PART 1: Missing Data Mechanisms:
Overview and Examples

Short Course, Colombia, July 2011
1. Examples of missing data

2. Full vs. observed data

3. Full-data model

4. Missing data mechanisms
   - MCAR
   - MAR
   - MNAR
Examples of missing data

Dropout in a clinical trial: GH Study

- Individuals scheduled for 3 measurements
- Target of inference: mean at time 3
- This is a full data parameter
- But several individuals are missing time 3 outcome

⇒ can define full-data parameter,

but cannot identify it from observed data
Examples of missing data

Dropout and mortality: HIV cohort study

- Consider HIV cohort study with long follow up
- Measure CD4 every six months for 5 years
- Target of inference: mean change in CD4 over 5 years
- Missing data due to:
  - Dropout and withdrawal
  - Death from various causes
- What is the *full data*?
Full vs observed data

When drawing inference from incomplete data, essential to distinguish between full data and observed data.

Full data:
- Data that was intended to be collected
- Data that you would like to use as basis for inference

Observed data:
- Data that was actually collected
- Data you have available to draw inferences
Full vs observed data ($J = 2$)

Consider simple case where $J = 2$, and $Y_2$ may be missing

**Full data:** $F = (Y, X, V, R)$

Includes *response and covariates*

- $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ = response variable
- $X$ = covariates of interest
- $V$ = auxiliary covariates

And includes *missing data indicator*

$$R = \begin{cases} 
0 & \text{if } Y_2 \text{ missing} \\
1 & \text{if } Y_2 \text{ observed}
\end{cases}$$
Full vs observed data ($J = 2$)

Assume all covariates fully observed

**Observed data:** \( O = (Y_{\text{obs}}, X, V, R) \)

\[
Y_{\text{obs}} = \begin{cases} 
(Y_1, Y_2)^T & \text{if } R = 1 \\
Y_1 & \text{if } R = 0
\end{cases}
\]
Full vs observed data: $J > 2$

For the general case, same notation applies

**Full data response and indicators**

$$
\begin{align*}
Y &= (Y_1, \ldots, Y_J)^T \\
R &= (R_1, \ldots, R_J)^T
\end{align*}
$$

where

$$
R_j = \begin{cases} 
0 & \text{if } Y_j \text{ missing} \\
1 & \text{if } Y_j \text{ observed}
\end{cases}
$$
Partition responses into observed and missing components

\[ Y_{\text{obs}} = \{ Y_j : R_j = 1 \} \]
\[ Y_{\text{mis}} = \{ Y_j : R_j = 0 \} \]

Sometimes write full-data response as

\[ Y = (Y_{\text{obs}}, Y_{\text{mis}}) \]
Full-data model

Recall full data: $F = (Y, R, X)$

- Full data response model:
  
  This is the target of inference

- Full data model:

- Relationship between the two

  $$p(y \mid x, \theta(\omega)) = \sum_{r \in R} p(y, r \mid x, \omega),$$

  where $R$ is the sample space of $r$.

- Notice also that $\theta = \theta(\omega)$
Inference about $p(y \mid x, \theta)$ depends critically on assumptions about $p(y, r \mid x, \omega)$.

However, observed data not sufficient to identify full-data model.

To see this, notice that

$$p(y, r \mid x, \omega) = p(y_{obs}, y_{mis}, r \mid x, \omega)$$

$$= p(y_{mis} \mid y_{obs}, r, x, \omega) \ p(y_{obs}, r \mid x, \omega)$$

Hence, untestable assumptions needed.

These are formalized in terms of missing data mechanisms.
Any full-data model can be factored as

\[ p(y, r \mid x, \omega) = p(y \mid x, \theta(\omega)) \cdot p(r \mid y, x, \psi(\omega)). \]

The factor

\[ p(r \mid y, x, \psi(\omega)) \]

is the missing data mechanism.

Useful to rewrite as

\[ p(r \mid y, x, \psi) = p(r \mid y_{\text{obs}}, y_{\text{mis}}, x, \psi) \]
Missing data mechanism: MCAR

Missing completely at random (MAR)

Data are missing completely at random (MCAR) when dropout is unrelated to all elements of $\mathbf{Y}$.

$$p(r \mid y, x, \psi) = p(r \mid x, \psi)$$
Missing data mechanisms

Example

- Longitudinal survey of school test scores
- Measure at years 1 and 2, record as \((Y_1, Y_2)\)
- Sample 1000 students
- \(Y_1\) observed for everyone
- \(Y_2\) observed for randomly chosen subsample of 800

\[
P(R_2 = 1 \mid Y_1, Y_2) = P(R_2 = 1)
\]
Missing at random (MAR)
Dropout unrelated to $Y_{mis}$, conditionally on $Y_{obs}$

$$p(r \mid y_{obs}, y_{mis}, x, \psi) = p(r \mid y_{obs}, x, \psi)$$
Missing data mechanism: MAR

Example (test scores, continued)

- Intend to collect \((Y_1, Y_2)\) on 1000 students
- Those with higher \(Y_1\) more likely to show up for second test
- Among those with given score \(Y_1\), score on \(Y_2\) does not predict missingness
- Hence \(P(R_2 = 1)\) depends on \(Y_1\) but not \(Y_2\)

\[
P(R_2 = 1 \mid Y_1, Y_2) = P(R_2 = 1 \mid Y_1)
\]
MAR and dropout

- MAR applied to dropout

\[ h(t_j \mid \overline{Y}_j) = h(t_j \mid \overline{Y}_{j-1}) \]

- In words: hazard of dropout at \( t_j \)
  - can depend on observable past responses
    \[ Y_1, \ldots, Y_{j-1} \]
  - conditional on the past, cannot depend on possibly missing current and future responses
    \[ Y_j, Y_{j+1}, \ldots, Y_J \]
Missing data mechanism: MNAR

Missing not at random (MNAR)

Dropout may depend on $\mathbf{y}_{\text{mis}}$ after conditioning on $\mathbf{y}_{\text{obs}}$

\[
p(r \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \mathbf{x}, \psi) \neq p(r \mid \mathbf{y}_{\text{obs}}, \mathbf{y'}_{\text{mis}}, \mathbf{x}, \psi)
\]

for $\mathbf{y}_{\text{mis}} \neq \mathbf{y'}_{\text{mis}}$.
Missing data mechanism: MNAR

Example: Smoking cessation study

- Smoking outcomes are \((Y_1, Y_2)\); suppose \(Y_2\) can be missing
- If MAR,

\[
P(R = 1 \mid Y_1, Y_2 = 1) = P(R = 1 \mid Y_1, Y_2 = 0),
\]

\(\implies\) dropout same for those who are smokers and nonsmokers at time 2
- Probably not true for smoking cessation studies
Model factorizations under MNAR I

Different ways are used to factor the full data distribution to construct models:

- Selection models (SM):
  \[ p(y, r) = p(r|y)p(y) \]

- (Pattern) Mixture models (MM):
  \[ p(y, r) = p(y|r)p(r) \]
Model factorizations under MNAR II

- Shared parameter models (SPM):

\[ p(y, r) = \int p(y, r|b) dF(b) \]

- often assume \( p(y, r|b) = p(y|b)p(r|b) \)

- parametric SMs and SPMs can be problematic and restrictive; MMs allow more flexibility (more later)
Sidebar about covariates

- **Model covariates are** $X$
  - These are the covariates of direct interest
  - Usually interested in $E(Y \mid X)$

- **Auxiliary covariates** $V$
  - Information is available but not of direct interest
  - $E(Y \mid X) \neq E(Y \mid X, V)$
  - Important (and problematic) if need to condition on these for MAR (see research talk)
Part 2

Inference under ignorable missingness: Definitions, issues and a case study

Short Course, Colombia, July 2011
1. Ignorable missingness

2. Dependence and Covariance models

3. Data Augmentation

4. Case Study I: GH
Ignorability

A missing data mechanism is **ignorable** if the following three conditions hold:

1. The missing data mechanism is **MAR**

2. The full data parameter $\omega$ can be decomposed as $\omega = (\theta, \psi)$, where
   - $\theta$ indexes the full-data response model $p(y | \theta)$, and
   - $\psi$ indexes the missing data mechanism $p(r | y, \psi)$.

3. The parameters $\theta$ and $\psi$ are a-priori independent; i.e.,

   $$p(\theta, \psi) = p(\theta)p(\psi).$$
Key implication of ignorability is that the likelihood can be factored over $\theta$ and $\psi$

$$L(\theta, \psi \mid y_{\text{obs}}, r) = L_1(\psi \mid r, y_{\text{obs}}) \cdot L_2(\theta \mid y_{\text{obs}}).$$

Observed-data posterior also factors as

$$p(\theta, \psi \mid y_{\text{obs}}, r) \propto \{p(\psi)L_1(\psi \mid r, y_{\text{obs}})\} \cdot \{p(\theta)L_2(\theta \mid y_{\text{obs}})\}$$

Observed-data posterior for $\theta$ is

$$p(\theta \mid y_{\text{obs}}) \propto p(\theta) \cdot L_2(\theta \mid y_{\text{obs}})$$

$\implies$ not a function of $\psi$
Ignorability: practical implications

Key implication: missing data mechanism factors out of observed-data likelihood

Consequence:

- missing data mechanism $p(r | y, \psi)$: does not have to be specified
- full-data model $p(y | x, \theta)$: must be specified correctly; dependence is very important
- But remember: no evidence to check this. We are relying on MAR

Does not mean: missing data are ignored
Importance of modelling the dependence

- mis-specification of dependence leads to biased mean parameters under ignorability (and nonignorability)
- different from complete data: often just an issue of loss of efficiency
- demonstrate via simple example: a bivariate response (GH, month 0,12) from a randomized controlled trial with binary treatment $x$ and response vector $\mathbf{Y}_i = (Y_{i1}, Y_{i2})^T$.
- $Y_{i1}$ is always observed, but $Y_{i2}$ may be missing
- missingness indicator,

$$R_i = \begin{cases} 
0 & Y_{i2} \text{ missing} \\
1 & Y_{i2} \text{ observed} 
\end{cases}$$
Importance of modelling the dependence II

- assume the full data model, $p(y; \theta)$ is a bivariate normal distribution

\[
Y_i \mid \mu, \Sigma, x \sim N(\mu(x), \Sigma(x))
\]

with the covariance matrix $\Sigma(x) = \{\sigma_{jk}(x)\}$.

