littell, R. C., Stroup, W. W., Milliken, G. A., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for mixed models*. SAS institute.

Long, J. D. (2011). *Longitudinal data analysis for the behavioral sciences using R*. Sage.

Milliken, G. A., & Johnson, D. E. (2009). *Analysis of messy data volume 1: designed experiments* (Vol. 1). CRC Press.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. CRC Press.

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.

Snijders, T.A. & Bosker, R. (2011) Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. 2nd Ed. SAGE Publications, London.

West, B. T., Welch, K. B., & Galecki, A. T. (2014). *Linear mixed models: a practical guide using statistical software*. CRC Press.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint arXiv:1308.5499*. URL: http://arxiv.org/pdf/1308.5499.pdf

Courses & Sites

Bristol Centre for Multilevel Modelling:
http://www.bristol.ac.uk/cmm/learning/

Bristol online LEMMA Course: Learning Environment for Multilevel Methods and Applications. URL: http://www.cmm.bris.ac.uk/lemma

Franklin Chang (2015) Introduction to linear models in R. University of Kent.
URL: https://www.kent.ac.uk/psychology/workshops/r

Deepayan Sarkar (2008) Fitting Mixed-Effects Models Using the lme4 Package in R
URL: http://master.bioconductor.org/help/course-materials/2008/PHSIntro/lme4Intro-handout-6.pdf

Journal Articles

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.
19%

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological review*, *118*(3), 438.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278. 29%

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369-382.

Davidson, DJ & Martin, AE 2013, 'Modeling accuracy as a function of response time with the generalised linear mixed effects model' Acta Psychologica, vol 144, no. 1, pp. 83-96., 10.1016/j.actpsy.2013.04.016

Janssen, D. P. (2012). Twice random, once mixed: Applying mixed models to simultaneously analyze random effects of language and participants. *Behavior Research Methods*, *44*(1), 232-247.

Locker, L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research.*Behavior Research Methods*, *39*(4), 723-730.

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416-420.

Named online resources and mailing lists

Mixed models listserv. URL: https://stat.ethz.ch/mailman/listinfo/r-sig-mixed-models

R-Lang mailing list. URL: https://mailman.ucsd.edu/mailman/listinfo/ling-r-lang-l 8%

R-websites/R-help/R-forums 8%

Stack Exchange

Stack Overflow 5%

Specific web addresses

http://www.cookbook-r.com/

http://www.bodowinter.com/tutorials.html

http://glmm.wikidot.com/faq

www.inside-r.org

http://resmeth.wikidot.com/

Quick-R: www.statmethods.net

http://www.stat.columbia.edu/~gelman/arm/

www.statisticshell.com/

http://talklab.psy.gla.ac.uk/

http://www.uvm.edu/~dhowell/StatPages/

Journals

Journal of Memory and Language	9%

Individuals

Harald Baayen	11%
Dale Barr	9%
Marc Brysbaert	
Matthew Goldrick	
David C. Howell	
Florian Jaeger	14%
Hugo Quene	
Stephen W. Raudenbush	
Shravan Vasishth	

Software available for power analysis

Jake Westfall's online power calculators for varying experimental designs
http://jakewestfall.org/, Web apps page.

Tom Snijders, Roel Bosker, and Henk Guldemond's PINT.
Power analysis IN Two-level designs.
https://www.stats.ox.ac.uk/~snijders/multilevel.htm

Optimal Design (OD) software available from Stephen Raudenbusch and colleagues
https://sites.google.com/site/optimaldesignsoftware/home

RMASS2 from Don Hedeker and colleagues
http://hedeker.people.uic.edu/ml.html

A3: Description of classification criteria for review of papers using LMMs

A3.1. Field/Topic: Papers were broadly tagged as Psychology, Psycholinguistics, Linguistics & Phonetics, and Neuroscience. Where papers covered more than one discipline (e.g. if they presented both behavioural and neuroscientific data), the classification was made on the basis of the 'take home point' or key hypotheses.

