A Comparison of Methods for Longitudinal Data with Missing Due to Truncation

Rong Liu

Merck Research Laboratories

Viswanathan Ramakrishnan

Virginia Commonwealth University

November 8, 2006, Savannah GA

Outline

- Introduction
- Implementation of EM algorithm for MDT method
- Comparison using simulations
- MDT method in conjunction with Rubin’s multiple imputation
- Summary and extensions

Introduction

Let $\mathbf{Y}$ be an N x p data matrix from a p-dimensional multivariate probability distribution $p(\mathbf{Y} \mid \boldsymbol{\theta})$ with parameter $\boldsymbol{\theta}$. Let $\mathbf{I}$ denote the N x p missing indicator matrix from a probability distribution $p(\mathbf{I} \mid \xi, \mathbf{Y})$ with parameter ξ, where

$$I_{\mathrm{i,j}} = \begin{cases} 1 & \text{if } y_{\mathrm{i,j}} \text{ is missing} \\ 0 & \text{if } y_{\mathrm{i,j}} \text{ is observed} \end{cases}.$$

If $\mathbf{Y}$ is not fully observed, denote the observed portion of $\mathbf{Y}$ by $\mathbf{Y}_{obs}$ and the missing portion by $\mathbf{Y}_{mis}$. Then the joint probability of $\mathbf{Y}$ and $\mathbf{I}$ could be written

$$\begin{aligned} p(\mathbf{Y}, \mathbf{I} \mid \boldsymbol{\theta}, \xi) &= p(\mathbf{Y} \mid \boldsymbol{\theta})\, p(\mathbf{I} \mid \mathbf{Y}, \xi) \\ &= p((\mathbf{Y}_{obs}, \mathbf{Y}_{mis}) \mid \boldsymbol{\theta})\, p(\mathbf{I} \mid (\mathbf{Y}_{obs}, \mathbf{Y}_{mis}), \xi) \end{aligned}$$

- Missing completely at random (MCAR): $p(\mathbf{I} \mid \mathbf{Y}, \xi) = \xi$
- Missing at random (MAR): $p(\mathbf{I} \mid \mathbf{Y}, \xi) = p(\mathbf{I} \mid \mathbf{Y}_{obs}, \xi)$
- Not Missing at random (NMAR): $p(\mathbf{I} \mid \mathbf{Y}, \xi) \neq p(\mathbf{I} \mid \mathbf{Y}_{obs}, \xi)$

Introduction (con't)

- Ignorability: MAR and $p(\boldsymbol{\theta},\xi)=p(\boldsymbol{\theta})p(\xi)$

- Likelihood-based inference with incomplete data
  - Under nonignorability: $L(\boldsymbol{\theta},\xi\mid\mathbf{Y}_{obs},\mathbf{I})\propto p(\mathbf{Y}_{obs},\mathbf{I}\mid\boldsymbol{\theta},\xi)$
  - Under ignorability: $L(\boldsymbol{\theta},\xi\mid\mathbf{Y}_{obs},\mathbf{I})\propto p(\mathbf{Y}_{obs}\mid\boldsymbol{\theta})$

$$
\begin{aligned}
p(\mathbf{Y}_{obs},\mathbf{I}\mid\boldsymbol{\theta},\xi) &= \int p((\mathbf{Y}_{obs},\mathbf{Y}_{mis}),\mathbf{I}\mid\boldsymbol{\theta},\xi)d\mathbf{Y}_{mis}\\
&= \int p(\mathbf{I}\mid\mathbf{Y}_{obs},\xi)\,p((\mathbf{Y}_{obs},\mathbf{Y}_{mis})\mid\boldsymbol{\theta})d\mathbf{Y}_{mis}\\
&= p(\mathbf{I}\mid\mathbf{Y}_{obs},\xi)\int p((\mathbf{Y}_{obs},\mathbf{Y}_{mis})\mid\boldsymbol{\theta})d\mathbf{Y}_{mis}\\
&= p(\mathbf{I}\mid\mathbf{Y}_{obs},\xi)\,p(\mathbf{Y}_{obs}\mid\boldsymbol{\theta}).
\end{aligned}
$$

