ESM Material

Not many diagnostics exist for multilevel mixed effects models, especially for those with binary outcomes. We propose a new method, based on the work of Perera et al (24) to derive a more intuitive analog of the Hosmer-Lemeshow (HL) statistic for evaluation of model calibration of multilevel models with binary outcomes. A multilevel model accounts for the fact that patients are nested within hospitals, and fits one regression line for each hospital, based on all of the patients at the given hospital. Given this understanding of a multilevel model, a natural extension of the HL statistic to multilevel models, then, is to apply the commonly used HL statistic to each nested result; i.e. to the predictions of each hospital separately. We use the non-adjusted significance level (since the Bonferroni correction is conservative) to evaluate model calibration across the hospitals.