A3.2 Model Type: Linear Mixed Effects Models (LMM), Generalised Linear Mixed Effects Models (GLMM, e.g. logit models), both LMM and GLMM (e.g. in the analysis of both reaction time and error data), Generalised Additive Models (GAMS), or Other (e.g. item response modeling).

A3.3 Approach: This criteria sought to capture the aim of the analysis approach. ANOVA testing for fixed effects via LRTs/model comparison: mixed models are used as a replacement for a factorial ANOVA, including a control for by subject and by item variance, and testing for significance by comparing models with and without factors of interest (i.e. significance according to an increased in explained variance). ANOVA testing with random effects of interest: as above, but the random effects are presented as the authors are interested in the variation across subjects, items or another grouping. Regression with control for subject / item variance: mixed models are used in place of a regression analysis, including explicit control for subject and item variance. Regression with multiple predictors and control variables: models are used in place of regression analysis with the focus on analyzing the relative contributions of a complex set of predictor and control variables. Regression with random effects of interest: mixed models are used in place of a regression analysis and the random effects are presented as the authors are interested in the variation across subjects, items or another grouping. Repeated measures / control for hierarchical sampling: the classical use of mixed models in which the analysis controls for repeated measures sampling designs, with individual data points grouped under several nested levels (e.g. children in classes, schools and towns). Repeated measures with random effects of interest: as above, but the random effects (i.e. the nested groupings) are presented are also presented and explored.

A3.4 Model Comparison: The means by which model iterations or nested models have been compared. Likelihood Ratio Tests (LRTs), Akakike Information Criterion (AIC), Bayesian Information Criterion (BIC), or multiple (LRTs & AIC/BIC), Descriptive: the paper describes a process of model comparison but does not report them as part of the method or results.

A3.5 Model Selection: Categories generated from information reported in manuscripts on how models were selected. For example, best fitting models compared with LRTs and AIC/BIC, minimal or maximal models.

A3.6 Model equation: is the underlying equation for the model reported as part of the results (Yes, No, for some and not others).

A3.7 Dependent variable: the data being analysed (see Table 8).

A3.8. Fixed effects 1: Whether the analysis tests independent variables (IVs) or independent variables alongside control variables (IV & Control).
Fixed effects 2: Whether the main effects or main effects and interactions are analysed.

A3.9. Random Effects Approach (RE Approach): If this was mentioned in the paper, the way in which the random effects structure had been decided upon. Comparing models with and without RE effects (LRTs, LRT & AIC/BIC), comparing models only for random effect slopes (LRTs/AIC for slopes), comparing 'empty' models to those with RE (LRTs against null), including the maximal structure as recommended by Barr et al, 2013 (Maximal structure), starting with the maximal structure and removing REs that do not contribute significantly to model fit (LRTs backwards from maximal), starting with a minimal RE structure and adding in REs that improve model fit (LRTs upwards from minimal).
Convergence / simplification: If mentioned, did initial models not converge and subsequent simplifications were made to the RE structure.

A3.10 Random effect intercepts and Random Effect Slopes: what RE intercepts and/or slopes were included in the final models: see Table 8.

A3.11 Random effect covariances: where intercepts and slopes were both included, was the correlation/covariance between them also modeled?

A3.12 Reporting Format: the way in which results were reported/presented, see Table 8. Note that 'table' was only included if the Table presented the model outputs (coefficients, standard errors etc.) and not when tables were used to present averages of raw data.

A3.13 Reporting Fixed Effects: what aspects of the model output were reported (see Table 8).

A3.14 Reporting Random Effects: what aspects of the random effects were reported (see Table 8).

A3.15 Model fit reported: was the goodness of fit for the model to the data reported, and if so, what measure of model fit? See Table 8.

A3.16 P values: If mentioned, we classified what method was used to evaluate the significance of the effects of interest. There were a number of methods in the literature. Authors sometimes assumed that t values greater than +/- 1.96 or 2 were significant (under the assumption of the t distribution approaching normality with large enough samples), or used modeling methods available with R packages (MCMC sampling) or specific means for calculating degrees of freedom (Sattherwaite, Kenward-Rogers). In some cases, the comparison of models with and without the predictor of interest was used to evaluate significance for that predictor (LRTs) and in others the model output was put through an ANOVA (F tests).