Introduction (con't)
An Example

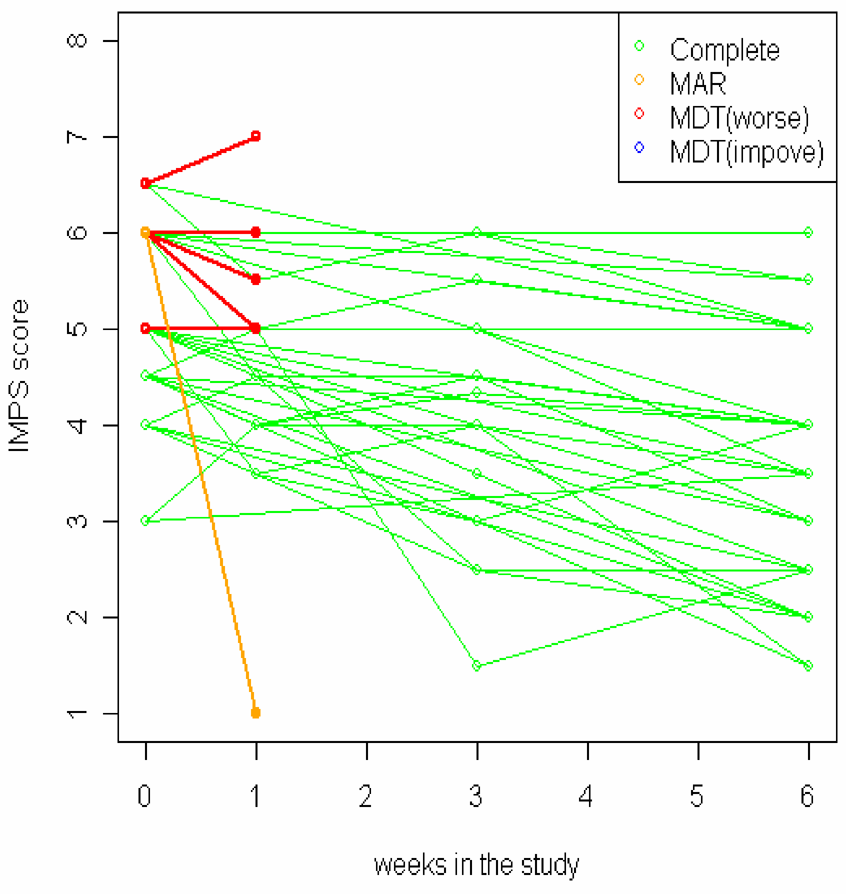
Table 1.2 IMPS (Inpatient Multidimensional Psychiatry Score) data summary

	Treatment (N=71)		Placebo (N=24)		Total (N=95)	
Completion Status	n	%	n	%	n	%
Complete	**61**	**86**	**18**	**75**	**79**	**83**
Incomplete	**10**	**14**	**6**	**25**	**16**	**17**
Worsening	2	20	5	83	7	44
Improved	1	10	0	0	1	6
MAR	7	70	1	17	8	50

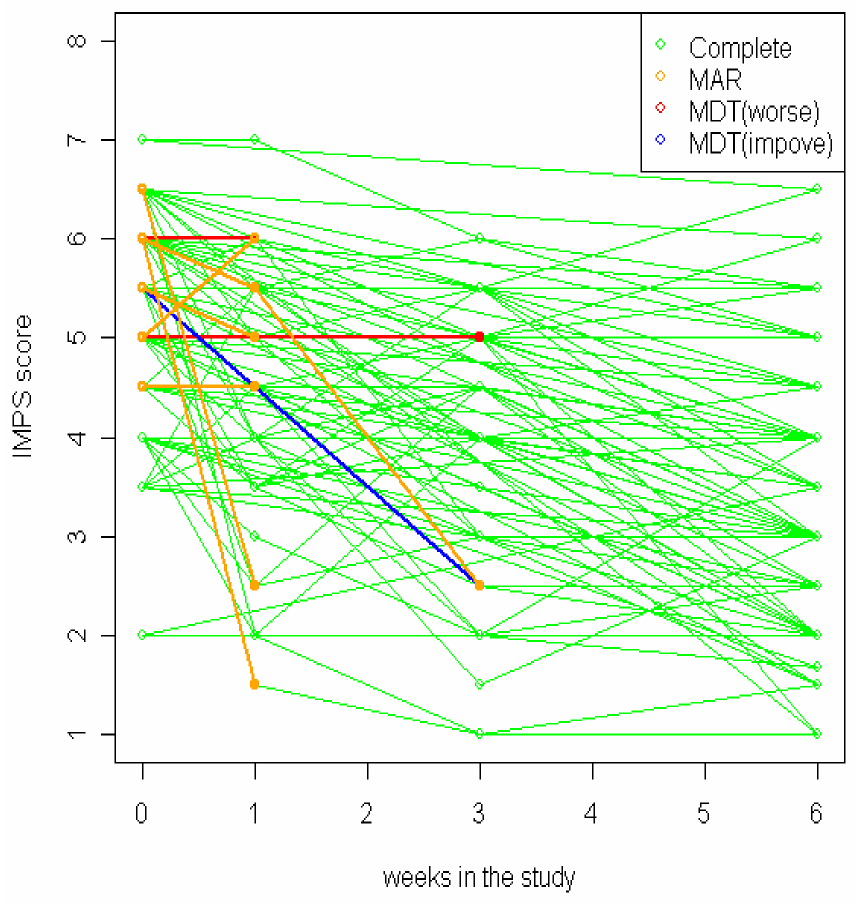
Introduction (con't)

IMPS Data

Placebo group



Treatment group



Introduction (con’t)
MDT Method

- Data matrix: missing occur at the upper end of the distribution and only at the last observed time point *T*.

