Paper 1467-2014: Missing Data Overview, Likelihood, Weighted Estimating Equations, and Multiple Imputation

Geert Molenberghs

geert.molenberghs@uhasselt.be & geert.molenberghs@med.kuleuven.be

Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat)

Universiteit Hasselt, Belgium & Katholieke Universiteit Leuven

www.censtat.uhasselt.be & www.kuleuven.ac.be/biostat/

Interuniversity Institute for Biostatistics and statistical Bioinformatics

Washington DC, March 24, 2014

Chapter 1
A Gentle Tour

- ▷ Orthodontic growth data
- ▷ Commonly used methods
- ▷ Survey of the terrain

1.1 Growth Data: An (Un)balanced Discussion

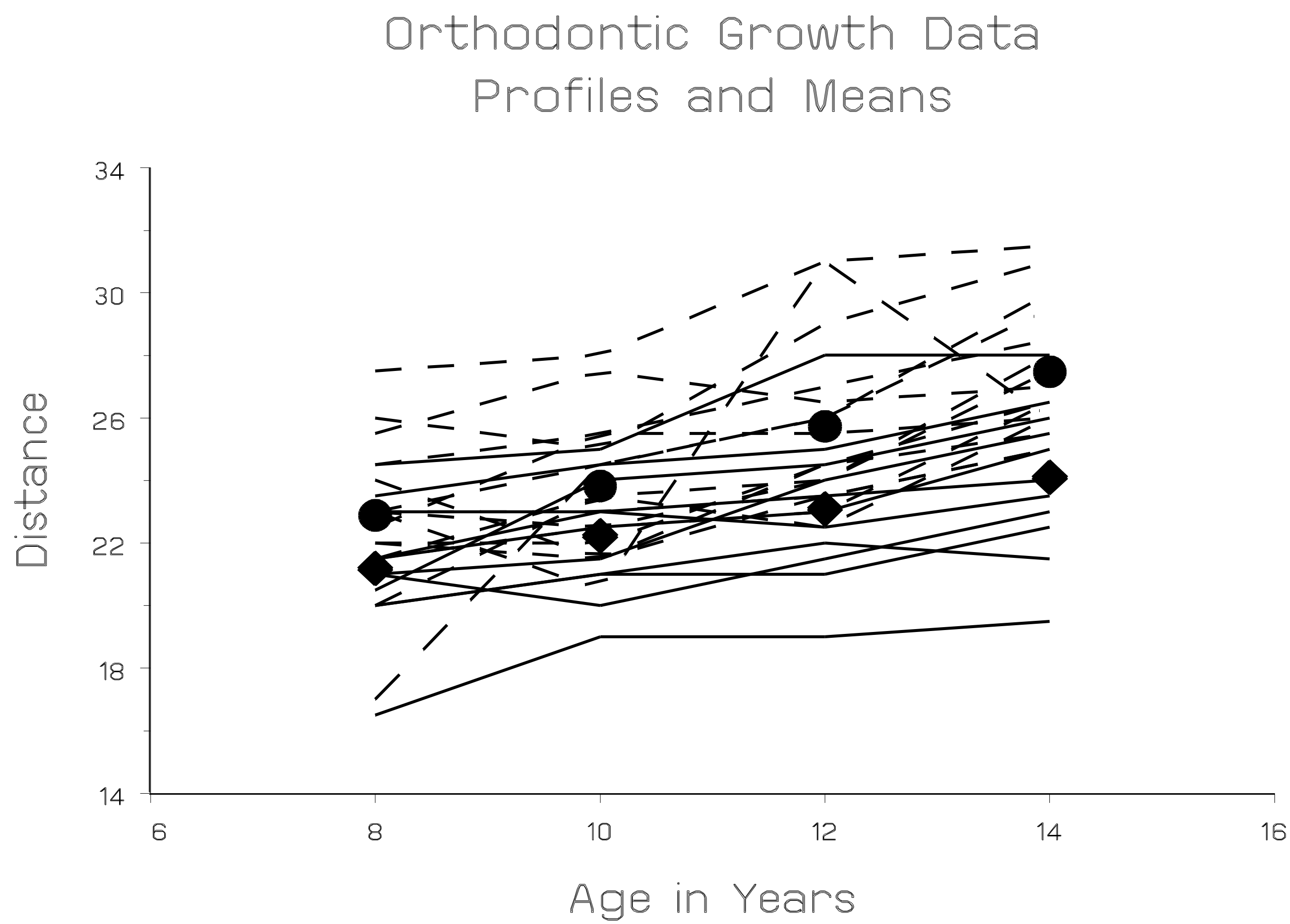
- Taken from Potthoff and Roy, Biometrika (1964)

- Research question:

Is dental growth related to gender ?