- recall that we can re-parameterize this model (based on the sequential factorization $p(y_1)p(y_2|y_1)$) using

\[
E[Y_1 \mid x] = \mu_1(x) \\
\text{Var}[Y_1 \mid x] = \sigma_{11}(x) \\
E[Y_2 \mid Y_1, x] = \beta_0 + \phi_{21}(x)y_1 \\
\text{Var}[Y_2 \mid Y_1, x] = \sigma_{22}^*(x)
\]
suppose the dependence parameter $\phi_{21}(x) = \sigma_{21}(x)/\sigma_{11}(x)$ is incorrectly assumed to be 0

$$\phi_{21}(x) = 0$$
Importance of modelling the dependence IV

Let $\mu_1^c(x) = E[Y_1 \mid x, R = 1]$.

Can show $E[\mu_2(x) \mid y_{obs}]$ has (asymptotic) bias

$$\phi_{21}(x) \left\{\mu_1(x) - \mu_1^c(x)\right\}$$

Degree of bias related to

- magnitude of $\phi_{21}(x)$
- how much the means at time 1 differ between completers and dropouts, $\mu_1(x) - \mu_1^c(x)$
Importance of modeling the dependence

- bias in treatment effect, \( \mu_2(1) - \mu_2(0) \)

\[-\phi_{21}(1)[\{\mu_1(1) - \mu_1^c(1)\} + \phi_{21}(0)[\{\mu_1(0) - \mu_1^c(0)\}] \]

- depends on relative differences between completers and dropouts on each treatment weighted by the respective dependence parameters

- if the dependence parameter and difference between completers and dropouts are all the same for each treatment, no bias

- if mis-specify as \( \phi_{21}(x) = \phi_{21} \), similar biases
Bias in Growth Hormone Study

For GH trial, consider month 0 and 12 on treatment EG

- $\mu^c_2 = 88$
- $\mu_1 - \mu^c_1 = 69 - 78 = -9$
- $\phi_{21} = .56$

So bias if assume $\phi_{21} = 0$ is

$$ -\phi_{21}[\mu_1 - \mu^c_1] = -.56 \times (-9) = 5.0 \text{ (Large!)} $$
Bayesian covariance models for longitudinal data

- focus on covariance models for multivariate normal model (MVN) and multivariate probit model (MVP)
- the latter requires models for a correlation matrix
- will review models based on the modified Choleski decomposition of a covariance matrix (Pourahmadi, 1999); related models have been explored by Zimmerman and Nunez-Anton (2000)
- also will review models based on re-parametrizing the off-diagonal elements of a correlation matrix (partial correlations) with partial autocorrelations
GARP/IV: the decomposition

Modified Choleski (GARP/IV) decomposition (Pourahmadi, 1999; Daniels and Pourahmadi, 2002):

- $\Sigma^{-1} = TD^{-1} T'$
  - $T$ unique unit lower triangular with 1's on its diagonal
  - $D$ diagonal with positive diagonal entries
elements are the autoregressive coefficients and prediction variances from the sequential conditional distributions,  
\[ p(y_j \mid y_1, \ldots, y_{j-1}) \]

- generalized autoregressive parameters (GARP) are the regression coefficient, \( \phi \) in

\[
E[Y_j \mid y_1, \ldots, y_{j-1}] = \mu_j + \sum_{k=1}^{j-1} \phi_{jk}(y_k - \mu_k)
\]

- innovation (prediction) variances,

\[
\text{Var}[Y_j \mid y_1, \ldots, y_{j-1}] = \sigma_j.
\]

for positive definite
Interpretation of GARP/IV II

- no restrictions on $\phi$
- $\sigma_j^2 > 0$
Properties

- a parameterization popular for covariance matrices due to
  - computational simplicity (conditional conjugacy)
  - interpretation of the elements
- but, obviously depends on the ordering (e.g., longitudinal data)
Correlations/variances

- $\Sigma = S(\sigma)R(\rho)S(\sigma)$
- $S(\sigma)$ is a diagonal matrix of standard deviations and $R(\rho)$ is a symmetric positive definite matrix (with ones on the diagonal) of marginal correlations (i.e., a correlation matrix)
- $R = \{\rho_{jk}\}$, with $\rho_{jk} = Cor(Y_j, Y_k)$
Partial autocorrelations (ordered)

- $\Sigma = S(\sigma)R(\pi)S(\sigma)$
- parameterize a correlation matrix $R = \{\rho_{jk}\}$ in terms of the lag-1 correlations

$$\pi_{j,j+1} \equiv \rho_{j,j+1}, j = 1, \cdots, p - 1$$

and the partial autocorrelations

$$\pi_{jk} = \rho_{jk|j+1,\cdots,k-1} : k - j \geq 2$$

- $\pi_{jk}$'s can vary freely in the interval $(-1, 1)$; domain of $\pi$ is the $p(p-1)/2$-dimensional hypercube
- suitable for ordered data
Let $\mathbf{R}[j : j + k]$ be the submatrix of the correlation matrix $\mathbf{R}$ containing rows and columns from $j$ to $j + k$, i.e.

$$\mathbf{R}[j : j + k] = \begin{pmatrix} 1 & r'_1(j, k) & \rho_{j,j+k} \\ r_1(j, k) & R_2(j, k) & r_3(j, k) \\ \rho_{j+k,j} & r'_3(j, k) & 1 \end{pmatrix},$$

where

$$r'_1(j, k) = (\rho_{j,j+1}, \ldots, \rho_{j,j+k-1}),$$

$$r'_3(j, k) = (\rho_{j+k,j+1}, \ldots, \rho_{j+k,j+k-1}),$$

and $\mathbf{R}_2(j, k)$ is the correlation matrix corresponding to components $(j + 1, \ldots, j + k - 1)$.
Connection to marginal correlations II

Then, the partial autocorrelations between $Y_j$ and $Y_{j+k}$ adjusted for the intervening variables ($\pi_{j,j+k} = g(\rho_{j,j+k}, R[j:j+k])$)

\[
\pi_{j,j+k} = \frac{\rho_{j,j+k} - r'_1(j, k) R_2(j, k)^{-1} r_3(j, k)}{D_{jk}}
\]

where

\[
D_{jk} = [1 - r'_1(j, k) R_2(j, k)^{-1} r_1(j, k)]^{1/2} * [1 - r'_3(j, k) R_2(j, k)^{-1} r_3(j, k)]^{1/2}
\]

$D_{jk}$ is a function of the partial variances.
Connection to marginal correlations III

- the function $g(\cdot)$ on the previous page that maps a correlation matrix $R$ into the partial autocorrelation matrix $\Pi$, is invertible,
- solving for $\rho_{j,j+k}$

$$\rho_{j,j+k} = r'_1(j, k)R_2(j, k)^{-1}r_3(j, k) + D_{jk} \pi_{j,j+k},$$

(2)
- easy to move between the two parameterizations
An Attractive Property

- the lag $k$ partial autocorrelations are based on conditional regressions where the conditioning sets have the same number of elements
- i.e., they are *exchangeable* in some sense (all conditioning on $k - 1$ intervening variables)
- unlike similar sequential partial autocorrelations (regression coefficients) from GARP/IV, $Y_j \mid y_1, \ldots, y_{j-1}$
- simplifies building models for the partial autocorrelations, $\pi_{j,j+k}$ as a function of lag $k$ and reducing the dimension in general
- BUT, not as computationally friendly as GARP/IV
Structures for longitudinal data based on GARP/IV I

- Structured GARP and IV models
  - model GARP
    \[ \phi_{jk} = Z_{jk} \gamma \]
  - model IV
    \[ \log(\sigma_j^2) = L_j \lambda \]
- model identification using regressograms (Pourahmadi, 1999; Pourahmadi and Daniels, 2002)
- potential structures
  - \( \phi_{jk} \) polynomial in \((j - k)\) (Toeplitz-like structure)
  - \( \phi_{j,j+k} \) polynomial in \(j\) for fixed lag \(k\) (non-stationary)
  - \( \log(\sigma_j^2) \) polynomial in \(j\)
  - can model freely without restrictions (constraints)
Specific Structures

- structure examples using $Z_{jk}$
  - linear in lag, $Z_{jk} = (1, |j - k|)$
  - stationary, $Z_{jk} = (I\{|j - k| = 1\}, \ldots, I\{|j - k| = p - 1\})$
  - lag 1 linear in time, Toeplitz, $Z_{jk} = I\{|j - k| = 1\} * j, I\{|j - k| = 2\}, \ldots, I\{|j - k| = p - 1\})$

- structure examples using $L_j$
  - linear in time, $L_j = (1, j)$
Figure: GARP versus lag
Figure: Lag 1 GARP versus time
Figure: IV versus time: ‘•’ is low severity
Structures based on partial autocorrelations/marginal variances

- model partial autocorrelations

\[ z(\pi_{jk}) = -\frac{1}{2} \log \frac{1 - \pi_{jk}}{1 + \pi_{jk}} = W_{jk} \gamma \]

- model marginal variances

\[ \log(\sigma_{jj}) = L_j \lambda \]

- similar structures to GARP/IV
Parsimonious modeling

- $\pi_{j,j+k}$ gauges the conditional (on the intervening variables) dependence between variables $k$ units apart
- expect them to be smaller for larger $k$.
- also expect a lot of partial autocorrelations to be zero
- e.g., AR(1) all partial autocor of lag 2 or higher are zero
For exploratory examine plots of

- $\{\pi_{j,j+k}; j = 1, \cdots, p - k\}$ versus lag $k = 1, \cdots, p - 1$
- $\{\pi_{j,j+k}; k = 1, \cdots, p - 1\}$ vs. $j$ for a fixed lag $k$ (nonstationary)
In general, regression models

\[ \phi_{i,jk} = w_{i,jk}^* \gamma. \]  

where \( w_{i,jk}^* \) is a unit-specific covariate vector that can induce structure and/or introduce unit-level covariates.
General Regression Model: Partial autocor