$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1T-1} & . \\ y_{21} & y_{22} & \cdots & y_{2T-1} & . \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{r_T 1} & y_{r_T 2} & \cdots & y_{r_T T-1} & . \\ y_{(r_T+1)1} & y_{(r_T+1)2} & \cdots & y_{(r_T+1)T-1} & y^*_{(r_T+1)T} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nT-1} & y^*_{nT} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{r_T} \\ \mathbf{y}^*_{r_T+1} \\ \vdots \\ \mathbf{y}^*_n \end{bmatrix}$$

- The *T* - 1 dimensional vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{r_T}$ are independent identically distributed multivariate variables.
- The *T* dimensional vectors $\mathbf{y}^*_{r_T+1}, \ldots, \mathbf{y}^*_n$ are independent identically distributed multivariate variables, where the distribution of the *T*th observations on the *n* individuals are considered to be truncated.

Ramakrishnan, Wang, 2005, Analysis of Data from Clinical Trials with Treatment Related Dropouts, Communication in Statistics

Introduction (con't)
MDT Method

- At time *t*, let the observation $y_n, i = 1, 2, \ldots, n$ represent a sample from a population with some specified pdf.

r_T : Number of cases MDT at last time point.

$\mu_{T/i}$ and $\sigma^2_{T/i}$: Mean and variance of $y_{i,T}$ conditioning on previous *T*-1 observations.

M : Threshold beyond which individuals drop out.

$\mu_T(\theta)$: A function representing the mean response of individuals at time *t*, where θ is an unknown, vector-valued parameter.

For example,

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma},$$

where X and Z are known design matrix, β is a fixed parameter vector and $\boldsymbol{\gamma}$ represents the random effects $\boldsymbol{\gamma} \sim N(0, G)$

Ramakrishnan, Wang, 2005, Analysis of Data from Clinical Trials with Treatment Related Dropouts, Communication in Statistics

Introduction (con’t)
MDT Likelihood

- The likelihood function under truncated normal distribution:

$$L(\mu_T, \sigma_T^2, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_T) \propto \frac{1}{\sigma_{T\,|\,i}^{(n-r_T)} \prod_{i=1}^{n-r_T} \Phi\left(\frac{M-\mu_{T\,|\,i}}{\sigma_{T\,|\,i}}\right)} \times$$

$$\exp\left\{\frac{1}{2\sigma_{T\,|\,i}^2} \sum_{i=1}^{n-r_T} \left(y^*_{r_T+i} - \mu_{T\,|\,i}\right)^2\right\} \times$$

$$\frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{\frac{1}{2} \sum_{i=1}^{n} \left(\mathbf{y}_i - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\mathbf{y}_i - \boldsymbol{\mu}\right)\right\} \times$$

$$\binom{n}{r_T} \left[1-\Phi\left(\frac{M-\mu_T}{\sigma_T}\right)\right]^{r_T} \left[\Phi\left(\frac{M-\mu_T}{\sigma_T}\right)\right]^{n-r_T} .$$

Ramakrishnan, Wang, 2005, Analysis of Data from Clinical Trials with Treatment Related Dropouts, Communication in Statistics

Introduction (con’t)
MDT Estimation

- An EM algorithm is used to simplify the maximum likelihood estimation of the parameters.
  - Observed data are incomplete data.
  - MDT and observed data form complete data.
  - In the E-step, MDT are estimated from conditional truncated normal distribution.
  - In the M-step, repeated measure method is applied to the complete data (PROC MIXED).
  - Iterate between E-step and M-step until convergence.
- The initial values for the mean and variance-covariance parameters could be obtained from a repeated measures model with missing as MAR.
- Initial estimate for *M* is given by the truncated mean and variance at time *T*:

$$M_0 = \mu_{0T} + \sigma_{0T}\Phi^{-1}\left(1 - \frac{r_T}{n}\right)$$

Ramakrishnan, Wang, 2005, Analysis of Data from Clinical Trials with Treatment Related Dropouts, Communication in Statistics

Introduction (con't)
MDT Estimation (2)

- This procedure could be extended for the case where the data MDT starts occurring at any time point *t* prior to *T*.
- The application of this procedure to the situation where the dropouts occur in the opposite end of the distribution is straightforward.
- Similarly, the extension of this procedure to the situation where the MDT occurs on both sides of the distribution are also possible.

Ramakrishnan, Wang, 2005, Analysis of Data from Clinical Trials with Treatment Related Dropouts, Communication in Statistics

Introduction (con't)

Other Nonignorable Missing Methods

Under the assumption that the subjects are modeled as independent, that is,

$$f(\mathbf{Y},\mathbf{I}\,|\,\mathbf{X},\boldsymbol{\theta},\xi)=\prod_{i=1}^{n} f(\mathbf{y}_{i,}\mathbf{I}_i\,|\,\mathbf{X}_i,\boldsymbol{\theta},\xi).$$

- Selection model

$$f(\mathbf{y}_{i,}\mathbf{I}_i\,|\,\mathbf{X}_i,\boldsymbol{\theta},\xi)= f(\mathbf{y}_i\,|\,\mathbf{X}_i,\boldsymbol{\theta})\,f(\mathbf{I}_i\,|\,\mathbf{X}_i,\mathbf{y}_{i,}\xi)$$

