Bayesian Computation for Semi-continuous Longitudinal Outcomes with Non-ignorable Missing Data

University of Manitoba

Dr. Depeng Jiang[1], Dr. Puying Zhao[2], Dr. Niansheng Tang[3]

1. University of Manitoba, Canada; 2. University of Waterloo, Canada; 3. Yunnan University, China

Introduction

Managing "the tail of the curve" - The course, predictive factors and work-related outcomes of injured workers: A prospective cohort study conducted at two Workplace Safety & Insurance Board (WSIB) specialty clinics.

- **Objectives**
  - To describe the course and outcomes of workers after attending the WSIB Specialty clinics for upper limb disorders.
  - To evaluate the modifiable factors associated with more successful outcomes.
- **Methods and Sample**
  - 614 clinic attendees recruited, 595 with useable baseline data.
  - Self-report variables on work status, work disability, physical functioning and perception of recovery were assessed at clinic attendance and 3, 6, and 12-months post visit.
  - At baseline, >30 known prognostic factors of return-to-work (RTW) outcomes were fielded.

Outcome measures

- At-work disability was measured by Work Limitation Questionnaire (WLQ, Amick et al., 2004; Lerner et al., 2002). WLQ-index scores were weighted sums of items from four domains (time management, physical demand, mental-interpersonal and output demands).
- WLQ-index scores ranged from 0 to 28.6, with higher score indicating most limitations.
- WLQ-index scores are relative measures --- loss in at-work productivity as a result of health disorder. Studies have shown that 1 point higher indicates about 0.9-1% productivity loss.

Missing data problems

- WLQ only captured for those workers who were working at the time of assessment and was missing for those off work (missing by design).
- Study dropouts
  - Attrition rate: 3-month (16.5%); 6-month (17.3%); 12-month (26.1%).
  - Mechanism for dropout was unknown (MAR or NMAR?)

Three-part mixture model

- Multiple latent classes are used to represent the missing mechanism (m part) and work status (u-part), as well as to represent the growth in the non-missing outcomes (y-part).
- The mixture is represented by c, a latent categorical variable with K classes.

- M1: latent class model for working status and dropout mechanism.

$$u_{it} = \begin{cases} 1 & \text{if at-work} \\ 0 & \text{if off-work} \end{cases} \qquad m_{it} = \begin{cases} 1 & \text{if dropout} \\ 0 & \text{if not dropout} \end{cases}$$

$$P(u_1 = 1, \cdots, u_4 = 1; m_2 = 0, \cdots, m_4 = 0) = \sum_{k=1}^{K} P(c = k) P(u_1 = 1 \mid c = k) \cdots P(m_4 = 0 \mid c = k)$$

- M2: linear mixed model for modeling change in at-work disability conditional on an individual being in class k.

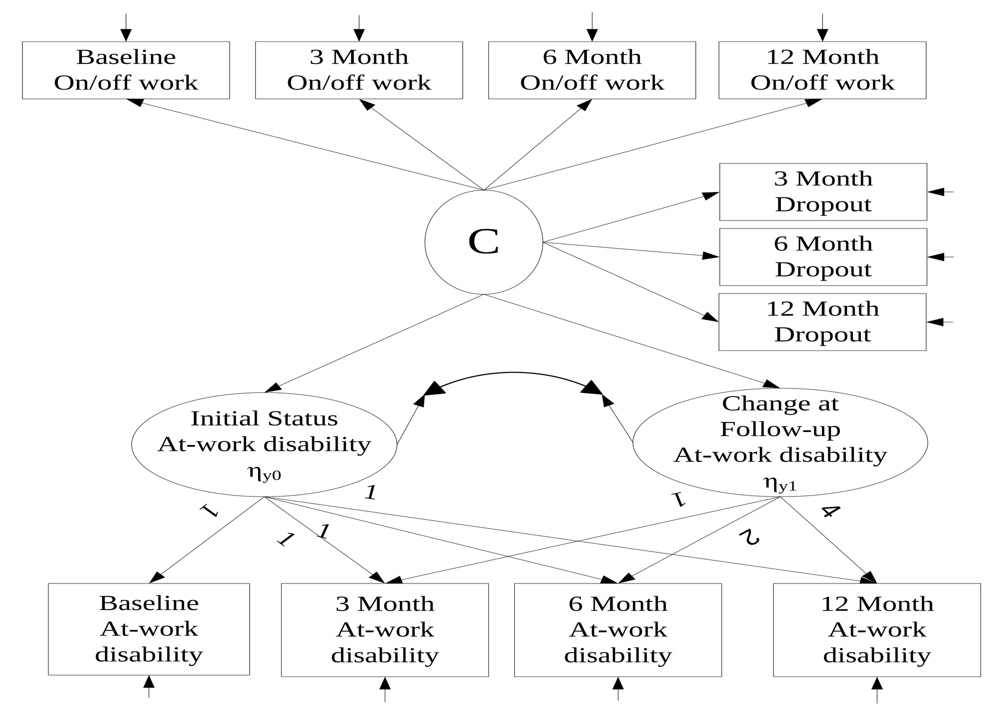
$$g(y_{it} \mid u_{it} = 1, m_{it} = 0, c = k) = \eta_{y0ik} + \eta_{y1ik} a_{it} + \varepsilon_{it}$$

