# Methods for intuitive presentations of hierarchical categorical models: An example predicting assessment of futile treatment in critical care

Joshua F. Wiley[1], Thanh N. Huynh[2] MD, Neil S. Wenger[3] MD

[1]UCLA Department of Psychology & Institute for Digital Research and Education [2]Department of Medicine, Division of Pulmonary and Critical Care Medicine [3]Department of Medicine, Division of General Internal Medicine and Health Services Research

## Background

Medical research often involves hierarchical data structures requiring the use of mixed effects models

Primary outcomes are often categorical, such as whether a patient develops a disease or not

Generalized linear mixed models (GLMMs) account for hierarchical structure and categorical outcomes, but are instable for rare outcomes and small samples. Confidence intervals for predicted probabilities are not well defined.

*Bayesian GLMMs incorporate prior information helping with small samples, and posterior predicted probabilities can be drawn from a distribution to easily obtain estimates and credible intervals*

## Method & Sample

- Patients (n = 1125) in critical care units in one health system had their treatment assessed over multiple days by treating physicians (n = 36)
- Physicians assessed treatment as *not futile*, *probably futile*, or *definitely futile* each day
- A total of 6,897 observations were cross classified in both patients and physicians
- Analyses used a cross classified random effects Bayesian ordered probit model using the MCMCglmm package (Hadfield, 2010) in R 2.15.2 (R Core Team, 2012)

## Technical Details

- **Predictor effects** are presented as the average marginal change in predicted probability for a one unit change in the predictor with 95% highest posterior density intervals

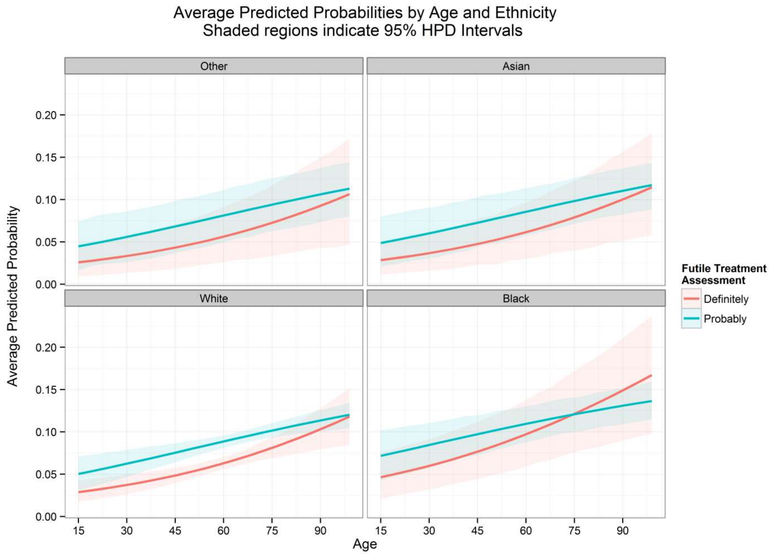
$$marginal\ effect = E\left[\frac{P(Y|X = x + h) - P(Y|X = x)}{h}\right]$$