- The distance from the center of the pituitary to the maxillary fissure was recorded at ages 8, 10, 12, and 14, for 11 girls and 16 boys

- Individual profiles:
  - ▷ **Un**balanced data
  - ▷ Balanced data



1.2 LOCF, CC, or Direct Likelihood?

Data:

$$\begin{array}{|c|c|}\hline 20 & 30 \\ \hline 10 & 40 \\ \hline \end{array} \quad \begin{array}{|c|}\hline 75 \\ \hline 25 \\ \hline \end{array}$$

LOCF:

$$\begin{array}{|c|c|}\hline 20 & 30 \\ \hline 10 & 40 \\ \hline \end{array} \quad \begin{array}{|c|c|}\hline 75 & 0 \\ \hline 0 & 25 \\ \hline \end{array} \quad \Longrightarrow \quad \begin{array}{|c|c|}\hline 95 & 30 \\ \hline 10 & 65 \\ \hline \end{array} \quad \Longrightarrow \quad \widehat{\theta} = \frac{95}{200} = 0.475 \; [0.406; 0.544] \text{ (biased \& too narrow)}$$

CC:

$$\begin{array}{|c|c|}\hline 20 & 30 \\ \hline 10 & 40 \\ \hline \end{array} \quad \begin{array}{|c|c|}\hline 0 & 0 \\ \hline 0 & 0 \\ \hline \end{array} \quad \Longrightarrow \quad \begin{array}{|c|c|}\hline 20 & 30 \\ \hline 10 & 40 \\ \hline \end{array} \quad \Longrightarrow \quad \widehat{\theta} = \frac{20}{100} = 0.200 \; [0.122; 0.278] \text{ (biased \& too wide)}$$

d.l.(MAR):

$$\begin{array}{|c|c|}\hline 20 & 30 \\ \hline 10 & 40 \\ \hline \end{array} \quad \begin{array}{|c|c|}\hline 30 & 45 \\ \hline 5 & 20 \\ \hline \end{array} \quad \Longrightarrow \quad \begin{array}{|c|c|}\hline 50 & 75 \\ \hline 15 & 60 \\ \hline \end{array} \quad \Longrightarrow \quad \widehat{\theta} = \frac{50}{200} = 0.250 \; [0.163; 0.337]$$

1.3 Direct Likelihood/Bayesian Inference: Ignorability

$$\boxed{\text{MAR}} : f(\boldsymbol{Y}_i^o | \boldsymbol{X}_i, \boldsymbol{\theta}) \cancel{f(\boldsymbol{r}_i | \boldsymbol{X}_i, \boldsymbol{Y}_i^o, \boldsymbol{\psi})}$$

$$\left.\begin{array}{r} \text{Mechanism is MAR} \\ \boldsymbol{\theta} \text{ and } \boldsymbol{\psi} \text{ distinct} \\ \text{Interest in } \boldsymbol{\theta} \\ \text{(Use observed information matrix)} \end{array}\right\} \Longrightarrow \quad \text{Lik./Bayes inference valid}$$

Outcome type	Modeling strategy	Software
Gaussian	Linear mixed model	SAS® MIXED
Non-Gaussian	Gen./Non-linear mixed model	SAS GLIMMIX, NLMIXED

1.4 Rubin, 1976

- Ignorability: Rubin (Biometrika, 1976): 35 years ago!
- Little and Rubin (1976, 2002)
- Why did it take so long?

1.5 A Vicious Triangle

$$\begin{array}{ccccc} & & \textbf{Industry} & & \\ & \nearrow\swarrow & & \searrow\nwarrow & \\ \textbf{Academe} & & \begin{array}{c}\longleftarrow \\ \longrightarrow\end{array} & & \textbf{Regulatory} \end{array}$$

- **Academe:** The R^2 principle
- **Regulatory:** Rigid procedures $\longleftrightarrow$ scientific developments
- **Industry:** We **cannot / do not want to** apply new methods

1.6 Terminology & Confusion

- The Ministry of Disinformation:

$\longleftarrow$ $\boxed{\textbf{All directions}}$

$\boxed{\textbf{Other directions}}$ $\longrightarrow$

- **MCAR, MAR, MNAR:** "What do the terms mean?"
- **MAR, random dropout, informative missingness, ignorable, censoring,. . .**
- **Dropout from the study, dropout from treatment, lost to follow up,. . .**
- **"Under MAR patients dropping out and patients not dropping out are similar."**