A3.17 Appendices: was a full presentation of data, model iterations and/or data and code given in supplementary materials or appendices?

A4: Review data from papers citing Baayen et al (2008) 2013-2016

A4.1: Model Type

	2013	2014	2015	2016	*Total*
LMM	54	49	41	49	*193*
GLMM	15	22	28	23	*88*
LMM & GLMM	24	24	22	25	*95*
GAM	1	1	2	1	*5*
GAM + LMM/GLMM	2	1	0	0	*3*
Other	1	2	0	0	*3*
Total	*97*	*99*	*93*	*98*	

A4.2: Approach

	2013	2014	2015	2016	*Total*
ANOVA (F1/F2)	23	14	11	13	*61*
ANOVA (F1/F2) + RE	0	1	1	0	*2*
Regression + RE control for subjects/items	52	64	73	73	*272*
Regression + RE of interest	1	1	1	6	*9*
Regression with multiple predictors	5	4	0	3	*12*
Regression with multiple predictors & controls	6	2	0	1	*9*
Repeated measures hierarchical sampling	9	11	6	0	*26*
Repeated measures + RE of interest	0	2	1	1	*4*
Other*	1	0	0	1	*2*
Total	*97*	*99*	*93*	*98*	

*2013 – unclassifiable; 2016 – multinomial model to normalize data

A4.3: Model selection

	2013	2014	2015	2016	*Total*
Statement on model selection					
"Best fit" with LRT or AIC comparisons	9	6	4	4	*23*
Minimal to maximal	0	0	2	6	*8*
Maximal to minimal	0	0	0	6	*6*
Minimal model (model simplification if FE/RE ns)	3	10	11	7	*31*
Backwards fit	1	0	4	2	*7*
Generic statement about best model	2	1	1	0	*4*
Separate models to parse interactions	2	0	0	0	*2*
Other	4	3	3	0	*10*
Total	***21***	***20***	***25***	***25***	
Convergence issues	2	8	14	21	*45*
Intercept only	0	2	1	3	*6*
Removed slopes	0	1	4	4	*9*
Removed correlations	0	0	0	2	*2*
Removed slopes & correlation	1	0	3	0	*4*
Include slopes with largest variance / remove those with smallest	1	1	3	1	*6*
Remove FEs	0	1	0	0	*1*
Remove RE interactions	0	3	0	0	*3*
General simplification / report fullest model that converged	0	0	3	11	*14*

A4.4: Model Comparison

	2013	2014	2015	2016	*Total*
LRT	40	23	37	29	*129*
AIC/BIC	2	5	3	2	*12*
LRT & AIC/BIC	6	4	3	7	*20*
Descriptive	5	16	18	15	*54*
LRT & descriptive	1	0	0	0	*1*
Other	2	0	0	0	*2*
Total	*56*	*48*	*61*	*53*	

A4.5: Fixed Effects

	2013	2014	2015	2016	*Total*
IV	69	74	62	66	*271*
IV & Control variables	24	24	31	30	*109*
Main effects	24	24	24	22	*94*
Main effects & interactions	71	73	69	74	*287*

A4.6: Random Effects

	2013	2014	2015	2016	*Total*
Approach					
LRTs	10	7	4	4	*25*
LRT & AIC/BIC	1	0	0	0	*1*
LRTs/AIC for slopes	0	2	0	1	*3*
Maximal structure	5	20	29	32	*86*
LRTs backwards from maximal	0	3	6	2	*11*
LRTs upwards from minimal	0	0	4	3	*7*
LRTs against null	0	1	0	0	*1*
Total	***16***	***33***	***43***	***42***	
RE structure					
Intercept					
Subject	12	23	10	19	*64*
Item/other	2	2	3	1	*8*
Subject & Item/other	71	68	66	72	*277*
Item/other & other	0	0	1	1	*2*
Subject & item & other	5	4	11	4	*24*
Unclear	2	1	0	0	*3*
Total	***92***	***98***	***91***	***97***	
Slopes					
FE over Subject	15	20	23	20	*78*
FE over Item/other	1	1	3	3	*8*
FE over Subject & Item/other	21	26	27	20	*94*
FE over Subject with interactions	2	1	3	2	*8*
FE over Item/other with interactions	0	1	1	1	*3*
FE over Subject & Item/other with interactions	2	1	10	12	*25*
Total	***41***	***50***	***67***	***58***	
Correlations/covariances					
Explicitly mentioned or noted	6	17	17	16	*56*
Not fit	47	2	9	5	*63*
Unclear	24	18	5	15	*62*
Reference to ‘maximal’ (*can assume this means correlations – but not always stated explicitly*)	5	20	29	32	*86*