In general, regression models

$$z(\pi_{i,jk}) = w_{i,jk}^* \gamma^*. \quad (4)$$

where $w_{i,jk}^*$ is a unit-specific covariate vector that can induce structure and/or introduce unit-level covariates.
Data augmentation and Posterior Inference

- often easier to sample from the full data response model using the complete data vs with observed (incomplete) data: \( p(\theta \mid y) \) vs. \( p(\theta \mid y_{\text{obs}}) \)
- at each iteration,
  1. \( y_{\text{mis}}^{(k)} \sim p(y_{\text{mis}} \mid y_{\text{obs}}, \theta^{(k-1)}) \) [Imputation step]
  2. \( \theta^{(k)} \sim p(\theta \mid y_{\text{obs}}, y_{\text{mis}}^{(k)}) \)
- step 2. just complete data tools
- step 1. importance of dependence is clear
Consider a trivariate normal (like GH), $\mathbf{Y} \sim N(\mu, \Sigma)$ with $y_3$ missing

Step 1. Distribution of $p(y_3 \mid y_1, y_2, \theta)$ takes the form

$$Y_3 \mid y_1, y_2, \theta \sim N(\mu_3^*, \sigma_{33}^*)$$

where

$$\mu_3^* = \beta_{03} + \phi_{31} y_1 + \phi_{32} y_2$$

so dependence of $y_{\text{mis},i}(y_3)$ on $y_{\text{obs},i}(y_1, y_2)$ is governed by $\phi_{31}$ and $\phi_{32}$ (functions of the covariance matrix $\Sigma$)
Consider treatment EG again
What is the data augmented mean for a subject with $Y_1 = 65$ and $Y_2 = 71$?

- $E[Y_3|Y_1, Y_2, \mu, \beta, \phi] = \beta_{03} + \phi_{31}(65) + \phi_{32}(75)$
- under MAR, $\beta_{03} = -3.4$, $\phi_{31} = .45$, $\phi_{32} = .65$

So,

$$E[Y_3|Y_1, Y_2, \mu, \Sigma] = -3.4 + .45 \times 65 + .65 \times 82 = 75$$

But if incorrectly assume independence,

$$E[Y_3|Y_1, Y_2, \mu, \Sigma] = \mu_3 = 88$$

A huge difference
Covariance modeling

- recall for $p(y; \theta)$ trivariate normal with mean $\mu$ and covariance matrix $\Sigma$, can rewrite as

$$p(y_1, y_2, y_3) = p(y_1)p(y_2 | y_1)p(y_3 | y_1, y_2)$$

with parameters $\phi$ and $\sigma^*$ where

$$E[Y_2|y_1] = \beta_0 + \phi_{21}y_1$$
$$Var[Y_2|y_1] = \sigma^*_{22}$$
$$E[Y_3|y_2, y_1] = \beta_0 + \phi_{32}y_2 + \phi_{31}y_1$$
$$Var[Y_3|y_2, y_1] = \sigma^*_{33}$$

- model dependence via 'regression' coefficients $\phi$ which are unconstrained; important to do carefully
Case study I: Growth Hormone trial

- primary outcome of interest: mean quadriceps strength at baseline, month 6, and month 12;
- for subject $i$, $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})^T$.
- randomized to one of four treatment groups: Growth Hormone plus Exercise (EG), Growth Hormone (G), Placebo plus Exercise (EP), Placebo (P)
- denote treatment group as $Z_i$, $\{1, 2, 3, 4\}$
- main inferential objective: mean quadriceps strength at 12 months for each of the four treatments.
**GH: data summary**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>$k$</th>
<th>$n_k$</th>
<th>0</th>
<th>6</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EG</strong></td>
<td>1</td>
<td>12</td>
<td>58</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>57</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>22</td>
<td>78</td>
<td>90</td>
<td>88</td>
</tr>
<tr>
<td>All</td>
<td>38</td>
<td></td>
<td>69</td>
<td>87</td>
<td>88</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td>1</td>
<td>6</td>
<td>68</td>
<td>81</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5</td>
<td>77</td>
<td>81</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>30</td>
<td>67</td>
<td>64</td>
<td>63</td>
</tr>
<tr>
<td>All</td>
<td>41</td>
<td></td>
<td>69</td>
<td>66</td>
<td>63</td>
</tr>
<tr>
<td><strong>EP</strong></td>
<td>1</td>
<td>7</td>
<td>65</td>
<td>86</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>87</td>
<td>86</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>31</td>
<td>65</td>
<td>81</td>
<td>73</td>
</tr>
<tr>
<td>All</td>
<td>40</td>
<td></td>
<td>66</td>
<td>82</td>
<td>73</td>
</tr>
<tr>
<td><strong>P</strong></td>
<td>1</td>
<td>8</td>
<td>66</td>
<td>62</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5</td>
<td>53</td>
<td>62</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>28</td>
<td>67</td>
<td>62</td>
<td>63</td>
</tr>
<tr>
<td>All</td>
<td>41</td>
<td></td>
<td>65</td>
<td>62</td>
<td>63</td>
</tr>
</tbody>
</table>
GH models

\[ \mathbf{Y}_i \mid Z_i = k \sim N(\mu_k, \Sigma_k), \]  

- \[ \mu_k = (\mu_{1k}, \mu_{2k}, \mu_{3k})^T \]
- important to correctly model the dependence structure
- differences in the covariance across treatments (see next slide)
- however, need to be careful to ensure \( \Sigma_k \) stays positive definite
## Covariance matrices by tx

<table>
<thead>
<tr>
<th></th>
<th>EG</th>
<th>G</th>
<th></th>
<th>EP</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Covariance matrix" /></td>
<td><img src="image" alt="Covariance matrix" /></td>
<td></td>
<td><img src="image" alt="Covariance matrix" /></td>
<td><img src="image" alt="Covariance matrix" /></td>
</tr>
<tr>
<td></td>
<td>697</td>
<td>680</td>
<td>754</td>
<td>697</td>
<td>680</td>
</tr>
<tr>
<td></td>
<td>0.73</td>
<td>1253</td>
<td>1120</td>
<td>0.73</td>
<td>1253</td>
</tr>
<tr>
<td></td>
<td>0.78</td>
<td>0.87</td>
<td>1335</td>
<td>0.78</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>552</td>
<td>498</td>
<td>449</td>
<td>552</td>
<td>498</td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>631</td>
<td>516</td>
<td>0.84</td>
<td>631</td>
</tr>
<tr>
<td></td>
<td>0.83</td>
<td>0.89</td>
<td>532</td>
<td>0.83</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>622</td>
<td>457</td>
<td>353</td>
<td>622</td>
<td>457</td>
</tr>
<tr>
<td></td>
<td>0.78</td>
<td>546</td>
<td>424</td>
<td>0.78</td>
<td>546</td>
</tr>
<tr>
<td></td>
<td>0.63</td>
<td>0.81</td>
<td>498</td>
<td>0.63</td>
<td>0.81</td>
</tr>
</tbody>
</table>
GH models: re-parameterized

\[ Y_{i1} \mid Z_i = k \sim N(\mu_{1k}, \sigma_{1k}^2) \]
\[ Y_{i2} \mid y_{i1}, Z_i = k \sim N(\beta_{0k} + \phi_{21,k}y_{i1}, \sigma_{2k}^2) \]
\[ Y_{i3} \mid y_{i2}, y_{i1}, Z_i = k \sim N(\beta_{1k} + \phi_{31,k}y_{i1} + \phi_{32,k}y_{i2}, \sigma_{3k}^2) \]

- have implicitly reparameterized the elements of \( \Sigma_k \), as
  - \( \phi_k = (\phi_{21,k}, \phi_{31,k}, \phi_{32,k}) : k = 1, \ldots, 4 \): set of generalized autoregressive parameters (GARP) for each treatment
  - \( \sigma_k^2 = (\sigma_{1k}^2, \sigma_{2k}^2, \sigma_{3k}^2) : k = 1, \ldots, 4 \): the set of innovation variances (IV) for each treatment.
  - \( \mu_{jk} = \beta_{j-2,k} + \sum_{l=1}^{j-1} \phi_{jl} \mu_{lk} \) for \( j = 2, 3 \)
## GARP/IV by tx

<table>
<thead>
<tr>
<th></th>
<th>EG</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>697</td>
<td>552</td>
</tr>
<tr>
<td></td>
<td>0.98 563</td>
<td>0.90 176</td>
</tr>
<tr>
<td></td>
<td>0.45 0.65 241</td>
<td>0.26 0.61 91</td>
</tr>
<tr>
<td></td>
<td>EP</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>741</td>
<td>622</td>
</tr>
<tr>
<td></td>
<td>0.89 199</td>
<td>0.74 203</td>
</tr>
<tr>
<td></td>
<td>0.21 0.59 82</td>
<td>−0.01 0.78 154</td>
</tr>
</tbody>
</table>
GH models: covariance models

- Model (1): \((\phi_k, \sigma^2_k)\) distinct and unstructured for each treatment \((k)\),
- Model (2) \(\{\phi_k, \sigma^2_k\} = (\phi, \sigma^2): k = 1, \ldots, 4\) common \(\Sigma\) across treatments
- Model (3),

\[
\begin{align*}
(\sigma^2_{1k}, \phi_{31,k}, \phi_{32,k}) &= (\sigma^2_1, \phi_{31}, \phi_{32}) & k = 1, \ldots, 4 \\
(\sigma^2_{2k}, \sigma^2_{3k}) &= (\sigma^2_2, \sigma^2_3) & k = 2, 3, 4 \\
\phi_{21,k} &= \phi_{21} & k = 1, 2, 3
\end{align*}
\]
GH models: priors