$$\eta_{0ik} = \alpha_{0k} + \mu_{0ik}$$

$$\eta_{1ik} = \alpha_{1k} + \mu_{1ik}$$

$$\varepsilon_{\mathrm{it}} \sim \mathrm{N}(0, \Psi_{\mathrm{k}}) \qquad \begin{pmatrix} \mu_{0\mathrm{ik}} \\ \mu_{1\mathrm{ik}} \end{pmatrix} \sim \mathrm{N}(0, \Phi_{\mathrm{k}})$$

Three-part mixture model: Structural Equation Modelling (SEM) Graphic Framework



Bayesian estimation

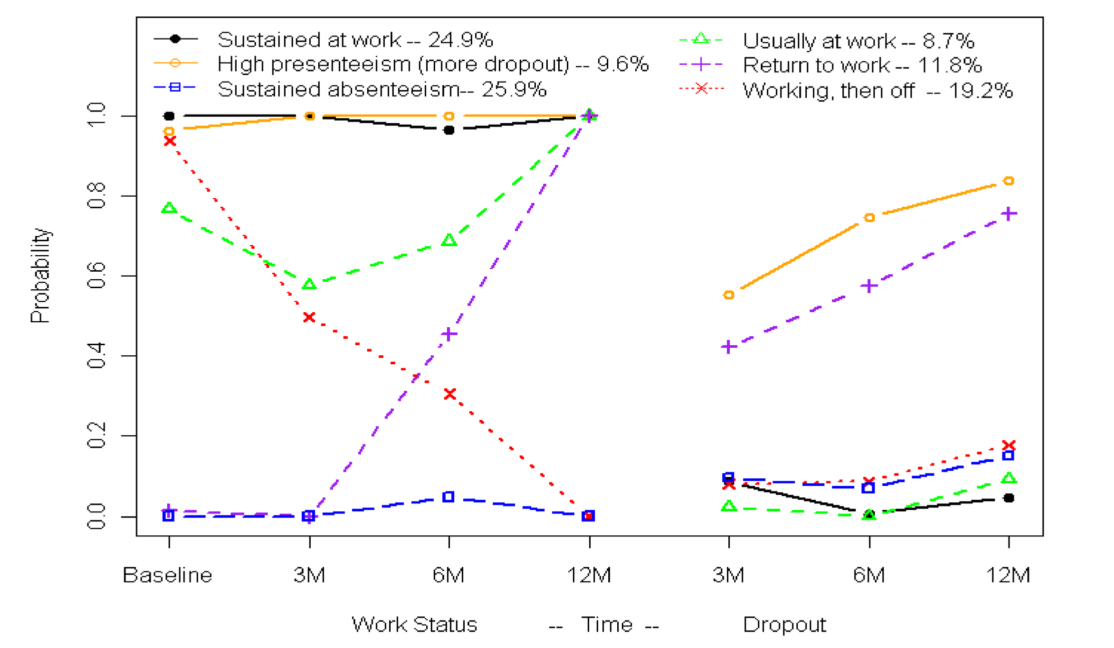
- Let θ be the overall parameter vector, Y and U be the observed data matrix for y-part and u-part, Ω be the matrix of latent vectors (η_0, η_1).
- We introduce a class label w_i for the i^{th} observation y_i as a latent allocation variable. Let W be the vector of allocation variables.
- The main task is to simulate a sufficiently large sample of observations from $[(\theta, \Omega, W) \mid (Y, U)]$. This is done by the Gibbs sampler (Geman and Geman, 1984) as follows. At the *r*th iteration with current values $[\theta^{(r)}, \Omega^{(r)}, W^{(r)}]$:
  - A. Generate $[W^{(r+1)}]$ from $[p(W) \mid (Y, U, \theta^{(r)})]$
  - B. Generate $[\Omega^{(r+1)}]$ from $[p(\Omega) \mid (Y, U, \theta^{(r)}, W^{(r+1)})]$
  - C. Generate $[\theta^{(r+1)}]$ from $[p(\theta) \mid (Y, U, \Omega^{(r+1)}, W^{(r+1)})]$
- We have used conjugate type prior distributions for various components of θ. The prior distribution of latent class proportions is taken as the symmetric Dirichlet distribution, that is, $\pi \sim D(\alpha, \alpha, \ldots, \alpha)$ with probability density function given by