1.7 A Virtuous Triangle

$$
\begin{tikzcd}
 & \textbf{Industry} \arrow[dl, shift left] \arrow[dr, shift right] & \\
\textbf{Academe} \arrow[ur, shift left] \arrow[rr, shift right] & & \textbf{Regulatory} \arrow[ul, shift right] \arrow[ll, shift right]
\end{tikzcd}
$$

- FDA/Industry Workshops
- DIA/EMA Meetings
- **The NAS Experience**

1.8 The NAS Experience: A Wholesome Product

- **FDA $\longrightarrow$ NAS $\longrightarrow$ the working group**
- **Composition**
- **Encompassing:**
  - ▷ terminology/taxonomy/concepts
  - ▷ prevention
  - ▷ treatment

1.9 Taxonomy

- **Missingness pattern:** complete — monotone — non-monotone
- Dropout pattern: complete — dropout — intermittent
- **Model framework:** SEM — PMM — SPM
- **Missingness mechanism:** MCAR — MAR — MNAR
- **Ignorability:** ignorable — non-ignorable
- **Inference paradigm:** frequentist — likelihood — Bayes

1.10 The NAS Panel

Name	Specialty	Affiliation
Rod Little	biostat	U Michigan
Ralph D'Agostino	biostat	Boston U
Kay Dickerson	epi	Johns Hopkins
Scott Emerson	biostat	U Washington
John Farrar	epi	U Penn
Constantine Frangakis	biostat	Johns Hopkins
Joseph Hogan	biostat	Brown U
Geert Molenberghs	biostat	U Hasselt & K.U.Leuven
Susan Murphy	stat	U Michigan
James Neaton	biostat	U Minnesota
Andrea Rotnitzky	stat	Buenos Aires & Harvard
Dan Scharfstein	biostat	Johns Hopkins
Joseph Shih	biostat	New Jersey SPH
Jay Siegel	biostast	J&J
Hal Stern	stat	UC at Irvine

1.11 Modeling Frameworks & Missing Data Mechanisms

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i \mid X_i, \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\boldsymbol{y}_i \mid X_i, \boldsymbol{\theta}) \; \boxed{f(\boldsymbol{r}_i \mid X_i, \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{\psi})}$

MCAR $\longrightarrow$ **MAR** $\longrightarrow$ **MNAR**

$f(\boldsymbol{r}_i \mid X_i, \boldsymbol{\psi})$ $\qquad$ $f(\boldsymbol{r}_i \mid X_i, \boldsymbol{y}_i^o, \boldsymbol{\psi})$ $\qquad$ $f(\boldsymbol{r}_i \mid X_i, \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{\psi})$

Pattern-mixture Models: $f(\boldsymbol{y}_i \mid X_i, \boldsymbol{r}_i, \boldsymbol{\theta}) \, f(\boldsymbol{r}_i \mid X_i, \boldsymbol{\psi})$

Shared-parameter Models: $f(\boldsymbol{y}_i \mid X_i, \boldsymbol{b}_i, \boldsymbol{\theta}) \, f(\boldsymbol{r}_i \mid X_i, \boldsymbol{b}_i, \boldsymbol{\psi})$

1.12 Frameworks and Their Methods

$$f(\boldsymbol{Y}_i, \boldsymbol{r}_i | X_i, \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\boldsymbol{Y}_i | X_i, \boldsymbol{\theta}) \boxed{f(\boldsymbol{r}_i | X_i, \boldsymbol{Y}_i^o, \boldsymbol{Y}_i^m, \boldsymbol{\psi})}$

$\boxed{\textbf{MCAR/simple}}$	$\longrightarrow$	$\boxed{\textbf{MAR}}$	$\longrightarrow$	$\boxed{\textbf{MNAR}}$
CC?		**direct likelihood!**		joint model!?
LOCF?		**direct Bayesian!**		sensitivity analysis?!
single imputation?		**multiple imputation (MI)!**		
⋮		**IPW** $\supset$ **W-GEE!**		
		d.l. $+$ **IPW** $=$ **double robustness! (consensus)**		

1.13 Frameworks and Their Methods: Start

$$f(\boldsymbol{Y}_i, \boldsymbol{r}_i | X_i, \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\boldsymbol{Y}_i | X_i, \boldsymbol{\theta}) \boxed{f(\boldsymbol{r}_i | X_i, \boldsymbol{Y}_i^o, \boldsymbol{Y}_i^m, \boldsymbol{\psi})}$

$\boxed{\textbf{MCAR/simple}} \longrightarrow \boxed{\textbf{MAR}} \longrightarrow \boxed{\textbf{MNAR}}$

direct likelihood!

direct Bayesian!

multiple imputation (MI)!