- Pattern-mixture model

$$f(\mathbf{y}_i,\mathbf{I}_i\,|\,\mathbf{X}_i,\boldsymbol{\theta},\xi)= f(\mathbf{y}_i\,|\,\mathbf{X}_i,\mathbf{I}_i,\boldsymbol{\theta})\,f(\mathbf{I}_i\,|\,\mathbf{X}_i,\xi)$$

  - The main difference between MDT method and the selection model is that the conditional distributions of $f(\mathbf{y}_i\,|\,\mathbf{X}_i,\mathbf{I}_i,\boldsymbol{\theta})$ don't depend on $\mathbf{I}_i$ in the selection model.

  - The MDT and the pattern-mixture model seem identical. However, there exists a fundamental difference in modeling the conditional distribution $f(\mathbf{y}_i\,|\,\mathbf{X}_i,\mathbf{I}_i,\boldsymbol{\theta})$. Pattern-mixture, in general, models the distribution for both $I_{ij}=0$ and $I_{ij}=1$ using the same family of distribution. In MDT case, the conditional distributions $\mathbf{y}_i$ given I_{ij} are allowed to come from a different family of distribution.

Introduction (con't)
Topics Addressed in This Talk

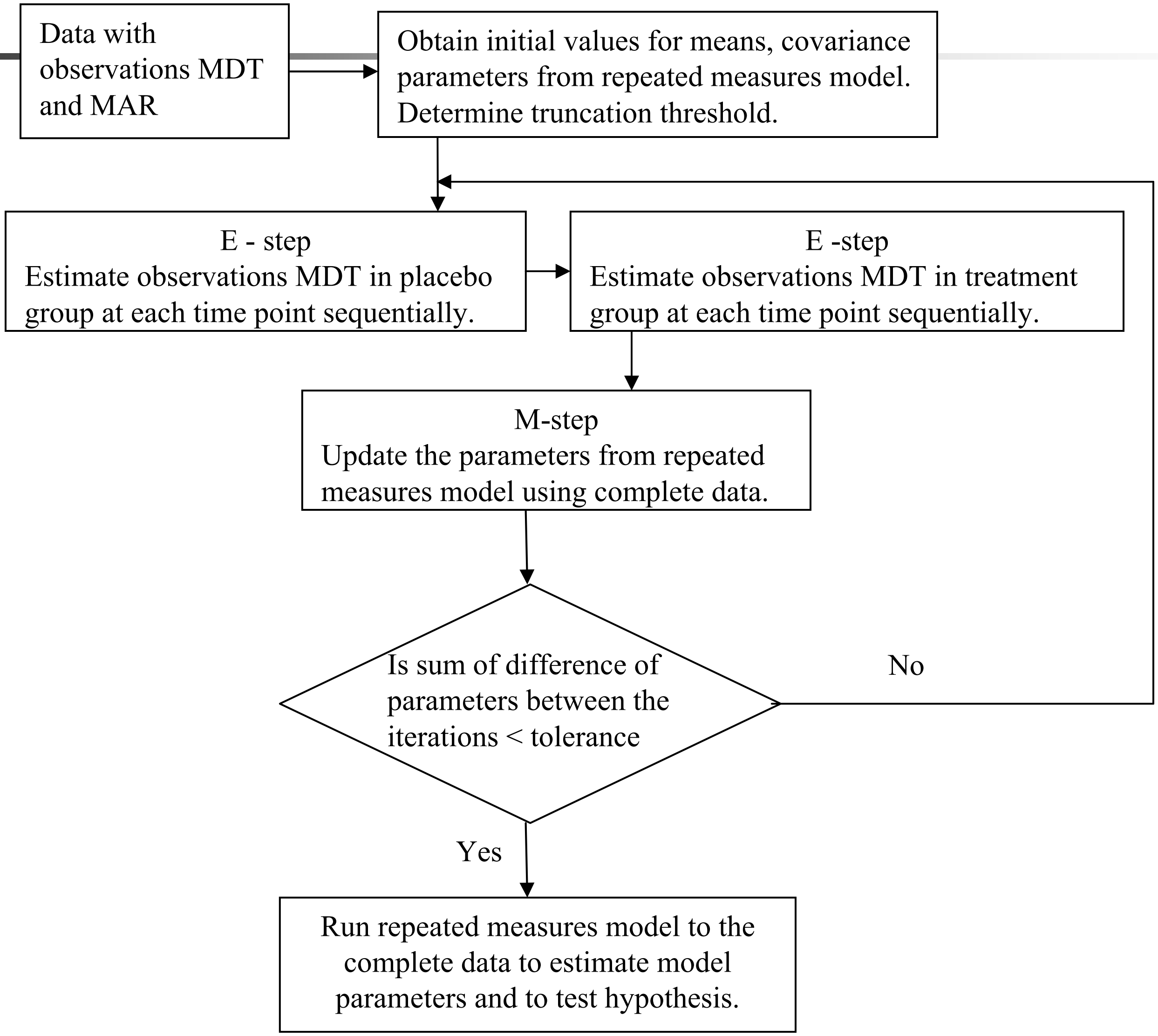
- Implementation of EM algorithm for MDT method
- Comparisons using simulation
- MDT method in conjunction with Multiple imputation method (MI)