- $\beta, \phi$: regression coefficients (for mean and dependence respectively)
- $\sigma^2$: variances
- diffuse priors

\[
\begin{align*}
\beta_{jk} & \sim N(0, 10^6 I_3), \quad j = 0, 1; \quad k = 1, \ldots, 4 \\
\phi_{jl,k} & \sim N(0, 10^6 I_3), \quad j = 2, 3; \quad l = 1, \ldots, j - 1; \quad k = 1, \ldots, 4 \\
\sigma_{jk} & \sim U(0, 100), \quad j = 1, \ldots, 3; \quad k = 1, \ldots, 4
\end{align*}
\]
Model
{
    # model for observable responses
    for (i in 1:N1) # observe on baseline
    {
        y[i,1]~dnorm(mu[trt[i]],tau1)
    }
    for (i in (N1+1):N2) # observe baseline and 6 months
    {
        y[i,1]~dnorm(mu[trt[i]],tau1)
        y[i,2]~dnorm(conmean1[i,trt[i]], tau2[trt1[i]])
        conmean1[i,trt[i]]<-beta0[trt[i]]+phi21[trt2[i]]*y[i,1]
    }
    for (i in (N2+1):N) # completers
    {
        y[i,1]~dnorm(mu[trt[i]],tau1)
        y[i,2]~dnorm(conmean2[i,trt[i]], tau2[trt1[i]])
        y[i,3]~dnorm(conmean3[i,trt[i]], tau3[trt1[i]])
        conmean2[i,trt[i]]<-beta0[trt[i]]+phi21[trt2[i]]*y[i,1]
        conmean3[i,trt[i]]<-beta1[trt[i]]+phi31*y[i,1]+phi32*y[i,2]
    }
}
# Calculation of marginal means mu2 and mu3 for each treatment

# Tx=1

# Tx=2

# Tx=3

# Tx=4
#priors on dependence parameters
phi21[1]~dnorm(0, 0.0001)
phi21[2]~dnorm(0, 0.0001)
phi31~dnorm(0, 0.0001)
phi32~dnorm(0, 0.0001)

#priors on mean parameters
for (k in 1:ntrt)
{
    mu[k]~dnorm(0, 0.00001)
beta0[k]~dnorm(0, 0.0001)
beta1[k]~dnorm(0, 0.0001)
}

# priors on the square root of innovation variances
sigma1~dunif(0,100)
sigma2[1]~dunif(0,100)
sigma2[2]~dunif(0,100)
sigma3[1]~dunif(0,100)
sigma3[2]~dunif(0,100)
Priors: WinBUGS II

# convert to precisions
tau1<-pow(sigma1,-2)
tau2[1]<-pow(sigma2[1],-2)
tau2[2]<-pow(sigma2[2],-2)
tau3[1]<-pow(sigma3[1],-2)
tau3[2]<-pow(sigma3[2],-2)
GH: Model Comparison: DIC

- based on **observed data likelihood**
- easy to construct based on factorization

\[ p(y_1, y_2, y_3) = p(y_1)p(y_2 | y_1)p(y_3 | y_2, y_1) \]

\[ \theta = (\mu, \phi, 1/\sigma^2) \]

<table>
<thead>
<tr>
<th>model</th>
<th>DIC</th>
<th>Dev((\theta))</th>
<th>(p_D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>2674</td>
<td>2599</td>
<td>37</td>
</tr>
<tr>
<td>(2)</td>
<td>2667</td>
<td>2631</td>
<td>18</td>
</tr>
<tr>
<td>(3)</td>
<td>2652</td>
<td>2609</td>
<td>21</td>
</tr>
</tbody>
</table>

- DIC says model (3) fits best
- for this model and parameterization, default DIC from WinBUGS is this DIC with this parameterization
### GH: Results

#### Cov Model (3)

<table>
<thead>
<tr>
<th>Time</th>
<th>EG</th>
<th>6</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>69 (62,77)</td>
<td>82 (71,94)</td>
<td>81 (70,92)</td>
</tr>
<tr>
<td>6</td>
<td>66 (58,74)</td>
<td>81 (73,90)</td>
<td>73 (65,80)</td>
</tr>
</tbody>
</table>

#### Cov Model (1)

<table>
<thead>
<tr>
<th>Month</th>
<th>EG</th>
<th>6</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>69 (4.2)</td>
<td>81 (6.0)</td>
<td>78 (6.3)</td>
</tr>
<tr>
<td>6</td>
<td>65 (4.2)</td>
<td>81 (4.4)</td>
<td>73 (3.7)</td>
</tr>
</tbody>
</table>

#### Completers only, Cov Model (1)

<table>
<thead>
<tr>
<th>Time</th>
<th>EG</th>
<th>6</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>78 (67,89)</td>
<td>90 (76,105)</td>
<td>88 (74,103)</td>
</tr>
</tbody>
</table>
GH: Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(3)</td>
<td>0.99</td>
<td>0.90</td>
<td>1.00</td>
<td>0.080</td>
<td>0.68</td>
<td>0.97</td>
</tr>
<tr>
<td>indep</td>
<td>1.00</td>
<td>0.98</td>
<td>1.00</td>
<td>0.068</td>
<td>0.52</td>
<td>0.94</td>
</tr>
</tbody>
</table>

- also differences between dependence models: for EG vs. EP, posterior probs were 0.90 and 0.80 for (1) and (2)
GH: conclusions

- the covariance model that allowed the individual GARP/IV parameters to vary by treatment, covariance model (3) was the preferred model for this analysis (as measured by the DIC)
- the EG arm had the greatest increases in quad strength at month 12
- EP had a significantly larger increase than P
- under assumption of ignorability, did not need to explicitly specify the MDM
Summary

- reviewed key features of ignorable models for incomplete data
- requires specification of only the full data response model
- for longitudinal data, care must be taken in modeling the dependence
PART 3: Analysis under Missing Not at Random (MNAR)

Short Course, Colombia, July 2011
1. Notation

2. Objectives of analysis

3. Modeling approaches
   - Extrapolation factorization
   - Sensitivity parameters

4. Modeling principles

5. Selection models

6. Summary
Notation I

\[ \mathbf{Y} = (Y_1, \ldots, Y_J)^T \]
\[ = \text{full-data vector of responses} \]

\[ \mathbf{R} = (R_1, \ldots, R_J)^T \]
\[ = \text{vector of observation indicators} \]

\[ R_j = \mathbf{1}\{Y_j \text{ observed}\} \]
\[ Y_{\text{obs}} = \{ Y_j : R_j = 1 \} = \text{observed responses} \]

\[ Y_{\text{mis}} = \{ Y_j : R_j = 0 \} = \text{missing responses} \]
Objectives of analysis

Full-data distribution of $\mathbf{Y}$ is

$$F(y \mid \theta)$$

Usual objective: infer some functional of $\theta$

This requires modeling the joint distribution

$$f(y, r \mid \omega)$$

where $\theta = g(\omega)$
Modeling approaches

1. Selection models

\[ p(y, r \mid \omega) = p(y \mid \theta) \, p(r \mid y, \psi) \]

2. Mixture models

\[ p(y, r \mid \omega) = p(y \mid r, \alpha) \, p(r \mid \phi) \]

3. Other approaches ...

More on each approach coming up ....
Extrapolation factorization

- Every full-data model uses an extrapolation

$$p(y, r \mid \omega) = p(y_{\text{obs}}, r \mid \omega_O) \times p(y_{\text{mis}} \mid y_{\text{obs}}, r, \omega_E)$$

= observed data dist’n × missing data extrapolation

where \( \omega = g(\omega_E, \omega_O) \).

- Under MAR:

$$p(y_{\text{mis}} \mid y_{\text{obs}}, r) = p(y_{\text{mis}} \mid y_{\text{obs}})$$

In words: extrapolation of missing data from observed data does not depend on missing data pattern.
Recall extrapolation factorization

\[ p(y, r | \omega) = p(y_{\text{obs}}, r | \omega_0) p(y_{\text{mis}} | y_{\text{obs}}, r, \omega_E) \]

**IF** there exists a reparameterization \( \xi(\omega) = (\xi_S, \xi_M) \) such that:

1. \( \xi_S \) is a nonconstant function of \( \omega_E \),
2. the observed-data likelihood
   \[ L(\xi_S, \xi_M | y_{\text{obs}}, r) \]
   is constant as a function of \( \xi_S \), and
3. at a fixed value of \( \xi_S \), the observed data likelihood is a nonconstant function of \( \xi_M \),

**THEN** we call \( \xi_S \) a **sensitivity parameter**.
Sensitivity parameters (in words)

1. $\xi_S$ indexes the extrapolation distribution.

2. The model fit is equivalent for every value of $\xi_S$.

3. When $\xi_S$ is fixed, the full data model is identified.
Specifying priors

- Define

\[ \xi_S = h(\xi_M, \Delta), \]

where \( \Delta \) captures the information about the missing data mechanism.

- To anchor the model at MAR, there must exist some \( \Delta_0 \) such that \( h(\xi_S, \Delta_0) \) implies MAR.
Specifying priors

- can specify $p(\xi_S, \xi_M)$ as

$$p(\xi_S, \xi_M) = p(\xi_S \mid \xi_M) p(\xi_M),$$  \hspace{1cm} (1)

where $p(\xi_S \mid \xi_M)$ captures assumptions about the missing data mechanism.
In terms of the sensitivity parameter $\Delta$, we can re-express $p(\xi_S \mid \xi_M)$ as

$$p(\xi_S \mid \xi_M) = p(\xi_S \mid \xi_M, \Delta) \ p(\Delta \mid \xi_M),$$

where

$$p(\xi_S \mid \xi_M, \Delta) = I\{\xi_S = h(\xi_M, \Delta)\}$$

is a point mass prior; that is, $\xi_S$ is a deterministic function of $(\xi_M, \Delta)$ as above.
Specifying priors

\[ p(\xi_S | \xi_M) = p(\xi_S | \xi_M, \Delta) p(\Delta | \xi_M), \]  \hspace{1cm} (3)

- The first part of the prior in (3) simply reparameterizes \( \xi_S \) in terms of departures from MAR
- The second part, \( p(\Delta | \xi_M) \), encodes the missing data mechanism \textit{and} the uncertainty about it.
Specifying priors

- A point mass prior

\[ p(\Delta \mid \xi_M) = I\{\Delta = \Delta^*\} \]

for a fixed value \(\Delta^*\) encodes a particular missing data mechanism such as MAR with absolute certainty.

