Interpretation and Effect Size Measures for Random Effects in Multilevel Survival Models

Julie Lorah, jlorah@iu.edu

Indiana University, Bloomington

Context

Effect size reporting represents an important step in data analysis, allowing the researcher to interpret and contextualize the findings. Multilevel models consider data at two or more levels of nesting, such as students nested within schools. Thus, an important result from multilevel models involves an effect size measure for the random effect of group.

The multilevel survival model, which measures a time to event outcome with nested data may be particularly useful for experimental and educational research. For example, Wao (2010) used a discrete multilevel survival model to investigate a time to doctorate outcome. Multilevel survival models may also be a useful tool for experimental research where predictors are based on assigned groups such as an experimental curriculum.

The present study considers one discrete multilevel survival model and one continuous multilevel survival model for comparison purposes. By considering the random effect size measure for each model using the same data, the implications of choosing a discrete versus continuous model can be considered.

The discrete multilevel survival model can be implemented using a multilevel logistic regression model as follows:

$$\text{logit}(p_{ij}) = \gamma_1 * t1_{ij} + \gamma_2 * t2_{ij} + \gamma_3 * t3_{ij} + \gamma_4 * t4_{ij} + \gamma_5 * t5_{ij} + u_{0j}$$
$$\text{where } logit(p) = ln\left(\frac{p}{1-p}\right) \quad (1)$$

Where p represents the probability of the event occurring, t1-t5 represent five discrete time points, γ represent slope coefficients, u represents the random effect for group, i represents level-1 individuals and j represents level-2 groups. Note that the preceding model includes all 5 time points and as such, no intercept is estimated. No predictors are added to the present formulation since the unconditional model is typically used to assess random effects.

The multilevel Cox regression model, considered a frailty model (Austin, 2017) can be implemented as follows:

$$h_i(t) = h_0(t)\exp(u_j) \quad (2)$$

Where all symbols are as defined previously and $h_0(t)$ represents the baseline hazard function.

To measure the effect of nesting for a multilevel logistic regression, the intraclass correlation coefficient (ICC) or the median odds ratio (MOR) can be computed; for the multilevel Cox regression, the median hazard ratio (MHR) can be computed (Austin et al., 2017). If σ^2 is given as the variance of the random effect, the ICC = $\sigma^2/(\sigma^2 + \pi^2/3)$ (Rodriguez & Elo, 2003); MOR = $\exp\left(\sqrt{2\sigma^2}\phi^{-1}(.75)\right)$ where ϕ^{-1} indicates the inverse of the standard normal cumulative distribution function (Austin et al., 2017);

and MHR for a model with random effects distributed according to the Gamma distribution is computed as upper quantile of $F(2\sigma^{-2}, 2\sigma^{-2})$ distribution (Austin et al., 2017).

Research question

How does the choice of a discrete versus continuous multilevel survival model relate to the information about random effect size provided for various datasets?

Methods

Monte Carlo simulation was used. All data was simulated and analyzed in R (R Core Team, 2019). Variables that vary by condition include level 1 sample size (i.e. group size, values of 5 and 30); level 2 sample size (i.e. number of groups, values of 10 and 40), and nesting (small, medium, and large group effect). A total of 2*2*3 = 12 conditions were simulated and for each condition, 500 simulations were used. A few values were not varied by condition including the average probability of censoring which was set at 0.3 and the number of time periods which was set at 5. Simulated datasets included a group (random) effect by varying the probability of censoring based on level 2 group membership.

Two empty multilevel models were estimated: the discrete survival model was estimated with a multilevel logistic regression model using glmer function in the lme4 package (Bates et al., 2015) and the continuous multilevel Cox (frailty) model with Gamma random effects distribution was estimated using the frailtyPenal function in the frailtypack package (Rondeau, Mazroui, & Gonzalez, 2012).