Topics Addressed in This Talk

- **Implementation of EM algorithm for MDT method**
- Comparisons using simulation
- MDT method in conjunction with Multiple imputation method (MI)

Implementation of EM algorithm for MDT method

Flowchart of MDT Method



Data with observations MDT and MAR

Obtain initial values for means, covariance parameters from repeated measures model. Determine truncation threshold.

E - step
Estimate observations MDT in placebo group at each time point sequentially.

E -step
Estimate observations MDT in treatment group at each time point sequentially.

M-step
Update the parameters from repeated measures model using complete data.

Is sum of difference of parameters between the iterations < tolerance

No

Yes

Run repeated measures model to the complete data to estimate model parameters and to test hypothesis.

Topics Addressed in This Talk

- Implementation of EM algorithm for MDT method
- **Comparisons using simulation**
- MDT method in conjunction with Multiple imputation method (MI)

Simulation Study for MDT Method

Purpose: To compare other relevant methods

- Data were generated from a multivariate normal distribution with four time points and MDT were created using a threshold *M* based on the dropout rate.
- Simulation characteristics
  - Linear and quadratic response functions with AR(1) error structure.
  - Missing data mechanisms: MDT and MAR.
  - AR(1) correlation: 0.2, 0.4 and 0.8 (with same variance of 2).
  - Dropout rate: one missing – 5%, 10% and 20% at time 4.

    two missing – 3%, 7% and 10% at time 3 with 5%, 10% and 20% at time 4 respectively.
  - Sample size: n = 50, 100, 200.
  - Simulation number: SN=100.
  - Other imputation methods: LOCF , REG and MIXED.

Simulation Study for MDT Method (con’t)

Other Relevant Methods

- Last observation carried forward method (LOCF)
  - Assigns the person’s last known observation to the missing value.
- Individual regression prediction method (REG)
  - Extrapolates the missing observations based on a regression fit between the outcome variable and time for each subject with missing value.
- Repeated measures mixed model method (MIXED)
  - Treats missing value as MAR.

Simulation Study for MDT Method (con’t)

Table B.1 Mean ($\mu_4 = 2.6$) estimates (s.e) from different methods for linear response & MDT at last time point

	$n = 50$			$n = 100$			$n = 200$		
Missing	5%	10%	20%	5%	10%	20%	5%	10%	20%
Method									
					$\rho = 0.2$				
MDT	2.652(0.188)	2.685(0.189)	2.747(0.197)	2.647(0.128)	2.679(0.128)	2.738(0.133)	2.642(0.087)	2.673(0.087)	2.731(0.089)
# of itera	4.030(1.359)	5.570(1.289)	8.190(1.426)	4.360(0.927)	5.770(0.839)	8.500(1.000)	4.570(0.700)	6.070(0.624)	8.190(1.426)
LOCF	2.790(0.171)	2.938(0.170)	3.178(0.185)	2.781(0.121)	2.928(0.119)	3.161(0.125)	2.773(0.083)	2.915(0.080)	3.1540.086)
REG	2.695(0.200)	2.764(0.218)	2.846(0.285)	2.692(0.148)	2.758(0.162)	2.828(0.206)	2.686(0.100)	2.747(0.110)	2.822(0.144)
MIXED	2.760(0.181)	2.881(0.181)	3.102(0.191)	2.752(0.124)	2.871(0.123)	3.084(0.129)	2.745(0.083)	2.861(0.081)	3.078(0.085)
					$\rho = 0.4$				
MDT	2.639(0.191)	2.663(0.192)	2.708(0.195)	2.634(0.128)	2.656(0.128)	1.868(0.127)	2.630(0.086)	2.652(0.085)	2.695(0.086)
# of itera	4.120(1.647)	5.770(1.413)	8.410(1.450)	4.450(0.914)	5.980(1.034)	8.500(1.000)	4.800(0.696)	6.380(0.722)	8.410(1.450)
LOCF	2.760(0.169)	2.879(0.167)	3.096(0.155)	2.749(0.115)	2.869(0.113)	1.833(0.135)	2.745(0.076)	2.861(0.074)	3.076(0.073)
REG	2.648(0.194)	2.666(0.219)	2.710(0.244)	2.643(0.145)	2.665(0.166)	2.295(0.275)	2.638(0.096)	2.654(0.108)	2.684(0.130)
MIXED	2.741(0.178)	2.846(0.182)	3.042(0.176)	2.732(0.119)	2.836(0.120)	1.828(0.133)	2.728(0.079)	2.829(0.078)	3.022(0.076)
					$\rho = 0.8$				
MDT	2.592(0.201)	2.582(0.205)	2.570(0.213)	2.595(0.146)	2.585(0.148)	2.571(0.150)	2.593(0.094)	2.584(0.096)	2.570(0.098)
# of itera	4.440(1.647)	6.290(1.066)	8.850(1.403)	4.900(1.159)	6.700(0.893)	9.310(1.152)	5.290(0.715)	7.160(0.678)	8.850(1.403)
LOCF	2.692(0.193)	2.771(0.180)	2.918(0.175)	2.690(0.137)	2.772(0.121)	2.916(0.118)	2.688(0.090)	2.775(0.079)	2.915(0.073)
REG	2.592(0.199)	2.584(0.206)	2.573(0.218)	2.592(0.154)	2.583(0.164)	2.570(0.176)	2.590(0.098)	2.579(0.100)	2.561(0.111)
MIXED	2.659(0.195)	2.704(0.188)	2.788(0.192)	2.660(0.139)	2.706(0.131)	2.787(0.131)	2.657(0.091)	2.706(0.085)	2.784(0.083)