- Priors that convey uncertainty about the missing data mechanism can also be used.

\[ \Delta \mid \xi_M \sim N(d, D), \]

- Center the missing data mechanism at \(\Delta = d\),
- Prior uncertainty quantified through a variance matrix \(D\).
Specifying priors

- assume $\Delta = \Delta_0$ implies MAR

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Condition 1</th>
<th>Condition 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAR (no uncertainty)</td>
<td>$E(\Delta</td>
<td>\xi_M) = \Delta_0$</td>
</tr>
<tr>
<td>MAR (uncertainty)</td>
<td>$E(\Delta</td>
<td>\xi_M) = \Delta_0$</td>
</tr>
<tr>
<td>MNAR (no uncertainty)</td>
<td>$E(\Delta</td>
<td>\xi_M) = \Delta^* \neq \Delta_0$</td>
</tr>
<tr>
<td>MNAR (uncertainty)</td>
<td>$E(\Delta</td>
<td>\xi_M) = \Delta^* \neq \Delta_0$</td>
</tr>
</tbody>
</table>
The pattern mixture model factorization makes them a
natural class of models for partitioning \( \omega \)
\[ p(y, r \mid \omega) = p(y \mid r, \alpha(\omega)) \ p(r \mid \phi(\omega)). \] (4)

can further factor the first term
\[
 p(y, r \mid \alpha, \phi) = p(y \mid r, \alpha) \ p(r \mid \phi) \\
= p(y_{\text{mis}} \mid y_{\text{obs}}, r, \alpha) \ p(y_{\text{obs}} \mid r, \alpha) \ p(r \mid \phi).
\]

the extrapolation model depends only on parameters \( \alpha \)
indexing the mixture components \( p(y \mid r, \alpha) \).

so the reparameterization of \( \omega = (\alpha, \phi) \) as \( \xi(\omega) = (\xi_S, \xi_M) \)
is relatively straightforward in many practical settings
(example next)
Mixture of bivariate normals

- bivariate response $\mathbf{Y}_i = (Y_{i1}, Y_{i2})^T$
- missingness in $Y_{i2}$, denoted by $R_i = 0$
- pattern mixture formulation

$$\mathbf{Y}_i \mid R_i = r \sim N(\mu^{(r)}, \Sigma^{(r)})$$
it is convenient to re-parametrize the model as

\[ Y_{i2} \mid y_{i1}, R_i = r \sim N(B_0^{(r)} + B_1^{(r)} y_{i1}, \sigma_{2|1}^{(r)}) \]
\[ Y_{i1} \mid R_i = r \sim N(\mu_1^{(r)}, \sigma_{11}^{(r)}) \]

where

\[ \alpha = \{ \beta_0^{(r)}, \beta_1^{(r)}, \sigma_{2|1}^{(r)}, \mu_1^{(r)}, \sigma_{11}^{(r)} : r = 0, 1 \} . \]

define \( \xi(\omega) = \xi(\alpha, \phi) = (\xi_S, \xi_M) \) such that

\[ \xi_S = (\beta_0^{(0)}, \beta_1^{(0)}, \sigma_{2|1}^{(0)}) , \]
\[ \xi_M = (\beta_0^{(1)}, \beta_1^{(1)}, \sigma_{2|1}^{(1)}, \mu_1^{(0)}, \mu_1^{(1)}, \sigma_{11}^{(1)}, \sigma_{11}^{(0)}, \phi) . \]

note that \( \xi_S \) is exclusively a function of \( \alpha \).
Bivariate mixture of normals 1

- Define the function $h(\xi_M, \Delta)$ as

$$
\begin{align*}
    h_1(\xi_M, \Delta) &= \beta_0^{(1)} + \Delta_0 \\
    h_2(\xi_M, \Delta) &= \beta_1^{(1)} + \Delta_1 \\
    h_3(\xi_M, \Delta) &= \Delta_2 \sigma_{2|1}^{(1)}.
\end{align*}
$$

- Now set $\xi_S = h(\xi_M, \Delta)$

$$
\begin{align*}
    \beta_0^{(0)} &= \beta_0^{(1)} + \Delta_0 \\
    \beta_1^{(0)} &= \beta_1^{(1)} + \Delta_1 \\
    \sigma_{2|1}^{(0)} &= \Delta_2 \sigma_{2|1}^{(1)}.
\end{align*}
$$
Bivariate mixture of normals II

- $(\Delta_0, \Delta_1, \Delta_2) = (0, 0, 1)$ yields MAR; non-degenerate priors on $\Delta$ to encode alternate missing data assumptions.
Bivariate mixture: Common variance

- In above example, 3 sensitivity parameters
- Can reduce dimension via modeling constraints
- Constrain $\Sigma^{(r)} = \Sigma : r = 0, 1$
- Then

$$
\beta_1^{(1)} = \beta_1^{(0)} = \beta_1 \\
\sigma_{2|1}^{(1)} = \sigma_{2|1}^{(0)} = \sigma_{2|1} \\
\sigma_{11}^{(1)} = \sigma_{11}^{(0)} = \sigma_{11}.
$$

- So only one sensitivity parameter

$$
\xi_S = \beta_0^{(0)} \\
\xi_M = (\beta_0^{(1)}, \beta_1, \sigma_{2|1}, \sigma_{11}, \phi).
$$
To center the model at MAR, let $\xi_S = \beta_0^{(0)} = h(\xi_M, \Delta)$, where

$$h(\xi_M, \Delta) = \beta_0^{(1)} + \Delta.$$

and $\Delta = 0$ is MAR
Bivariate mixture: Common variance

- Specify prior as follows

\[ p(\xi_S, \xi_M, \Delta) = p(\xi_S | \xi_M, \Delta) p(\Delta | \xi_M) p(\xi_M). \]

- For the first factor

\[ \beta_0^{(0)} | \Delta, \xi_M \sim I\{\beta_0^{(0)} = \beta_0^{(1)} + \Delta\}. \]
Bivariate mixture: Common variance

\[ p(\xi_S, \xi_M, \Delta) = p(\xi_S | \xi_M, \Delta) p(\Delta | \xi_M) p(\xi_M). \]

- can calibrate the informative prior for \( \Delta \) given \( \xi_M \) (second factor) using identified model parameters; for example,

\[ \Delta | \xi_M \sim N(d, \sigma^2_{2|1}). \]

- If \( d = 0 \), the prior is centered at MAR.
- Variance \( \sigma^2_{2|1} = \text{var}(Y_2 | Y_1) \) is chosen because the sensitivity parameter \( \xi_S = \beta_0^{(0)} \) characterizes the conditional mean \( E(Y_2 | Y_1, R = 0) \); example of this in case studies.
Modeling principles I

1. Represent missing data assumptions that cannot be verified by data in terms of unidentified (sensitivity) parameters.

- Assumptions about missing data mechanism are fully encoded by parameters that are not identified by observed data.

- Can move smoothly through the space of full-data models with MNAR missingness by varying the value of sensitivity parameters.
Modeling principles II

- Changing the value of the sensitivity parameter does not affect observed data likelihood
  - (All models being considered have equivalent fit to observed data as measure by DIC or PPL.)
  - Hence can check fit of observed-data part of the model

- Center the model at MAR
Modeling principles III

2. Formalize certainty or uncertainty about missing data mechanism using prior distributions for sensitivity parameters.

- Can use informative prior if expert opinion or historical information is available
- Can use vague or flat (but proper) prior otherwise
- Properly reflects uncertainty about missing data mechanism in final inferences
Modeling principles IV

3 Sensitivity Analysis vs. Informative Priors

**Sensitivity analysis**
- Can be viewed as repeating the analysis across a series of point mass priors.
- Individual analyses do not reflect uncertainty about missing data mechanism.

**Informative prior**
- Gives a summary inference that incorporates prior beliefs about missing data mechanism.
- Requires clear communication about and justification of priors; they are part of the model.
Selection model I

Foregoing used mixture model parameterization. What about selection model?

\[ Y \sim F(y; \theta) \]
\[ R \mid Y = y \sim \text{Ber}(\pi(y)), \]

with, for example,

\[ g(\pi(y)) = \psi_0 + \psi_1 y \]
Selection model II

Properties of this model:

1. It is fully identified when $F(y; \theta)$ is a specified parametric model (e.g., normal)

2. If $F(y; \theta)$ specified, cannot generally parameterize in terms of a sensitivity parameter

3. Can do this (approximately) if $p(y_{\text{obs}})$ or $p(y)$ is nonparametric
Illustration: Assume \( Y \sim N(\mu, \sigma^2) \)

If MDM is

\[
\text{logit}\{\pi(y)\} = \psi_0 + \psi_1 y
\]

then only one value of \((\psi_0, \psi_1)\) fills in the right tail according to a normal distribution.
Parametric selection models

- Can be useful if confident in model specification

- Sensitivity analysis is redundant: one model will fit observed data best; all parameters are identifiable

- May be useful to examine more than one selection mechanism

- Parametric shared parameter models have the same problem
Summary

- Best to focus on classes of models that allow sensitivity analysis and informative priors
  - mixture models – usually work ok
  - selection models – parametric versions usually will not work

- We illustrate these concepts in a case study next
PART 4: Case Studies

Short Course, Colombia, July 2011
Overview

Analysis of Growth Hormone Study

- Mixture of MV normals
- Fit under MAR
- Sensitivity to departures from MAR
- Calibration of MNAR priors

OASIS Study

- Repeated binary data
- Construction and use of elicited priors
Growth Hormone Study: Description

- Longitudinal trial of 4 treatments for increasing muscle strength in elderly
- 160 participants randomized to
  - Growth hormone (G)
  - *Growth hormone plus exercise (EG)*
  - Placebo (P)
  - *Placebo plus exercise (EP)*
- Outcome: Quadriceps strength (QS)
  - recorded at baseline, 3 months, 6 months
  - measured in foot-pounds of torque exerted against resistance device
- Objective: Compare mean QS at one year following randomization
## GH Study: Data summary

