# Pattern Mixture Models for the Analysis of Repeated Attempt Designs

M. J. Daniels, Wei Feng, Dan Jackson, Ian White

UT-Austin, UF, MRC Biostat-Cambridge (UK)

NISS, November 2015

# Setting and Problem

- follow-up studies where multiple attempts are made to collect a measurement after baseline (called *repeated attempt designs*)
- often reasonable to view information about repeated attempts as representing a 'continuum of resistance' to providing data; repeated attempt models (RAMs) exploit this feature
- current literature focuses on selection model factorizations (Alho, 1990; Wood et al., 2006; Jackson et al., 2012; Qin and Follman, 2014); refer to as RAM-SM
  - problem: not allow sensitivity parameters (and sensitivity analysis) which is an essential component of inference for missing data

# Motivating trial I

- QUATRO trial was a single-blind, multi-center randomized controlled trial of the effectiveness of adherence therapy for participants with schizophrenia.
- trial included 409 participants in four centers
- randomized to receive either adherence therapy (intervention) or health education (control).
- assessments were undertaken at baseline and at a follow-up of 52 weeks.

# Motivating trial II

- Objective: Assess the impact of adherence therapy on self-reported quality-of-life of people with severe mental illness
- investigators made multiple attempts to collect the quality-of-life outcome at '52 weeks' (as many as nine attempts) but there were still individuals whose response could not be collected.
  - concern that those with less favorable outcomes may be less likely to provide data

# Notation

- $Y$: the outcome of interest (here, self reported QoL at 52 weeks)
- $X$: the set of baseline covariates (here, baseline QoL and center)
- $R$: the number of attempts until the outcome is successfully collected
  - assume up to $K$ attempts to collect ($R = K + 1$ corresponds to the outcome not being collected after the maximum number of attempts)
- $Z$: randomized intervention (here, $Z = 1$ is adherence therapy)
- $[Y|X, Z, R = K + 1]$: (unidentified) extrapolation distribution

# Inferential goal

The quantity of interest is the treatment effect on the mean outcome, unconditional on $X$ and $R$,

$$\theta = E(Y|Z=1) - E(Y|Z=0).$$

# Model - RAM-PMM I

- use a pattern mixture formulation of the RAM (RAM-PMM); patterns are defined by the values of $R$.
- For pattern $R = k$ $(k = 1, \ldots, K+1)$ and arm, $Z = z$ $(z = 0, 1)$, we consider the following model for the conditional distribution of $Y$

$$Y|Z = z, X = x, R = k \sim N\{\mu(z, x, k), \sigma^2(z, x, k)\}$$

  - allows the mean and variance of the outcome data to depend on covariates and patterns

# Model - RAM-PMM II

- In our modeling, we assume

$$\mu(z, x, k) = \alpha_z^{(k)} + \beta_1 x, k \leq K$$

  and a constant variance, $\sigma^2(z, x, k) = \sigma^2$.
- estimating the treatment effect, $\theta$ also requires identifying $\mu(z, x, K+1)$, i.e., the mean of the *extrapolation distribution* (do via priors in what follows)

# Model - RAM-PMM III

- also need a model for the conditional distribution of the pattern indicator,

$$\text{logit}\{\pi_{kz}(x)\} = \lambda_{0k} + \lambda_z z + \boldsymbol{\lambda}_x x,$$

  where $\pi_{kz}(x) = P(R = k | R \geq k, Z = z, X = x)$.
- and a model for $[X|Z] = [X]$ (by randomization)
- factorization respects the fact, which is sometimes overlooked in pattern mixture models, that the distribution of the baseline outcome (which is included in $X$ here) does not depend on $Z$.

# Model - RAM-PMM IV

- for the RAM-PMM,

$$\theta = E[Y|Z=1] - E[Y|Z=0] = \int\int \mu(1,x,k)dF(k|x,Z=1)dF(x)$$
$$- \int\int \mu(0,x,k)dF(k|x,Z=0)dF(x)$$

# Priors I

- identified parameters, $(\{\alpha_z^{(k)} : k = 1, \ldots, K\}, \boldsymbol{\beta}_1, \sigma, \boldsymbol{\lambda})$, are given diffuse priors
- Unidentified parameters: $\alpha_z^{(K+1)}$ (sensitivity parameters)
  - exploit the repeated attempt design and assume a functional relationship between the intercept parameters for the observed outcomes, $\{\alpha_z^{(k)} : k = 1, \ldots, K\}$ and the number of attempts $(k)$ to identify these parameters
  - in particular, specify a prior for $\alpha_z^{(K+1)}$ conditional on $\bar{\alpha}_z^{(K)} = (\alpha_z^{(1)}, \ldots, \alpha_z^{(K)})^T$, i.e. $p(\alpha_z^{(K+1)} | \bar{\alpha}_z^{(K)})$.

# Priors II

- center this prior at its prediction based on implicitly fitting the regression

$$\alpha_z^{(k)} = h_z(k;\boldsymbol{\zeta}) + \epsilon_{zk}, \quad k = 1, \ldots, K.$$

  - here, set $h_z(k;\boldsymbol{\zeta}) = \zeta_{0z} + \zeta_{1z}k, \quad k = 1, \ldots, K$ and compute the least squares estimate of $(\zeta_{0z}, \zeta_{1z})$ to obtain

$$\alpha_z^{(K+1)} | \bar{\alpha}_z^{(K)} \sim N\{\hat{\zeta}_{0z} + \hat{\zeta}_{1z}(K + C), \tau^2\}.$$

    where $\hat{\zeta}_{jz}$ are functions of $\{\alpha_z^{(1)}, \ldots, \alpha_z^{(K)}\}$.

