Extreme Sawtooth-Sign in Motor Neuron Disease (MND) suggests Laryngeal Resistance to Forced Expiratory Airflow

Britton, D., Kain, A., Chen, Y-W, Wiedrick, J., Benditt, J.O., Merati, A.L., & Graville, D.

Appendix C: Supplemental Statistical Methods

Mixed-effects model specification (expiration data)

For the expiration trials, a linear mixed-effects model [1] was fit for additive effects of disease status (MND vs control) and task (cough vs. FVC), with linear adjustments for age, sex, and trial duration, and incorporating random intercepts for participants. This is a standard random-offset model for repeated measures, where participant means after adjustment for the model effects are assumed to vary normally about a common mean. For added flexibility, task-specific heteroskedastic residuals were modeled, where within-subject variances are modeled for each task as independent normal deviations about the participant trial means for the task. For parameter covariance calculations, the commonly-used Huber-White robust ("sandwich") covariance estimator[2,3] was employed over the repeated measures for the 24 participants in order to adjust for possible violations of the standard model covariance assumptions.

Hurdle model specification (inspiration data)

The inspiration data was not amenable to a standard mixed-modeling treatment due to the large number of zero responses: of the 144 trials, only 28 (19%) logged any nonzero response for the presence of periodicity during inspiration. The zero-inflated inspiration data was therefore handled using an exponential Cragg hurdle model,[4] a standard analytic choice for zero-inflated data of this type. A hurdle model is a joint

regression model that contains a mean response submodel for the value of an observation conditional on it being nonzero, and a selection submodel for the probability of having a nonzero response as a function of covariates. For the mean response submodel, the same covariates as in the model for expiration data were used with an additional assessment of the importance of a sex-by-disease interaction term. Neither age nor trial duration were important predictors of mean response for any inspiration-phase outcome. For the selection submodel, disease status, sex, age, and trial duration were included with an evaluation of the relative strength of their influence on the probability of observing periodicity in the airflow signal. As in the linear mixed-effects model, the cluster-robust Huber-White sandwich covariance estimator was employed to account for the correlation in responses induced by the clustering of repeated measures for individual participants.

Effects as contrasts of model predictions

Comparisons between factor levels were estimated using conditional marginal effects (i.e. contrasts of model predictions, integrated over all other model parameters) with confidence intervals calculated via the Korn and Graubard population-variance estimator.[5] These comparisons are consistent estimates of the cross-sectional difference in response in two independent populations caused by having different levels of the factor in each population.

Estimating correlations among measures

Within-subject correlations among the periodicity, magnitude, and kurtosis measures on expiration events were estimated using Bland and Altman's method,[6] implemented as a fixed-effects within-subjects regression[7] of one variable on the other

(the order is arbitrary); degrees of freedom for the significance calculations were taken as $nm/(1+(m-1)\cdot ICC)-3$, where $n=24$ is the number of subjects, $m=6$ is the number of trials per subject (across both the cough and FVC tasks), and ICC represents the intraclass correlation coefficient estimated by the within-subject model. Correlations between periodicity and vocal fold kinematics and airflow measures were explored only for the cough tasks because vocal fold kinematic and airflow data was not collected for the FVC tasks in the Britton et al. (2014) study.[8] The correlations were estimated as partial correlations from structural equation models[9] where the two measures to be correlated were adjusted for the known factors of disease status, task, sex, and age in order to enable assessment of the potential influence of unmeasured factors on both variables; this influence represents unexplained correlation between the variables. The structural equation models were estimated using full-information maximum likelihood with the robust Huber-White sandwich covariance estimator, as above.

Measures of significance and effect size

Effect significance was reported in terms of a z-score representing the directional deviation from the null (zero) value for the association in question, as measured in units of the effect's standard error. For all comparisons, the sign of the z-score indicates the direction of departure of the MND group away from the control group value, or of the FVC task from the cough task, or of females from males; for example, if the z-score is negative then the MND group has lower values of the outcome. Z-scores larger than approximately 1.65 in magnitude (in either direction) are suggestive of significant effects, and any larger than about 3 in magnitude are highly salient in the sample. In the case of regression coefficients or predictive contrasts from one of the models mentioned

above, the z-score for the effect is calculated in the usual way, dividing the effect estimate by its model-based standard error. In the case of partial correlations estimated from the within-subject models or structural equation models, the correlation value was transformed using the Fisher z-transformation (i.e. hyperbolic arctangent) and scaling by the degrees of freedom calculated as described above. It is important to remember that z-scores themselves are not measures of effect size, but rather effect significance (in the limited statistical sense of "surprising if the null hypothesis is true"). To allow comparisons of effect size in a standardized way that does not depend on the units of measurement, the Hedges's g measure of effect size[10] was reported, as it is a typical choice of effect size metric for studies of our size due to its robust small-sample bias correction.

REFERENCES

1. Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Second Edition ed. Thousand Oaks, CA: Sage Publications, Inc.; 2002.
2. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. Paper presented at: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics; 1967, 1967; Berkeley, Calif.
3. White H. Maximum Likelihood Estimation of Misspecified Models. *Econometrica.* 1982;50(1):1-25.
4. Cragg JG. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica.* 1971;39(5):829-844.
5. Korn EL, Graubard BI. *Analysis of Health Surveys.* New York: John Wiley & Sons, Inc.; 1999.
6. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 1--Correlation within subjects. *BMJ.* 1995;310(6977):446.
7. Allison PD. *Fixed Effects Regression Models.* Thousand Oaks, CA: Sage Publications, Inc.; 2009.
8. Britton D, Benditt JO, Merati AL, et al. Associations between laryngeal and cough dysfunction in motor neuron disease with bulbar involvement. *Dysphagia.* 2014.
9. Kline RB. *Principles and Practice of Structural Equation Modeling.* Fourth Edition ed. New York: Guilford Publications, Inc.; 2015.

10. Hedges LV. Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics.* 1981;6(2):107-128.