A Two-part Mixed-effects Model for Longitudinal Outcomes with "Don't Know" Category: Implication for Program Evaluation

Beili Huang[1], Depeng Jiang[1,2]

[1] Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba

[2] The George Fay Yee Center for Healthcare Innovation, University of Manitoba, Winnipeg, Manitoba

Background

- "Don't Know (DK)" category has been increasingly used in surveys as an alternative for respondents
- Higher frequencies of DKs have been consistently reported among respondents with vulnerable characteristics in health-related research, especially socio-demographics
- Challenges: Although reducing item non-responses and guess work in some extent, DKs create unique challenges in data analysis and program evaluation, especially for longitudinal data
- Current Limitations: Common approaches such as treating DKs as missing data will not only lose meaningful information, but also lead to biases and loss of precision on the estimates of intervention effects

Possible Reasons of DK

- Lack of knowledge
- Lack of involvement
- Lack of interest
- Satisficing
- Self-image protection
- ...

Objectives

- Develop a statistical model for longitudinal responses with DK category
- Conduct simulation studies to illustrate the merits of the proposed model over other alternative methods
- Apply the proposed model in program evaluation and compare it with other methods, e.g., treating DKs as missing data

Methods

Model Development

- A two-part mixed-effects model will be developed (i: subject, t: time point):
  - One part characterizing DKs (Pr(y='DK')) over time using a logistic mixed model (LMM)
  - The other part modeling the evolving patterns in longitudinal outcomes (y|y≠'DK') using a non-linear mixed model (NLMM)
  - The random effects of two parts (αi & βi) are assumed to be correlated (correlation coefficient: ϱ) to account for associations between DKs and longitudinal outcomes

$$\text{LMM: } logit(Pr(y_{it} = 'DK')) = X\lambda + \alpha_i$$

$$\text{NLMM: } E(\eta(y_{it}) \,|\, y_{it} \neq 'DK') = X\gamma + \beta_i$$

$$\text{Correlation: } \begin{pmatrix}\alpha_i \\ \beta_i\end{pmatrix} \sim N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2\end{pmatrix}\right)$$

Simulation Studies

- Simulation studies will be conducted to illustrate the advantages of the proposed model on parameter estimates over other methods under different conditions of sample size, DK proportions, and correlation strengths.

Application in Intervention Program

- The proposed model will be applied to evaluate an intervention program, and compared with other approaches treating DKs as missing data, including deleting DKs and multiple imputation of DKs
- Data: PROJECT 11, a school-based mental health prevention program in Manitoba.

Significance

- Deeper understanding of special item non-responses like DKs in longitudinal research
- Improvements on the conclusion quality in program evaluation, and potential to identify marginal / vulnerable populations and inform more targeted preventions and interventions

Acknowledgements

- This project has been presented at Canadian Statistical Student Conference (CSSC) and got third prize in Master's group.
- I would like to acknowledge the support of the Visual and Automated Disease Analytics (VADA) Program for their funding
- I appreciate my supervisor Dr. Depeng Jiang for his guidance.

For any questions, please contact Beili Huang:
E-mail: huangb2@myumanitoba.ca