<table>
<thead>
<tr>
<th>Treatment</th>
<th>$k$</th>
<th>$n_k$</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>EP</td>
<td>1</td>
<td>7</td>
<td>65 (32)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>87 (52)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>31</td>
<td>65 (24)</td>
</tr>
<tr>
<td>All</td>
<td>40</td>
<td>66 (26)</td>
<td>82 (26)</td>
</tr>
<tr>
<td>EG</td>
<td>1</td>
<td>12</td>
<td>58 (26)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>57 (15)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>22</td>
<td>78 (24)</td>
</tr>
<tr>
<td>All</td>
<td>38</td>
<td>69 (25)</td>
<td>87 (32)</td>
</tr>
</tbody>
</table>
Analysis Plan

Compare two arms:

- Exercise + rhGH (EG)
- Exercise + placebo (EP)

Three analyses using pattern mixture models

- MAR
- MNAR at fixed points (‘sensitivity analysis’)
- Using informative MNAR prior
GH Trial: Full data model

\((Y_1, Y_2, Y_3)^T\) = quad strength measures at months 0, 6, 12

\(Z = \) treatment group indicator
\((0 = \text{GH}, 1 = \text{GH+E})\)

\((R_1, R_2, R_3)^T\) = response indicators
\((1 = \text{observed}, 0 = \text{missing})\)

\(S = \sum_j R_j\)
Pattern Mixture Model: Set up

- $p_s(y_1)$ identified for $s = 1, 2, 3$
- For $p_s(y_2 | y_1)$ and $p_s(y_3 | y_2, y_1)$, have

| $S = 1$ | $p_1(y_2 | y_1)$ | $p_1(y_3 | y_1, y_2)$ |
|---------|------------------|------------------|
| $S = 2$ | $p_2(y_2 | y_1)$ | $p_2(y_3 | y_1, y_2)$ |
| $S = 3$ | $p_3(y_2 | y_1)$ | $p_3(y_3 | y_1, y_2)$ |

- Non-identified components in red.
Model specification for $Y \mid S$ (by TX)

**Model of $Y_1 \mid S$**

\[
(Y_1 \mid S = 1) \sim N(\mu_1, \sigma_1^2) \\
(Y_1 \mid S = 2) \sim N(\mu_2, \sigma_2^2) \\
(Y_1 \mid S = 3) \sim N(\mu_3, \sigma_3^2)
\]
Model specification for $\mathbf{Y} \mid S$ (by TX)

Model for $Y_2 \mid Y_1, S$

$$(Y_2 \mid Y_1, S = 1) \sim N(\alpha_{2.1}^{(0)} + \beta_{2.1} Y_1, \sigma_{2.1}^2)$$

$$(Y_2 \mid Y_1, S \geq 2) \sim N(\alpha_{2.1}^{(1)} + \beta_{2.1} Y_1, \sigma_{2.1}^2)$$
Model specification for $Y \mid S$ (by TX)

Model for $Y_3 \mid Y_1, Y_2, S$

$$(Y_3 \mid Y_1, Y_2, S = 1, 2) \sim N(\alpha_{3.21}^{(0)} + \beta_{3.21} Y_1 + \gamma_{3.21} Y_2, \sigma_{3.21}^2)$$

$$(Y_3 \mid Y_1, Y_2, S = 3) \sim N(\alpha_{3.21}^{(1)} + \beta_{3.21} Y_1 + \gamma_{3.21} Y_2, \sigma_{3.21}^2)$$
Model for $S$ (by TX)

Pattern indicator has multinomial distribution

$$S \sim \text{Mult}(\phi)$$

where $\phi = (\phi_1, \phi_2, \phi_3)$. 
Parameterize Departures from MAR

Model parameters in terms of \((\xi_M, \xi_S)\)

\[
\xi_S = \alpha_{2.1}^{(0)}, \alpha_{3.21}^{(0)}
\]

\[
\xi_M = \left\{ \begin{array}{l}
\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \mu_3, \sigma_3^2, \\
\alpha_{2.1}^{(1)}, \beta_{2.1}, \sigma_{2.1}, \\
\alpha_{3.21}^{(1)}, \beta_{3.21}, \gamma_{3.21}, \sigma_{3.21}^2, \\
\phi_1, \phi_2, \phi_3.
\end{array} \right.
\]
Parameterize Departures from MAR

- Introduce $h$ functions to relate identified, non-identified, and sensitivity parameters:

For $\alpha$ parameters:

\[ \alpha_{2.1}^{(0)} = h_1(\xi_M, \Delta) = \alpha_{2.1}^{(1)} + \Delta_1 \]

\[ \alpha_{3.21}^{(0)} = h_2(\xi_M, \Delta) = \alpha_{3.21}^{(1)} + \Delta_2 \]

MAR induced by setting $\Delta_1 = \Delta_2 = 0$
Interpretation of sensitivity parameters

Separately by treatment,

\[ \Delta_1 = E(Y_2 \mid Y_1, S = 1) - E(Y_2 \mid Y_1, S \geq 2) \]

\[ \Delta_2 = E(Y_3 \mid Y_2, Y_1, S = 1, 2) - E(Y_3 \mid Y_2, Y_1, S = 3) \]
Effect of sensitivity parameter on means

Full-data mean at time 2 is

\[ E_\Delta(Y_2) = \sum_{s=1}^{3} \phi_s E_\Delta(Y_2 \mid S = s). \]

Effect of departures from MAR quantified by

\[ E_\Delta(Y_2) - E_0(Y_2) = \phi_1 \Delta \]

Hence the contribution of the sensitivity parameter to the shift in \( \mu_2 \) is proportional to the fraction dropping out at \( S = 1 \).
Effect of sensitivity parameter on means

Similarly for time 3,

\[ E_\Delta(Y_3) - E_0(Y_3) = \Delta\{(\phi_1 + \phi_2) + \phi_1\beta_{3.21}\}. \]
Constrain space of sensitivity parameters

Can further constrain sensitivity parameter

\[ \Delta_1 = \Delta_2 = \Delta \]

We allow \( \Delta \) to have different values across treatment
Priors: General form

Recall:

\[ \xi_M = \text{identified model parameters} \]
\[ \xi_S = \text{non-identified model parameters} \]
\[ \Delta = \text{parameters indicating departure from MAR} \]
\[ \xi_S = h(\xi_M, \Delta) \]

General form of prior:

\[ p(\xi_S, \xi_M, \Delta) = p(\xi_S | \xi_M, \Delta) p(\Delta | \xi_M) p(\xi_M) \]

Prior for \( \xi_S \) takes form:

\[ p(\xi_S | \xi_M, \Delta) = I\{\xi_S = h(\xi_M, \Delta)\} \]
MAR and MNAR Priors

MAR prior: point mass at MAR values for $\Delta$

$$p(\Delta \mid \xi_M) = I\{\Delta = 0\}$$

Departures from MAR:

- Allow $p(\Delta \mid \xi_M)$ to take general form

- Vary $\Delta$ over a fixed range $\mathcal{D}$

$$\{p(\Delta \mid \xi_M) = I\{\Delta = \Delta^*\} : \Delta^* \in \mathcal{D}\}$$
Next few slides

1. MNAR analysis over fixed points in $\mathcal{D}$
2. Advice about how to specify $\mathcal{D}$
3. MNAR analysis over more general priors
4. Compare to MAR analysis
Sensitivity analysis

- Specify a range $D$ for $\Delta$, and repeat inferences for $\Delta \in D$.

- Relies on contextual interpretation

- Example: $D = [-20, 0]$
  - Conditional mean for dropouts at $t$ is 0 to 20 points lower than for non-dropouts

- Equivalent to repeating analysis over a set of point-mass priors for $\Delta$

- Yields contour plots that can be useful
Calibrate prior for $\Delta$ 

Priors for this analysis

$\alpha^{(0)} \mid \alpha^{(1)}, \beta, \gamma, \sigma, \phi, \Delta \sim I(\alpha^{(0)} = \alpha^{(1)} + \Delta)$

$\Delta \mid \alpha^{(1)}, \beta, \gamma, \sigma, \phi \sim \text{Unif}(-\sigma, 0)$

$(\alpha^{(1)}, \beta, \gamma, \sigma, \phi) \sim \text{vague priors}$

Calibrates range of $\Delta$ using

$\sigma = \text{SD}(Y_2 \mid Y_1) = \text{SD}(Y_3 \mid Y_2, Y_1)$
Table: Residual SD estimates from OLS regression on observed data

<table>
<thead>
<tr>
<th>Model</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>((Y_2 \mid Y_1, S \geq 2))</td>
<td>EP ((z = 0))</td>
</tr>
<tr>
<td></td>
<td>14</td>
</tr>
<tr>
<td>((Y_3 \mid Y_1, Y_2, S = 3))</td>
<td>8.9</td>
</tr>
</tbody>
</table>
Define $\theta = E(Y_3 \mid Z = 1) - E(Y_3 \mid Z = 0)$

1. Sensitivity analysis: Contours of
   - $E(\theta \mid Y_{\text{obs}}, \Delta)$
   - $P(\theta > 0 \mid Y_{\text{obs}}, \Delta)$

2. Fully Bayesian inference using calibrated priors on $\Delta$
\[ E(\theta | \Delta_EG, \Delta_EP) \]

\[ \Delta_EG \]

\[ \Delta_EP \]

\[ \text{MAR} \]

Short Course, Colombia, July 2011

PART 4: Case Studies
\[ P(\theta > 0 | \Delta_{EG}, \Delta_{EP}) \]
## Comparative Analysis

<table>
<thead>
<tr>
<th>Trt</th>
<th>Month</th>
<th>MVN</th>
<th>Pattern Mixture Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MVN</td>
</tr>
<tr>
<td>EP</td>
<td>0</td>
<td>65 (4.2)</td>
<td>66 (4.6)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>81 (4.4)</td>
<td>82 (4.7)</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>73 (3.7)</td>
<td>73 (4.0)</td>
</tr>
<tr>
<td>EG</td>
<td>0</td>
<td>69 (4.2)</td>
<td>69 (4.4)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>81 (6.0)</td>
<td>81 (6.4)</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>78 (6.3)</td>
<td>78 (6.7)</td>
</tr>
<tr>
<td>Diff. at 12 mos.</td>
<td>5.7 (7.3)</td>
<td>5.4 (7.8)</td>
<td>2.6 (9.3)</td>
</tr>
</tbody>
</table>
Summary: GH Study