IPW $\supset$ W-GEE!

d.l. $+$ IPW $=$ double robustness!

1.14 Frameworks and Their Methods: Next

$$f(\boldsymbol{Y}_i, \boldsymbol{r}_i | X_i, \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\boldsymbol{Y}_i | X_i, \boldsymbol{\theta}) \boxed{f(\boldsymbol{r}_i | X_i, \boldsymbol{Y}_i^o, \boldsymbol{Y}_i^m, \boldsymbol{\psi})}$

MCAR/simple $\longrightarrow$ **MAR** $\longrightarrow$ **MNAR**

~~joint model!?~~

sensitivity analysis!

PMM

MI (MGK, J&J)

local influence

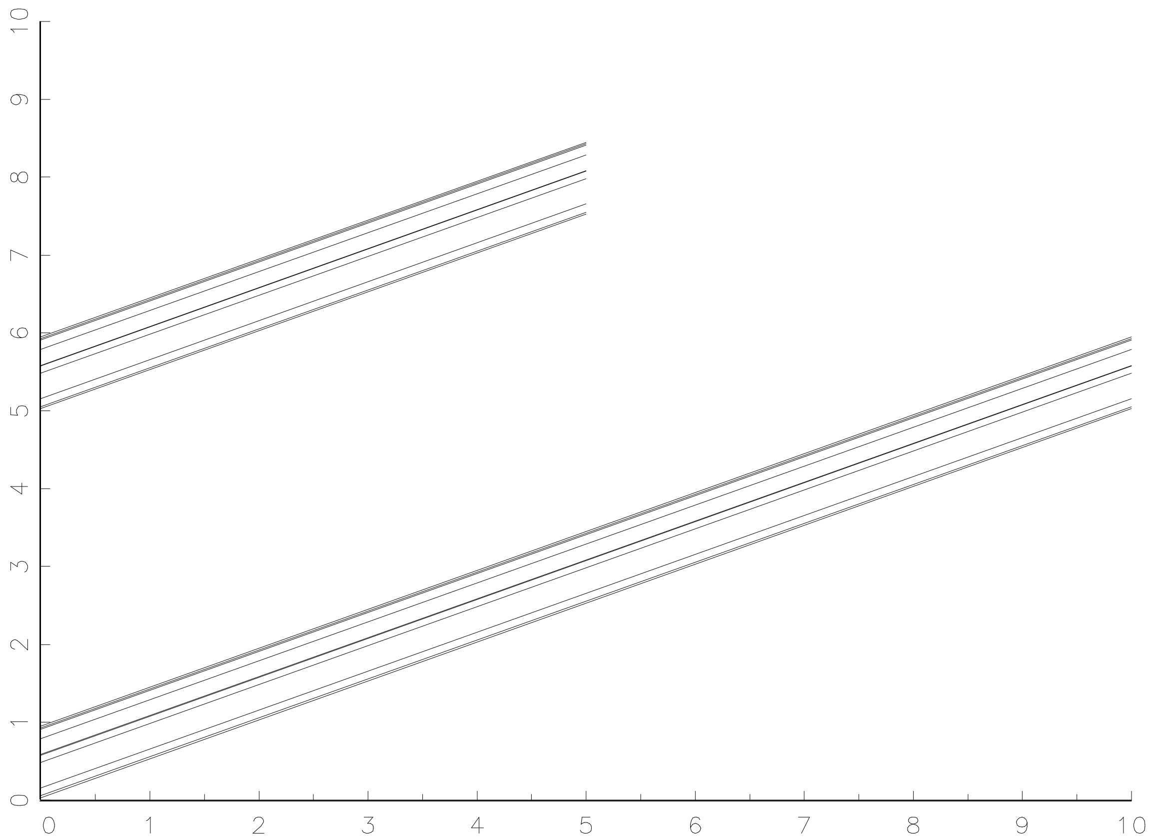
interval ignorance

IPW based

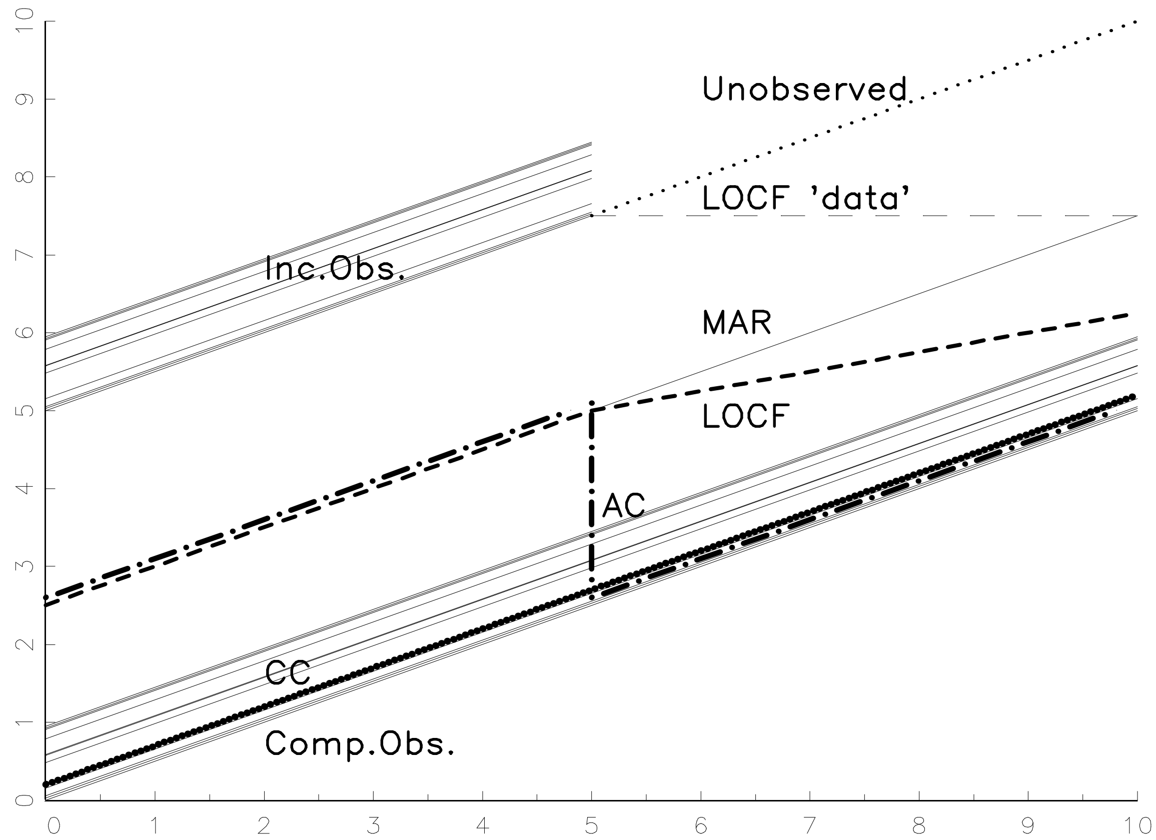
1.15 Overview and (Premature) Conclusion

MCAR/simple	CC LOCF single imputation	biased inefficient not simpler than MAR methods
MAR	direct lik./Bayes IPW/d.r. multiple imputation	easy to conduct Gaussian & non-Gaussian
MNAR	variety of methods	strong, untestable assumptions most useful in sensitivity analysis