Results

Table 1 shows the average value for ICC, MOR, and MHR for each condition, as well as the standard deviation of the simulations which represents standard error for the given statistic. Table 2 shows the correlations among the effect size values and the conditions. The ICC, based on the discrete survival model represents the proportion of variance at the group level. For example, for the first condition (i.e. 1st row on Table 1), the ICC = 0.04 indicating that about 4% of the variance in the outcome measure is at the group level. The ICC doesn't tend to vary much based on sample size although Table 2 indicates a small positive correlation between ICC and sample size. As expected, the ICC is positively related to the true group variance. The MOR represents the ratio of median odds for an individual who moves from a low-risk to a high-risk cluster (Austin et al., 2017). For example, for the first condition (i.e. first row of Table 1), the odds of event occurrence for a high-risk group are 1.33 times those for a low-risk group. The MOR also shows a small positive correlation with sample size and is positively related to the true group variance. The MHR represents the median increase in hazard when moving from a low-risk to a high-risk cluster. For example, the first condition indicates that the median increase in hazard for event occurrence is 3% when moving from a low-risk to high-risk group. The MHR shows a small negative relationship with sample size and a modest positive relationship with the true group variance.

Conclusions

Based on the present results, the degree of nesting may be under-represented with the MHR based on the Cox regression model compared with the MOR based on the logistic regression model. With the present conditions, particularly with only 5 time points, the estimates based on the continuous Cox model appear to be somewhat unstable. The full paper will additionally include conditions with varying number of time points which should help researchers assess, based on their data, whether a discrete or

continuous model may be more appropriate. The full paper additionally provides a demonstration of interpretation of these measures of effect size for random effects in multilevel survival models.

References

Austin, P. C. (2017). A tutorial on multilevel survival analysis: Methods, models and applications. *International Statistical Review, 85*(2), 185-203. doi:10.1111/insr.12214.

Austin, P. C., Wagner, P., & Merlo, J. (2017). The median hazard ratio: A useful measure of variance and general contextual effects in multilevel survival analysis. *Statistics in Medicine, 36*, 928-938. DOI: 10.1002/sim.7188.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rodriguez, G. & Elo, I. (2003). Intra-class correlation in random-effects models for binary data. *The Stata Journal, 3*(1), 32-46.

Rondeau, V., Mazroui,Y., & Gonzalez, J. R. (2012). frailtypack: An R Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation or Parametrical Estimation. *Journal of Statistical Software, 47*(4), 1-28. URL http://www.jstatsoft.org/v47/i04/.

Wao, H. O. (2010). Time to doctorate: Multilevel discrete-time hazard analysis. *Educational Assessment, Evaluation and Accountability, 22*, 227-247. DOI 10.1007/s11092-010-9099-6.

Table 1

Average value of ICC, MOR, & MHR for each of 12 conditions.

Each value based on 500 replications. SE represents the standard deviation of the given measure. NL1 = group size; NL2 = number of groups; L2SD = levels of nesting.

NL1	NL2	L2SD	ICC	SE	MOR	SE	MHR	SE
5	10	0.2	0.04	0.05	1.33	0.38	1.03	0.07
30	10	0.2	0.04	0.03	1.39	0.22	1.09	0.07
5	40	0.2	0.04	0.03	1.39	0.24	1.01	0.04
30	40	0.2	0.04	0.02	1.45	0.1	1.13	0.05
5	10	0.3	0.08	0.08	1.61	0.59	1.1	0.16
30	10	0.3	0.08	0.07	1.68	0.46	1.14	0.08
5	40	0.3	0.08	0.05	1.67	0.31	1.08	0.11
30	40	0.3	0.09	0.03	1.74	0.2	1.05	0.08
5	10	0.4	0.12	0.11	1.93	0.8	1.19	0.22
30	10	0.4	0.14	0.11	2.12	1.03	1.16	0.09
5	40	0.4	0.14	0.06	2	0.36	1.23	0.16
30	40	0.4	0.16	0.06	2.13	0.36	1.01	0.04

Table 2

Correlations between conditions and effect size values.

NL1 = group size; NL2 = number of groups; L2SD = levels of nesting. Each value based on 12 conditions * 500 replications = 6000 total replications.

	NL1	NL2	L2SD	ICC	MOR	MHR
ICC	0.06	0.05	0.53			
MOR	0.08	0.05	0.47	0.97		
MHR	-0.04	-0.13	0.26	0.49	0.44	