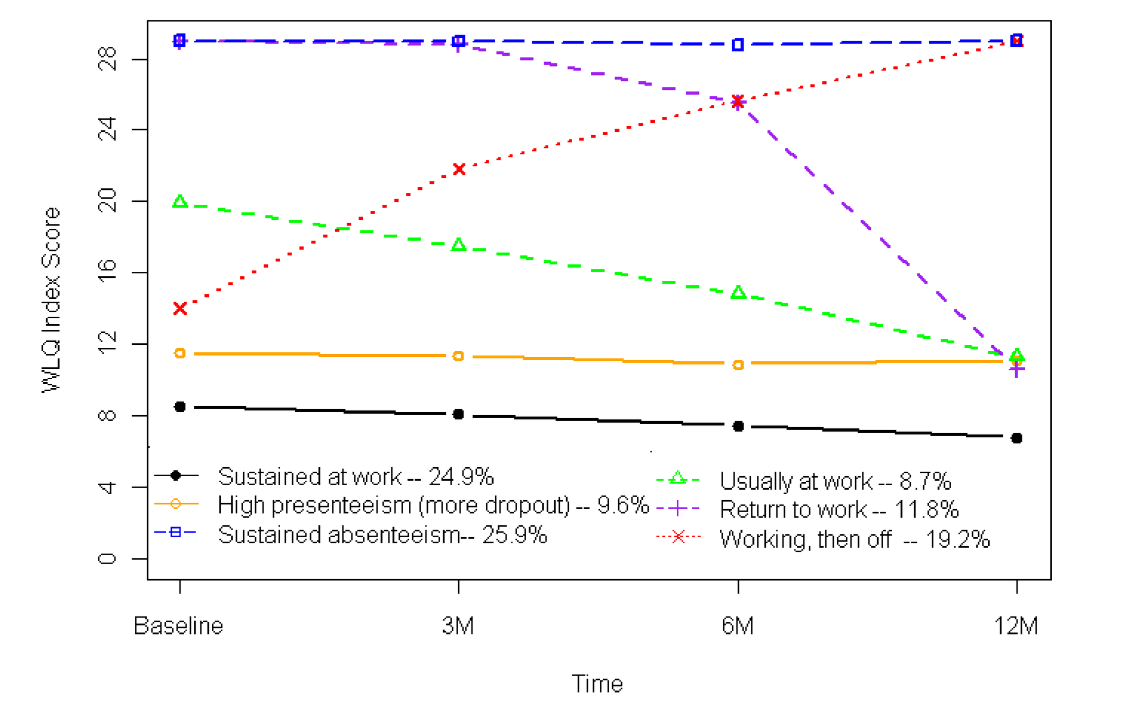
$$p(\pi) = \frac{\Gamma(K\alpha)}{\Gamma(\alpha)} \pi_1^{\alpha} \cdots \pi_{\mathrm{K}}^{\alpha}$$

- We will use the he Bayes factor (Kass and Rafety, 1995) to compare two mixture models with different number of latent class. But the path sampling procedure to calculate the Bayes factor is still on going.

Results

- Jointly modeling of u, m, and y parts and further explorations resulted in 6-class model considering predictive validity and practice usefulness.





	Class	Dropout rate at 3-, 6-, 12M	At-work disability	P-value for change in at-work disability
1	Sustained at work (24.9%)	<10%, <10%, <15%	Decreased 20.7% over 12 months	0.006
2	High presenteeism (9.6%)	54%, 75%,86%	No significant change	0.92
3	Sustained off work (25.9%)	<10%, <10%, <15%	Can not evaluate (no measures)	NA
4	Working at most time points (8.7%)	<10%, <10%, <15%	Decreased 44.3% over 12 months	<.001
5	Initially off work with RTW (11.8%)	46%, 60%, 76%	Decreased 80.5% over 12 months	<.001
6	Initially at work, later off (19.2%)	<10%, <10%, <15%	Increased over 60% in 3 months	<.001

Conclusion

- **Dropout is non-ignorable missing. The likelihood of dropout, the level of presenteeism and level of at-work disability were significantly associated.**
- **Bayesian SEM – mixture model - jointly model the missing mechanism (m-part), binary part (u-part) and continuous part (y-part) of outcome.**

Acknowledgements

We would like to express our thanks to Dr. Dorcas Beaton at St Michael's Hospital, Toronto for data use, and the support from the Manitoba Health Research Council (MHRC) Establishment Award and Research Manitoba Applied Health Services Program.