- MAR analyses show some evidence that GH increases quad strength
  - Posterior mean (SD) is 5.7 (7.3)

- MNAR analyses with uniform prior on $\Delta$ shows smaller effect
  - Posterior mean (SD) is 2.6 (9.3)

- MNAR assumption has greater effect on EG arm
  - Posterior mean at month 12 changes from 78 to 72
  - Largely due to higher dropout rate
Summary: GH Study

- General PMM has large number of sensitivity parameters
  - Dimension needs to be reduced
  - Use qualitative considerations

- Must decide on a range $\mathcal{D}$ for sensitivity parameters
  - Choice can be subjective
  - Can calibrate using variance (not SE!) from MAR model
  - Want to ensure $Y_{\text{mis}}$ does not get extrapolated beyond reasonable range

- Contour plots are conditional posteriors
  - Alternative is to use informative prior distributions
OASIS Study: Smoking Cessation Trial

**Description**

- Randomized clinical trial of two interventions for smoking cessation
- Enrolled 298 substance abusers, randomized to
  - Standard intervention (ST)
  - Enhanced intervention (ET)
    - increased phone contact
- Assess smoking status at 1, 3, 6, 12 months following randomization

**Objective**

- Compare smoking rates at month 12 using intention to treat
OASIS Study: Missing Data

- Dropout high on both arms
  - 40% on ST
  - 55% on ET

- Question: can these data even be useful??

- Analyses suggest dropout related to prior smoking status at each time
OASIS Study: Analysis Plan

1. Use of informative priors under MNAR
   Basic principle: embed MAR model in larger set of full-data models that assume MNAR
   Priors on unidentified parameters capture departures from MAR

2. Analysis of OASIS data
   Pattern mixture model set up
   Elicitation of priors for distribution of dropouts
   Comparative analysis via selection model

3. Summary and discussion
### OASIS: Summary Statistics

<table>
<thead>
<tr>
<th>Trt</th>
<th>Month</th>
<th>1</th>
<th>3</th>
<th>6</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quit</td>
<td></td>
<td>.18</td>
<td>.09</td>
<td>.11</td>
<td>.11</td>
</tr>
<tr>
<td>ET (n = 149)</td>
<td></td>
<td>.83</td>
<td>.47</td>
<td>.42</td>
<td>.34</td>
</tr>
<tr>
<td>ET Smoking</td>
<td></td>
<td>—</td>
<td>.44</td>
<td>.46</td>
<td>.55</td>
</tr>
<tr>
<td>ET Missing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ET S/(S+Q)</td>
<td></td>
<td>.83</td>
<td>.84</td>
<td>.79</td>
<td>.76</td>
</tr>
<tr>
<td>ET (S+M)/(S+M+Q)</td>
<td></td>
<td>.83</td>
<td>.91</td>
<td>.89</td>
<td>.89</td>
</tr>
<tr>
<td>ST (n = 149)</td>
<td></td>
<td>.15</td>
<td>.09</td>
<td>.10</td>
<td>.07</td>
</tr>
<tr>
<td>ST Quit</td>
<td></td>
<td>.85</td>
<td>.54</td>
<td>.52</td>
<td>.52</td>
</tr>
<tr>
<td>ST Missing</td>
<td></td>
<td>—</td>
<td>.37</td>
<td>.38</td>
<td>.40</td>
</tr>
<tr>
<td>ST S/(S+Q)</td>
<td></td>
<td>.85</td>
<td>.85</td>
<td>.84</td>
<td>.88</td>
</tr>
<tr>
<td>ST (S+M)/(S+M+Q)</td>
<td></td>
<td>.85</td>
<td>.91</td>
<td>.90</td>
<td>.93</td>
</tr>
</tbody>
</table>
OASIS: PM Model Specification

1. Most general model \((s = 1, 2, 3, 4)\)

\[
Y_1 \mid S = s \sim \text{Ber}(\mu^{(s)})
\]

\[
Y_j \mid Y_{j-1}, \ldots, Y_1, S = s \sim \text{Ber}(\phi_j^{(s)})
\]

2. Within pattern, assume first-order dependence

\[
\text{logit}(\mu^{(s)}) = \beta^{(s)}
\]

\[
\text{logit}(\phi_j^{(s)}) = \alpha_j^{(s)} + \gamma_j^{(s)} Y_{j-1}
\]

- Simplified structure relative to saturated model
- Can separate observed data likelihood from extrapolation
Sensitivity parameters

Full data model for $Y_j \mid Y_{j-1}$

$$\text{logit} \left( \phi_j^{(s)} \right) = \alpha_j^{(s)} + \gamma_j^{(s)} Y_{j-1}$$

Represent departures from MAR: for $j > s$,

$$\alpha_j^{(s)} = \alpha_j^{(\geq j)} + \Delta_0$$
$$\gamma_j^{(s)} = \gamma_j^{(\geq j)} + (\Delta_1 - \Delta_0)$$

$\alpha_j^{(\geq j)}$ and $\gamma_j^{(\geq j)}$ are logistic regression parameters for the collapsed patterns

$$(Y_j \mid Y_{j-1}, \mathcal{S} \geq j)$$
Sensitivity parameters

Each $\Delta$ compares odds of smoking between dropouts and non-dropouts, conditioning on $Y_{j-1} = y$

$$\Delta_y = \log \left\{ \frac{\text{odds}(Y_j = 1 \mid Y_{j-1} = y, S < j)}{\text{odds}(Y_j = 1 \mid Y_{j-1} = y, S \geq j)} \right\}.$$

- MAR holds when $\Delta_0 = \Delta_1 = 0$
- $\Delta$’s do not appear in observed data likelihood
- Simplifying assumption: $\Delta$’s are time-independent (not necessary)
### What is identified?

| $S = 1$ | 0 | $E(Y_2 | Y_1 = y)$ | $E(Y_3 | Y_2 = y)$ | $E(Y_4 | Y_3 = y)$ |
|---------|---|----------------|----------------|----------------|
|         | 0 |                |                |                |
|         | 1 |                |                |                |
| $S = 2$ | 0 | $\alpha^{(2)}_2$ |                |                |
|         | 1 | $\alpha^{(2)}_2 + \gamma^{(2)}$ |                |                |
| $S = 3$ | 0 | $\alpha^{(3)}_2$ | $\alpha^{(3)}_3$ |                |
|         | 1 | $\alpha^{(3)}_2 + \gamma^{(3)}$ | $\alpha^{(3)}_3 + \gamma^{(3)}$ |                |
| $S = 4$ | 0 | $\alpha^{(4)}_2$ | $\alpha^{(4)}_3$ | $\alpha^{(4)}_4$ |
|         | 1 | $\alpha^{(4)}_2 + \gamma^{(4)}$ | $\alpha^{(4)}_3 + \gamma^{(4)}$ | $\alpha^{(4)}_4 + \gamma^{(4)}$ |
### Add Sensitivity Parameters

<table>
<thead>
<tr>
<th>$S$</th>
<th>$y$</th>
<th>$E(Y_2 \mid Y_1 = y)$</th>
<th>$E(Y_3 \mid Y_2 = y)$</th>
<th>$E(Y_4 \mid Y_3 = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>$\alpha_2^{(\geq 2)} + \Delta_0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$\alpha_2^{(\geq 2)} + \gamma_2^{(\geq 2)} + \Delta_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$\alpha_2^{(2)}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$\alpha_2^{(2)} + \gamma_2^{(2)}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>$\alpha_2^{(3)}$</td>
<td>$\alpha_3^{(3)}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$\alpha_2^{(3)} + \gamma_2^{(3)}$</td>
<td>$\alpha_3^{(3)} + \gamma_3^{(3)}$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>$\alpha_2^{(4)}$</td>
<td>$\alpha_3^{(4)}$</td>
<td>$\alpha_4^{(4)}$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$\alpha_2^{(4)} + \gamma_2^{(4)}$</td>
<td>$\alpha_3^{(4)} + \gamma_3^{(4)}$</td>
<td>$\alpha_4^{(4)} + \gamma_4^{(4)}$</td>
</tr>
</tbody>
</table>
## Add Sensitivity Parameters

<table>
<thead>
<tr>
<th>$y$</th>
<th>$E(Y_2 \mid Y_1 = y)$</th>
<th>$E(Y_3 \mid Y_2 = y)$</th>
<th>$E(Y_4 \mid Y_3 = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>$\alpha^{(2)}_2 + \Delta_0$</td>
<td>$\alpha^{(3)}_3 + \Delta_0$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\alpha^{(2)}_2 + \gamma^{(2)}_2 + \Delta_1$</td>
<td>$\alpha^{(3)}_3 + \gamma^{(3)}_3 + \Delta_1$</td>
<td></td>
</tr>
<tr>
<td>$S = 2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>$\alpha^{(2)}_2$</td>
<td>$\alpha^{(3)}_3$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\alpha^{(2)}_2 + \gamma^{(2)}_2$</td>
<td>$\alpha^{(3)}_3 + \gamma^{(3)}_3$</td>
<td></td>
</tr>
<tr>
<td>$S = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>$\alpha^{(3)}_2$</td>
<td>$\alpha^{(3)}_3$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\alpha^{(3)}_2 + \gamma^{(3)}_2$</td>
<td>$\alpha^{(3)}_3 + \gamma^{(3)}_3$</td>
<td></td>
</tr>
<tr>
<td>$S = 4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>$\alpha^{(4)}_2$</td>
<td>$\alpha^{(4)}_3$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\alpha^{(4)}_2 + \gamma^{(4)}_2$</td>
<td>$\alpha^{(4)}_3 + \gamma^{(4)}_3$</td>
<td>$\alpha^{(4)}_4 + \gamma^{(4)}_4$</td>
</tr>
</tbody>
</table>
### Add Sensitivity Parameters