Simulation Study for MDT Method (con’t)

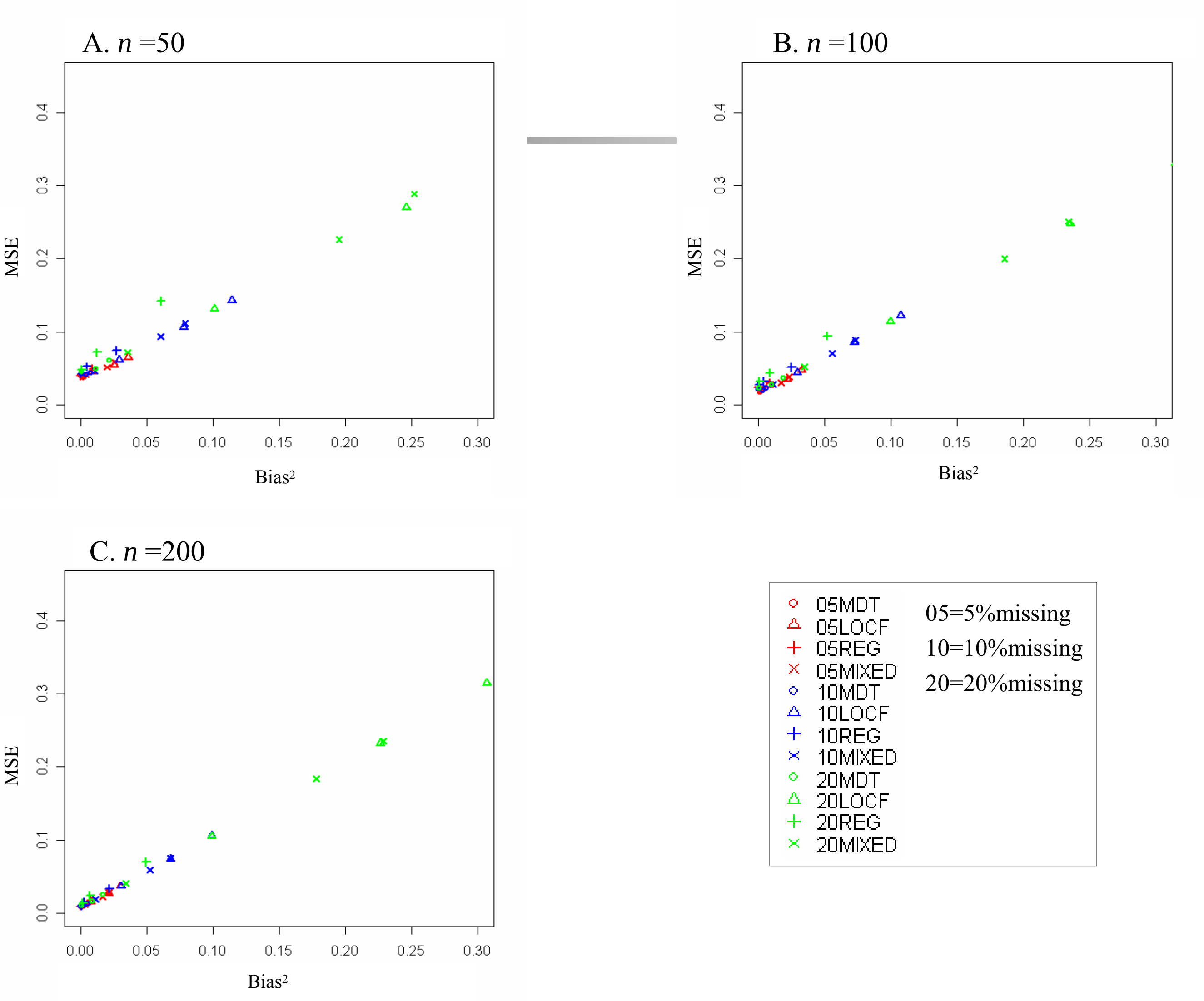


Figure B.1 Mean (μ_4) estimates from different methods for linear response, MDT at last time point and AR(1)=0.2, 0.4 and 0.8.