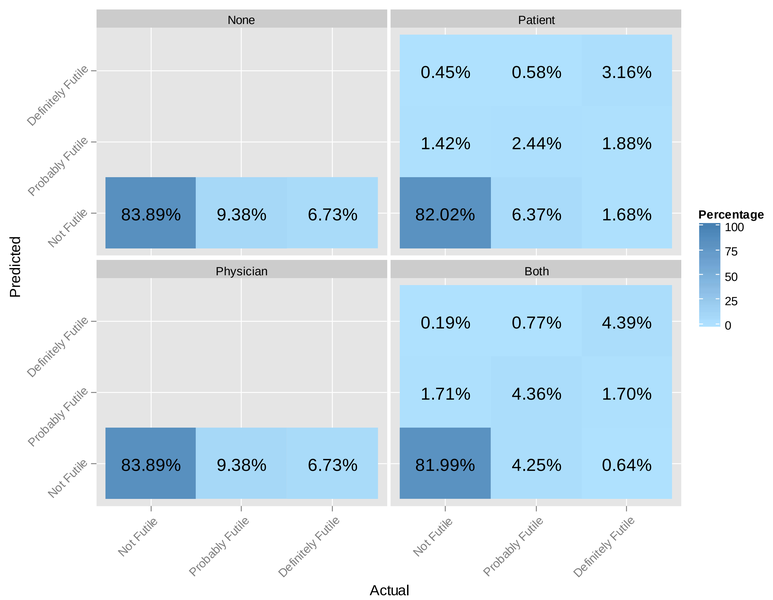
- **Predicted values** are presented as predicted probabilities with 95% HPD intervals
- **Model accuracy** is presented as the percentage of correctly predicted futility assessments. Classifications were based on the highest model predicted probability
- **Random effects** are often presented as variance of log odds. We start with fixed effects associated with a probability of .5 of not receiving futile treatment. The random effects are added to these and then transformed to probabilities and the standard deviation of these computed

An `R` package, **postMCMCglmm** was developed to facilitate calculation of each of these measures. It is available from https://github.com/jwiley/postMCMCglmm

## Average Marginal Predicted Probabilities

| | Not Futile | Probably Futile | Futile |
|---|---|---|---|
| Age | -.02 [-.03, -.01] | .01 [.0, .01] | .01 [.01, .02] |
| Other | .02 [-.06, .09] | -.01 [-.04, .03] | -.01 [-.04, .04] |
| Asian | 0 [-.06, .07] | -.00 [-.03, .02] | -.00 [-.04, .04] |
| Black | -.06 [-.12, .01] | .02 [-.00, .04] | .04 [-.01, .08] |




- *Top*: Predicted probabilities by race and age. Probabilities were calculated by setting everyone in the sample to each age and race, and then averaging across the probabilities
- *Bottom*: Percent correctly and incorrectly classified varying whether patient and physician random effects are included.

## Random Effects

For a baseline probability of non futile assessment, of .5 the average between patient variability on the probability scale was SD = .396, 95% HPD [.383, .409]

The average between physician variability on the probability scale was SD = .053, 95% HPD [.045, .061]

Including both random effects, the average variability on probability scale was SD = .399, 95% HPD [.386, .413]

## Conclusion

For categorical data, commonly used probit or logit scales are difficult to interpret.

An interpretational difficulty with probabilities is that they are nonlinear, we address this by calculating the *average marginal change*, which is the average effect in the sample.

Drawing from the posterior distribution from bayesian GLMMs, it is conceptually straightforward to calculate point estimates such as average marginal change, or average predicted probabilities, and to provide credible intervals on how precise the estimates are.

For ordinal probit models in `R`, `recycler()`, `predict2()`, and `stdranef()` from the **postMCMCglmm** package make it easy to compute average marginal probabilities, predicted probabilities, and random effects, respectively, on the probability scale, instead of the default probit scale.

*For categorical data, presenting results on the probability scale facilitates interpretation and is more familiar to researchers and practitioners*

## References

- Hadfield, J. D. (2010). *MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R package*. Journal of Statistical Software, 33(2), 1-22, URL http://www.jstatsoft.org/v33/i02/.
- Huynh, T. N., Kleerup, E. C., Wiley, J. F., Savitsky, T. D., Guse, D., Garber, B. J., & Wenger, N. S. (submitted for review). *The frequency and cost of futile treatment in critical care*.
- R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Funding/Support: This project was supported by Award Number T32GM084903 from the National Institute of General Medical Science of the National Institutes of Health. The content is solely my responsibility and does not necessarily represent the official views of the National Institute of General Medical Science or the National Institutes of Health.
Data collection was supported by a donation from Mary Kay Farley to RAND Health
Contact information is available at the author's website: http://joshuawiley.com/