<table>
<thead>
<tr>
<th>$y$</th>
<th>$E(Y_2 \mid Y_1 = y)$</th>
<th>$E(Y_3 \mid Y_2 = y)$</th>
<th>$E(Y_4 \mid Y_3 = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S = 1$</td>
<td>0</td>
<td>$\alpha_2^{(\geq 2)} + \Delta_0$</td>
<td>$\alpha_3^{(\geq 3)} + \Delta_0$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$\alpha_2^{(\geq 2)} + \gamma_2^{(\geq 2)} + \Delta_1$</td>
<td>$\alpha_3^{(\geq 3)} + \gamma_3^{(\geq 3)} + \Delta_1$</td>
</tr>
<tr>
<td>$S = 2$</td>
<td>0</td>
<td>$\alpha_2^{(2)}$</td>
<td>$\alpha_3^{(3)}$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$\alpha_2^{(2)} + \gamma_2^{(2)}$</td>
<td>$\alpha_3^{(3)} + \gamma_3^{(3)}$</td>
</tr>
<tr>
<td>$S = 3$</td>
<td>0</td>
<td>$\alpha_2^{(3)}$</td>
<td>$\alpha_3^{(3)}$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$\alpha_2^{(3)} + \gamma_2^{(3)}$</td>
<td>$\alpha_3^{(3)} + \gamma_3^{(3)}$</td>
</tr>
<tr>
<td>$S = 4$</td>
<td>0</td>
<td>$\alpha_2^{(4)}$</td>
<td>$\alpha_3^{(4)}$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$\alpha_2^{(4)} + \gamma_2^{(4)}$</td>
<td>$\alpha_3^{(4)} + \gamma_3^{(4)}$</td>
</tr>
</tbody>
</table>
## Add Sensitivity Parameters

<table>
<thead>
<tr>
<th>S = 1</th>
<th>( y )</th>
<th>( E(Y_2 \mid Y_1 = y) )</th>
<th>( E(Y_3 \mid Y_2 = y) )</th>
<th>( E(Y_4 \mid Y_3 = y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>( \alpha_2^{(\geq 2)} + \Delta_0 )</td>
<td>( \alpha_3^{(\geq 3)} + \Delta_0 )</td>
<td>( \alpha_4^{(4)} + \Delta_0 )</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>( \alpha_2^{(\geq 2)} + \gamma_2^{(\geq 2)} + \Delta_1 )</td>
<td>( \alpha_3^{(\geq 3)} + \gamma_3^{(\geq 3)} + \Delta_1 )</td>
<td>( \alpha_4^{(4)} + \gamma_4^{(4)} + \Delta_1 )</td>
</tr>
<tr>
<td>S = 2</td>
<td>0</td>
<td>( \alpha_2^{(2)} )</td>
<td>( \alpha_3^{(\geq 3)} + \Delta_0 )</td>
<td>( \alpha_4^{(4)} + \Delta_0 )</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>( \alpha_2^{(2)} + \gamma_2^{(2)} )</td>
<td>( \alpha_3^{(\geq 3)} + \gamma_3^{(\geq 3)} + \Delta_1 )</td>
<td></td>
</tr>
<tr>
<td>S = 3</td>
<td>0</td>
<td>( \alpha_2^{(3)} )</td>
<td>( \alpha_3^{(3)} )</td>
<td>( \alpha_4^{(4)} )</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>( \alpha_2^{(3)} + \gamma_2^{(3)} )</td>
<td>( \alpha_3^{(3)} + \gamma_3^{(3)} )</td>
<td></td>
</tr>
<tr>
<td>S = 4</td>
<td>0</td>
<td>( \alpha_2^{(4)} )</td>
<td>( \alpha_3^{(4)} )</td>
<td>( \alpha_4^{(4)} )</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>( \alpha_2^{(4)} + \gamma_2^{(4)} )</td>
<td>( \alpha_3^{(4)} + \gamma_3^{(4)} )</td>
<td>( \alpha_4^{(4)} + \gamma_4^{(4)} )</td>
</tr>
</tbody>
</table>
Formulating Priors

Map non-identified parameters to identified ones

\[ \gamma_j^{(s)} = h_1 j(\xi_M) = \alpha_j^{(\geq j)} + \Delta_0 \]
\[ \theta_j^{(s)} = h_2 j(\xi_M) = \gamma_j^{(\geq j)} + (\Delta_1 - \Delta_0) \]

General form of prior

\[ p(\xi_S, \xi_M, \Delta) = p(\xi_S | \xi_M, \Delta) p(\Delta | \xi_M) p(\xi_M), \]

where, as usual,

\[ p(\xi_S | \xi_M, \Delta) = I\{\xi_S = h(\xi_M, \Delta)\}. \]
Formulating Priors

Missing data assumptions encoded via

\[ p(\Delta_0, \Delta_1 \mid \xi_M) \]

Under MAR,

\[ p(\Delta_0, \Delta_1 \mid \xi_M) = \mathbb{I}\{(\Delta_0, \Delta_1) = (0, 0)\}. \]

Otherwise, formulate other priors
Results under MAR: point mass at $(\Delta_0, \Delta_1) = (0, 0)$

<table>
<thead>
<tr>
<th>Month</th>
<th>All avail</th>
<th>MAR constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.83</td>
<td>.83</td>
</tr>
<tr>
<td>ET</td>
<td>3</td>
<td>.84</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>.79</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>.76</td>
</tr>
<tr>
<td>1</td>
<td>.85</td>
<td>.85</td>
</tr>
<tr>
<td>ST</td>
<td>3</td>
<td>.85</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>.84</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>.88</td>
</tr>
<tr>
<td>OR</td>
<td>12</td>
<td>.51 (0.19, 1.06)</td>
</tr>
</tbody>
</table>
Eliciting Priors

1. Convened panel of 4 experts in smoking cessation
2. Elicited *best guess* and *90% interval* for probability of smoking at \( j + 1 \) among those with \( j \) measurements,

\[
\begin{align*}
\Pr(Y_{j+1} = 1 \mid Y_j = 1, S = j) \\
\Pr(Y_{j+1} = 1 \mid Y_j = 0, S = j)
\end{align*}
\]

3. Using observed smoking rates among non-dropouts, converted best guess and interval to log odds scale separately for each time point, then averaged over time.

\[ \Rightarrow \text{for each expert, gives best guess and 90\% interval for both } \Delta_0 \text{ and } \Delta_1 \]
Eliciting MNAR Priors

4 Matched each interval to a *skew normal* distribution for $\Delta_0, \Delta_1$

5 Prior on each $\Delta$ is a 4-component mixture of these skew-normal distributions
TABLE: Encoding prior beliefs about smoking rate among dropouts:

Among those who were smoking (abstinent) at occasion t, what is your best guess and 90% range for smoking rate among dropouts at occasion t+1?

<table>
<thead>
<tr>
<th>STEP 1</th>
<th>STEP 2</th>
<th>STEP 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consider all individuals at time t. Divide them into ...</strong></td>
<td><strong>Now suppose that among those who continue to the next measurement occasion, the smoking rate is</strong></td>
<td><strong>Considering the smoking rate among those who continue, provide information about the smoking rate among those who dropped out. Provide:</strong></td>
</tr>
</tbody>
</table>

- **a best guess** (single number percentage such as 75%)
- **a range** that you think would have a 90 percent chance of capturing the true smoking rate (e.g. 60% to 95%)

<table>
<thead>
<tr>
<th></th>
<th>Best Guess</th>
<th>90 Percent Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smokers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abstainers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE: Encoding prior beliefs about smoking rate among dropouts:

Among those who were smoking (abstinent) at occasion t, what is your best guess and 90% range for smoking rate among dropouts at occasion t+1?

<table>
<thead>
<tr>
<th><strong>STEP 1</strong></th>
<th><strong>STEP 2</strong></th>
<th><strong>STEP 3</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consider all individuals at time t. Divide them into ...</strong></td>
<td><strong>Now suppose that among those who continue to the next measurement occasion, the smoking rate is</strong></td>
<td><strong>Considering the smoking rate among those who continue, provide information about the smoking rate among those who dropped out. Provide:</strong></td>
</tr>
<tr>
<td></td>
<td>Best Guess</td>
<td>90 Percent Range</td>
</tr>
<tr>
<td>Smokers</td>
<td>80%</td>
<td>95%</td>
</tr>
<tr>
<td></td>
<td>85%</td>
<td>95%</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>100%</td>
</tr>
<tr>
<td>Abstainers</td>
<td>30%</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>40%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>60%</td>
<td>80%</td>
</tr>
</tbody>
</table>
Figure 4: Comparison of posterior distributions between $E(Y_t | Y_{t-1} = 0, U = 1)$ and $E(Y_t | Y_{t-1} = 0, U \geq t)$ by different priors on $\delta_0$ (solid line: $U = 1$, dot line: $U \geq t$, and first column: $E(Y_2 | Y_1 = 0)$, middle column: $E(Y_3 | Y_2 = 0)$, last column: $E(Y_4 | Y_3 = 0)$): Enhancements arm

$U=1$ vs. $U \geq 2$

$U=1$ vs. $U \geq 3$

$U=1$ vs. $U = 4$
Figure 5: Comparison of posterior distributions between $E(Y_t \mid Y_{t-1} = 0, U = 1)$ and $E(Y_t \mid Y_{t-1} = 0, U \geq t)$ by different priors on $\delta_0$ (solid line: $U = 1$, dot line: $U \geq t$, and first column: $E(Y_2 \mid Y_1 = 0)$, middle column: $E(Y_3 \mid Y_2 = 0)$, last column: $E(Y_4 \mid Y_3 = 0)$):
Results

Posterior mean and 95% CI of treatment OR at week 12

MAR: .51 (.2, 1.1)

MNAR: .65 (.2, 1.4)
Analyses of incomplete data require untestable assumptions

Proposal here: encode them using prior distributions

Transparency as it relates to:
1. Lack of information for inference
2. Source of information for assumptions about missing data
3. Uncertainty about those assumptions

Mixture models are convenient platform

Easier for clinical trials, harder for complex models
Email: mdaniels@stat.ufl.edu


Website: Datasets, etc.