Simulation Study for MDT Method (con’t)

Simulation Results Summary

- When missing proportion is small all the methods perform reasonably well.
- Although regression method estimates the means accurately for linear response function, it typically over estimates the variance and correlation especially when the correlation is low.
- The LOCF and regression are both sensitive to the forms of response function.
- when the missing are not MAR, the bias of the estimates from MIXED method is large for most cases.
- When the data are missing due to truncation, MDT method performs best for all the parameters regardless of missing proportion and the forms of response function.

Simulation Study for MDT Method (con’t)

Simulation Conclusion

- In practice, the choice of the method for dealing with the missing data is important especially when large proportion is missing. The MDT method should be used if the form of the model is unknown and there is reason to believe the assumption of truncated normal distribution is appropriate.
- When the missing mechanism is unknown, the application of MDT method is not recommended.

Topics Addressed in This Talk

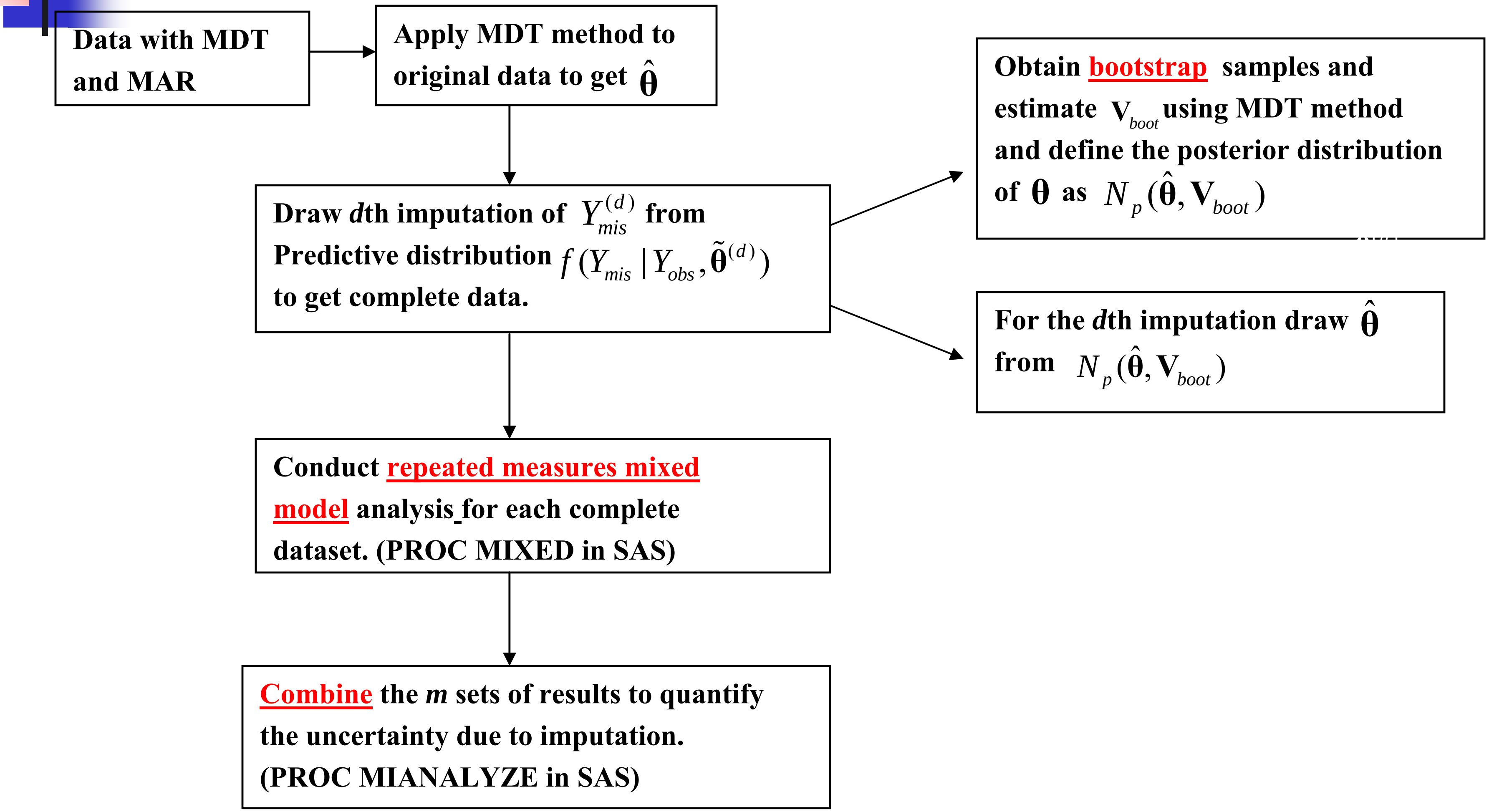
- Implementation of EM algorithm for MDT method
- Comparisons using simulation
- **MDT method in conjunction with Multiple imputation method (MI)**

MDT Method in conjunction with Multiple Imputation (MI)

- First developed by Rubin (1977) but became more popular in the past 10 years as the computer power grew.
- The idea is to replace each missing value with two or more accepted values so that the uncertainty about the right value to impute could incorporated into the analysis.
- The procedure is to,
  - Create *m* ($m \geq 2$) complete data sets by replacing each missing value with *m* repeated random draws from a predictive distribution of the missing data
  - Analyze each of the *m* complete data sets using standard complete data procedures.
  - Combine *m* sets of the point and variance estimates by ‘Rubin’s rule’ to make valid inferences.
- Majority of MI procedures involve the use of ignorable missing models. However, MI can also be used with nonignorable missing data.