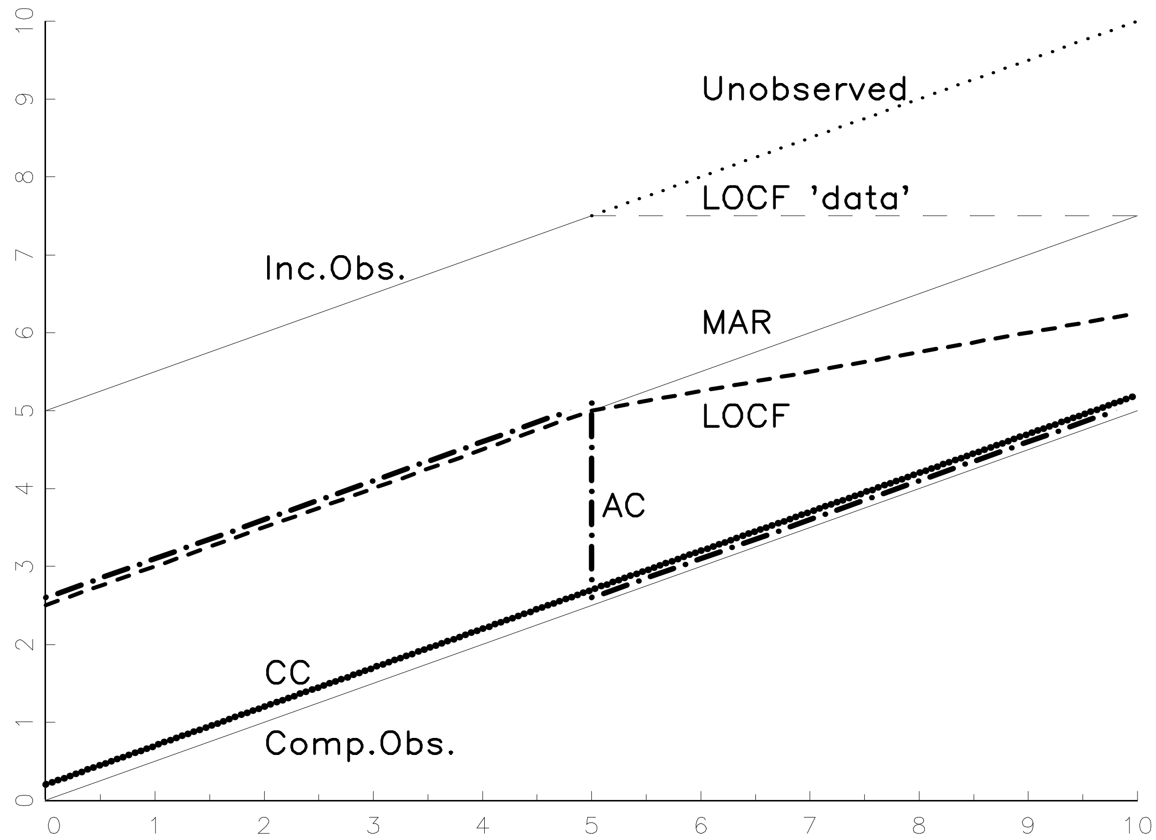
1.16 Incomplete Longitudinal Data



Data and Modeling Strategies

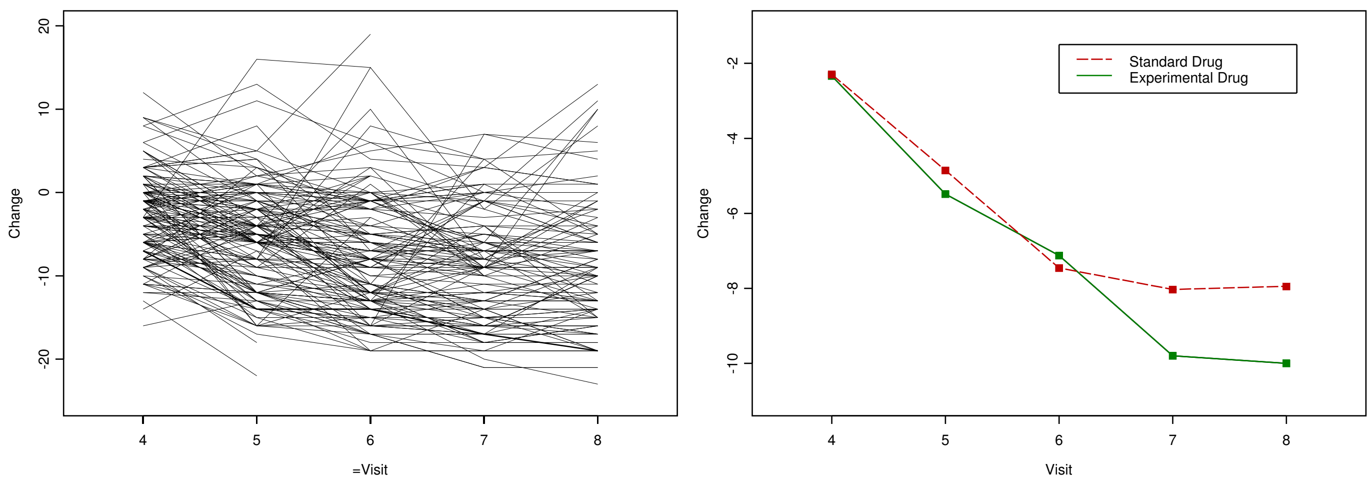


Modeling Strategies



1.17 The Depression Trial

- Clinical trial: experimental drug *versus* standard drug
- 170 patients
- Response: change versus baseline in $HAMD_{17}$ score
- 5 post-baseline visits: 4–8