A4.7 Model equation reported

	2013	2014	2015	2016	*Total*
Yes	7	6	26	22	*61*
For some & not others	2	3	0	3	*8*
No	88	89	67	73	*317*
Total	***97***	***98***	***93***	***95***	

A4.8 Reporting p values

	2013	2014	2015	2016	*Total*
Assume t is normal / > 1.96 or 2	12	13	8	19	*52*
Sig based on coeff 2*SE / CI	2	1	0	0	*3*
P values from LRT change	1	4	19	16	*40*
MCMC	33	17	9	4	*63*
MCMC & assume t>2	3	0	1	0	*4*
MCMC & F tests	4	0	0	0	*4*
F Tests	0	5	4	3	*12*
F Tests Kenward Roger	0	2	1	4	*7*
Sattherwaite	0	1	5	14	*20*
Df = no. obs - FE	0	0	1	1	*2*
Total	*55*	*43*	*48*	*61*	*207*

A4.9 Reporting Model Fit

	2013	2014	2015	2016	*Total*
AIC/BIC	0	1	4	6	*11*
AIC/BIC & LogLik	1	0	7	9	*17*
LogLik	2	2	2	1	*7*
Deviance	0	2	0	0	*2*
Deviance & AIC/BIC	0	1	0	0	*1*
R2	8	1	3	4	*8*
R2 & AIC/BIC/LogLik	0	0	1	0	*1*
Other	0	1	0	1	*2*
Not reported	86	91	76	77	*330*
Total	*97*	*99*	*93*	*98*	

A4.10 Reporting format

	2013	2014	2015	2016	*Total*
Text only	20	14	4	14	*52*
Text & Tables	35	17	22	11	*85*
Text & Figures	15	30	22	27	*94*
Text, Tables & Figures	26	38	42	45	*151*
Tables	0	0	1	1	*2*
Figures	0	0	1	0	*1*
Table & Figures	0	0	1	0	*1*
Total	*96*	*99*	*93*	*98*	

A4.11 Reporting Fixed Effects

	2013	2014	2015	2016	*Total*
Coefficients	2	0	2	0	*4*
Coefficients, SE/CI	1	1	3	4	*9*
Coefficients, SE/CI, p	10	8	9	12	*39*
Coefficients, SE/CI, t/z	13	17	17	15	*52*
Coefficients, SE/CI, t/z, p	23	31	39	35	*128*
Coefficients, p	2	10	7	6	*25*
Coefficients, t/z	1	2	0	0	*3*
Coefficients, t/z, p	18	10	5	8	*41*
P	2	2	1	2	*7*
t/z	0	1	0	0	*1*
t/z, p	18	8	3	6	*35*
Model comparison only	2	0	0	0	*2*
Total	*92*	*90*	*86*	*88*	

A4.12 Reporting Random Effects

	2013	2014	2015	2016	Total
Variance	13	13	13	12	51
Variance & Correlation/covariance	4	7	7	5	23
Variance & Significance	0	0	1	0	1
Variance only for slopes of interest	0	0	1	0	1
Figures	0	0	0	2	2
Not reported	80	77	69	78	304

A4.13 Use of appendices to report full models and/or model iterations

	2013	2014	2015	2016	*Total*
Papers counted	6	3	9	7	*25*

A4.14 Contingency table for Model Approach and evaluating significance

Framing of approach	LRTs or F Tests	Coefficients	Both
ANOVA	30	12	
Regression	31	122	2
Repeated measures hierarchical sampling	4	7	
Total	*65*	*141*	*2*