- We assume that the intercepts $\alpha_z^{(k)}$ follow a linear trend over patterns that provide outcome data (recall continuum of resistance)

# Priors III

$$\alpha_z^{(K+1)} | \bar{\alpha}_z^{(K)} \sim N\{\hat{\zeta}_{0z} + \hat{\zeta}_{1z}(K + C), \tau^2\}.$$

- sensitivity parameters are $C$ and $\tau$
  - $C$ represents how far we should extrapolate the linear trend to describe the missing outcome data (i.e., how resistant are those that have not provided outcome data by the $K$th attempt)
  - $\tau$ represents our uncertainty about the precision of the extrapolation.
- This approach does *not* put any modelling restrictions on the observed data, but still attempts to use information in an intuitive manner from the repeated attempt design.

# Analysis of QUATRO I

| | $Y$ observed after $k$ attempts | | | | | | | | | $Y$ not obs |
|---|---|---|---|---|---|---|---|---|---|---|
| $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| Control | 77(42.4) | 94(41.3) | 7(38.7) | 7(34.7) | 3(34.2) | 2(32.9) | 1(40.7) | 1(62.98) | 0(NA) | 13 |
| Treatment | 73(40.7) | 90(40.2) | 7(38.6) | 1(45.7) | 3(35.0) | 0(NA) | 0(NA) | 1(30.3) | 0(NA) | 29 |

- Up to 9 attempts were made to collect the 52 week outcome for participants.
- sparsity of subjects with 3 to 9 attempts on each arm; merged those subjects into one pattern

# Analysis of QUATRO II

| | $Y$ observed | | | $Y$ missing | |
|---|---|---|---|---|---|
| $R$ | 1 | 2 | 3 | $<4$ | 4 |
| Control ($n = 205$) | 77(42.4) | 94(41.3) | 21(37.4) | 8 | 5 |
| Treatment ($n = 204$) | 73(40.7) | 90(40.2) | 12(37.6) | 20 | 9 |

- an overall decreasing outcome mean with # of attempts
- the number of attempts, $R$ takes values in $\{1, 2, 3, 4\}$ (i.e., $K = 3$).
  - $R = 4$ corresponds to the pattern that $Y$ is not observed even after all attempts.
  - Individuals with $Y$ missing, but fewer than three attempts, have $R$ censored; 8 and 20 subjects censored respectively on the two arms
- Covariates $X$ are indicators of the four centers and the baseline response.

# Analysis of QUATRO III

- vary $C$ between
  - 0 (missing subjects are comparable to the last responders) and
  - 3 (missing subjects differ from the last responders as much as the last responders differ from the first responders)
- For all values of $C$ considered, we observed a negative effect of adherence therapy on self-reported 52 week QoL

# Analysis of QUATRO IV

| $C$ | parameter | mean | 95% CI |
|---|---|---|---|
| 0 | $\theta$ | -0.6 | (-2.9, 1.7) |
| | $E(Y\|Z=0)$ | 40.9 | (39.2, 42.5) |
| | $E(Y\|Z=1)$ | 40.2 | (38.4, 42.1) |
| 1 | $\theta$ | -0.7 | (-3.1, 1.8) |
| | $E(Y\|Z=0)$ | 40.7 | (39.1, 42.4) |
| | $E(Y\|Z=1)$ | 40.0 | (38.0, 42.1) |
| 2 | $\theta$ | -0.7 | (-3.5, 2.0) |
| | $E(Y\|Z=0)$ | 40.5 | (38.8, 42.3) |
| | $E(Y\|Z=1)$ | 39.8 | (37.4, 42.1) |
| 3 | $\theta$ | -0.8 | (-3.8, 2.2) |
| | $E(Y\|Z=0)$ | 40.4 | (38.5, 42.2) |
| | $E(Y\|Z=1)$ | 39.6 | (36.9, 42.2) |

# Analysis of QUATRO V

$$\alpha_z^{(K+1)} | \bar{\alpha}_z^{(K)} \sim N\{\hat{\zeta}_{0z} + \hat{\zeta}_{1z}(K + C), \tau^2\}.$$

- the slope of the priors, $\zeta_{10}$ and $\zeta_{11}$ are both negative, with posterior means (95% credible intervals) of $-1.9$ $(-4.8, 0.97)$ and $-1.7$ $(-5.2, 1.7)$, respectively
- the slope for those on adherence therapy was slightly more extreme (i.e., doing worse as the number of attempts increases)

# Conclusions I

- We have proposed a pattern mixture model for a repeated attempt design that allows sensitivity parameters
- recommend consideration of the RAM-PMM in general as it
  - allows for sensitivity analysis, unlike the RAM-SM
  - handles the missing data similar to the RAM-SM in terms of a continuum of resistance (not shown here)
  - does not have the issue of a potential large impact on inferences of modeling choices in the missing data mechanism

# Conclusions II

- In the QUATRO study, we found minimal evidence of a significant effect of the intervention (adherence therapy) on 52 week self-reported QoL with the models considered.
- RAM-SM for QUATRO corresponded to an unreasonably extreme value of $C$ suggesting that very extreme values of the QoL were needed to make the full data response look normal; this particular flaw can be avoided to some extent using semiparametric approaches (e.g., Qin and Follman, 2014).

# Extensions

- more complex forms for the mean and variance functions, $\mu(z, x, k)$ and $\sigma^2(z, x, k)$ and less parametric specification in general (e.g., BNP)
- For repeated attempt studies with sparse patterns, adapt the ideas from Roy (2003) and Roy and Daniels (2008) to combine/collapse patterns in a data-dependent way.
- longitudinal extension