MDT Method in conjunction with Multiple Imputation (con't)

Flowchart of MDT method in conjunction with MI



Data with MDT and MAR

Apply MDT method to original data to get $\hat{\boldsymbol{\theta}}$

Draw *d*th imputation of $Y_{mis}^{(d)}$ from Predictive distribution $f(Y_{mis} \mid Y_{obs}, \tilde{\boldsymbol{\theta}}^{(d)})$ to get complete data.

Obtain bootstrap samples and estimate $\mathbf{V}_{boot}$ using MDT method and define the posterior distribution of $\boldsymbol{\theta}$ as $N_p(\hat{\boldsymbol{\theta}}, \mathbf{V}_{boot})$

For the *d*th imputation draw $\hat{\boldsymbol{\theta}}$ from $N_p(\hat{\boldsymbol{\theta}}, \mathbf{V}_{boot})$

Conduct repeated measures mixed model analysis for each complete dataset. (PROC MIXED in SAS)

Combine the *m* sets of results to quantify the uncertainty due to imputation. (PROC MIANALYZE in SAS)

MDT Method in conjunction with Multiple Imputation (con't)

Application to IMPS Data

Table 1.2 IMPS data summary

	Treatment (N=71)		Placebo (N=24)		Total (N=95)	
Completion Status	n	%	n	%	n	%
Complete	**61**	**86**	**18**	**75**	**79**	**83**
Incomplete	**10**	**14**	**6**	**25**	**16**	**17**
Worsening	2	20	5	83	7	44
Improved	1	10	0	0	1	6
MAR	7	70	1	17	8	50

MDT Method in conjunction with Multiple Imputation (con’t)

Analysis Results (1)

Table 5.1 Variance Information using MDT Method in conjunction with MI

	Variance				Relative	Fraction
					Increase	Missing
Parameter	Between	Within	Total	DF	in Variance	Information
Intercept	0.00002	0.073	0.073	280.93	0.0003	0.0003
group	0.00002	0.094	0.094	280.95	0.0003	0.0003
stime	0.00111	0.021	0.023	245.12	0.0567	0.0542
group*stime	0.00148	0.029	0.031	246.24	0.0554	0.0530

Stime: square root transformation of TIME variable

MDT Method in conjunction with Multiple Imputation (con't)

Analysis Results (1)

Table 5.2 Parameter Estimates using MDT Method in conjunction with MI

Parameter	Estimate	Std Error	DF	Minimum	Maximum	*t* for H_0:θ=0	Pr > |t|
Intercept	5.194	0.270	280.93	5.188	5.203	19.23	<.0001
group	0.037	0.307	280.95	0.030	0.043	0.12	0.9045
stime	-0.465	0.151	245.12	-0.513	-0.385	-3.08	0.0023
group* stime	-0.325	0.176	246.24	-0.419	-0.284	-1.85	0.0655

Stime: square root transformation of TIME variable

MDT Method in conjunction with Multiple Imputation (con't)

Comparison with Multiple Imputation Assuming MAR

Table 5.4 Multiple Imputation parameter Estimates using PROC MI

Parameter	Estimate	Std Error	DF	Minimum	Maximum	*t* for H_0:θ=0	Pr > |t|
Intercept	5.252	0.283	280.71	5.245	5.265	18.56	<.0001
group	0.016	0.316	280.78	0.007	0.027	0.05	0.9605
Stime	-0.589	0.132	258.49	-0.611	-0.564	-4.47	<.0001
group*stime	-0.159	0.152	249.89	-0.199	-0.130	-1.04	0.2972

Stime: square root transformation of TIME variable

Summary and Future Work

- Extend MDT method to multivariate outcomes
  - Subject drops out of the study due to the subject exceeding threshold
  - Due to all the outcomes exceeding the thresholds
  - Due to some of the outcomes exceeding the thresholds
- Allow for the threshold to be random
  - Thresholds may vary among subjects
  - Assume subject-specific threshold is a random variable from a uniform distribution based on clinical knowledge
  - Define the threshold depending on the subject's previous observations
- Estimate missing values from the same subject simultaneously
  - When the MDT occurs at multiple time points
  - Estimate the MDTs simultaneously using multivariate truncated normal distribution.
  - The estimation involves multivariate normal CDFs.

Questions?