1.18 Analysis of the Depression Trial

- Treatment effect at visit 8 (last follow-up measurement):

Method	Estimate	(s.e.)	p-value
CC	-1.94	(1.17)	0.0995
LOCF	-1.63	(1.08)	0.1322
MAR	-2.38	(1.16)	0.0419

Observe the slightly significant p-value under the MAR model

Chapter 2
Multiple Imputation

- Valid under MAR

- An alternative to direct likelihood and WGEE

- Three steps:

  1. The missing values are filled in M times $\Longrightarrow$ M complete data sets

  2. The M complete data sets are analyzed by using standard procedures

  3. The results from the M analyses are combined into a single inference

- Rubin (1987), Rubin and Schenker (1986), Little and Rubin (1987)

2.1 Use of MI in Practice

- Many analyses of the same incomplete set of data
- A combination of missing outcomes and missing covariates
- As an alternative to WGEE: MI can be combined with classical GEE
- MI in SAS:

Imputation Task:	$\boxed{\text{PROC MI}}$
	$\downarrow$
Analysis Task:	$\boxed{\text{PROC “MYFAVORITE”}}$
	$\downarrow$
Inference Task:	$\boxed{\text{PROC MIANALYZE}}$

2.2 Age-related Macular Degeneration Trial

- Pharmacological Therapy for Macular Degeneration Study Group (1997)
- An occular pressure disease which makes patients progressively lose vision
- 240 patients enrolled in a multi-center trial (190 completers)
- Treatment: Interferon-α (6 million units) versus placebo
- Visits: baseline and follow-up at 4, 12, 24, and 52 weeks
- Continuous outcome: visual acuity: $\#$ letters correctly read on a vision chart
- Binary outcome: visual acuity versus baseline ≥ 0 or ≤ 0

- Missingness:

Measurement occasion					
4 wks	12 wks	24 wks	52 wks	Number	%
Completers					
O	O	O	O	188	78.33
Dropouts					
O	O	O	M	24	10.00
O	O	M	M	8	3.33
O	M	M	M	6	2.50
M	M	M	M	6	2.50
Non-monotone missingness					
O	O	M	O	4	1.67
O	M	M	O	1	0.42
M	O	O	O	2	0.83
M	O	M	M	1	0.42

2.3 MI Analysis of the ARMD Trial

- $M = 10$ imputations

- GEE:

$$\text{logit}[P(Y_{ij} = 1|T_i, t_j)] = \beta_{j1} + \beta_{j2}T_i$$

- GLMM:

$$\text{logit}[P(Y_{ij} = 1|T_i, t_j, b_i)] = \beta_{j1} + b_i + \beta_{j2}T_i, \qquad b_i \sim N(0, \tau^2)$$

- $T_i = 0$ for placebo and $T_i = 1$ for interferon-α

- t_j $(j = 1, \ldots, 4)$ refers to the four follow-up measurements

- Imputation based on the continuous outcome

- Results:

Effect	Par.	GEE	GLMM
Int.4	β_{11}	-0.84(0.20)	-1.46(0.36)
Int.12	β_{21}	-1.02(0.22)	-1.75(0.38)
Int.24	β_{31}	-1.07(0.23)	-1.83(0.38)
Int.52	β_{41}	-1.61(0.27)	-2.69(0.45)
Trt.4	β_{12}	0.21(0.28)	0.32(0.48)
Trt.12	β_{22}	0.60(0.29)	0.99(0.49)
Trt.24	β_{32}	0.43(0.30)	0.67(0.51)
Trt.52	β_{42}	0.37(0.35)	0.52(0.56)
R.I. s.d.	τ		2.20(0.26)
R.I. var.	τ^2		4.85(1.13)

2.4 Generalized Estimating Equations

Liang and Zeger (1986)

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} [D_i]^T \, [V_i(\boldsymbol{\alpha})]^{-1} \, (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

- $\boldsymbol{V}_i(.)$ is not the true variance of $\boldsymbol{Y}_i$ but only a plausible guess
- the score equations are solved in a standard way
- Asymptotic distribution:

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \;\sim\; N(\mathbf{0}, I_0^{-1} I_1 I_0^{-1})$$

$$I_0 = \sum_{i=1}^{N} D_i^T [V_i(\boldsymbol{\alpha})]^{-1} D_i \qquad I_1 = \sum_{i=1}^{N} D_i^T [V_i(\boldsymbol{\alpha})]^{-1} \mathsf{Var}(\boldsymbol{Y}_i) [V_i(\boldsymbol{\alpha})]^{-1} D_i$$

2.5 Weighted GEE

$$\pi_i = \prod_{\ell=2}^{n_i}(1 - p_{i\ell})$$

$$\pi_i' = \left[\prod_{\ell=2}^{d_i-1}(1 - p_{i\ell})\right] \cdot p_{id_i}$$

$$p_{i\ell} = P\left(D_i = \ell | D_i \geq \ell, Y_{i\overline{\ell}}, X_{i\overline{\ell}}\right)$$

$$R_i = 1 \text{ if subject } i \text{ is complete}$$

$$R_i = 0 \text{ if subject } i \text{ is incomplete}$$

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{R_i}{\pi_i} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'} V_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

$$S(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{N} \frac{1}{\pi_i'} \frac{\partial \boldsymbol{\mu}_i^o}{\partial \boldsymbol{\beta}'} (V_i^o)^{-1} (\boldsymbol{y}_i^o - \boldsymbol{\mu}_i^o) \;=\; \mathbf{0}$$

2.6 General Definition of Pseudo-likelihood

- General definition: $(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)$ i.i.d. common density depending on $\boldsymbol{\Theta}_0$

$$p\ell \quad := \quad \sum_{i=1}^{N} \sum_{s \in S} \delta_s \ln f_s(y_i^{(s)}; \Theta_i)$$

- Special case: bivariate pseudo-likelihood $\quad f(y_1|y_2)f(y_2|y_1)$

$$\delta_{(1,1)} = \quad 2 \qquad\qquad \delta_{(1,0)} = -1 \qquad\qquad \delta_{(0,1)} = -1$$

- Special case: Likelihood $\quad f(y_1, \ldots, y_n)$

$$\delta_s = \begin{cases} 1 & \text{if } s = (1, \ldots, 1) \\ 0 & \text{otherwise} \end{cases}$$

2.7 Asymptotic Results

- $\widetilde{\boldsymbol{\Theta}}_N \xrightarrow{P} \boldsymbol{\Theta}_0$.

- $\sqrt{N}(\widetilde{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta}_0) \xrightarrow{D} N_p(\mathbf{0}, J(\boldsymbol{\Theta}_0)^{-1} K(\boldsymbol{\Theta}_0) J(\boldsymbol{\Theta}_0)^{-1})$

$$J_{k\ell} = -\sum_{s \in S} \delta_s E_{\boldsymbol{\Theta}} \left(\frac{\partial^2 \ln f_s(\boldsymbol{y}^{(s)}; \boldsymbol{\Theta})}{\partial \theta_k \partial \theta_\ell} \right)$$

$$K_{k\ell} = \sum_{s,t \in S} \delta_s \delta_t E_{\boldsymbol{\Theta}} \left(\frac{\partial \ln f_s(\boldsymbol{y}^{(s)}; \boldsymbol{\Theta})}{\partial \theta_k} \frac{\partial \ln f_t(\boldsymbol{y}^{(t)}; \boldsymbol{\Theta})}{\partial \theta_\ell} \right).$$

- For likelihood $J^{-1} K J^{-1}$ reduces to the inverse of Fisher's information matrix I.

- Cramèr-Rao implies $J^{-1}KJ^{-1} \geq I^{-1}$

2.8 Marginal (Pairwise) Pseudo-likelihood

Full likelihood	$\ln f(y_{i1},\ldots,y_{in})$
First pseudo-likelihood	$p\ell_i = \sum_{j<k} \ln f(y_{ik}, y_{ij})$
Second pseudo-likelihood	$p\ell_i^* = \sum_{j<k} \ln f(y_{ik}, y_{ij})/(n_i-1)$

- **Factor** $1/(n_i-1)$:
  - ▷ Each response occurs (n_i-1) times
  - ▷ PL reduces to ML under independence
- **Computational ease over GEE2**
  - ▷ No evaluation of 3rd and 4th order probabilities.
  - ▷ No explicit working assumptions required.
  - ▷ Only bivariate Plackett distribution is needed.

2.9 General Naive, IPW, and Doubly Robust Estimating Equations

$$
\begin{aligned}
\boldsymbol{U}_{\text{naive, CC}} &= \sum_{i=1}^{N} R_i \boldsymbol{U}_i(\boldsymbol{Y}_i) \\
\boldsymbol{U}_{\text{naive, AC}} &= \sum_{i=1}^{N} \boldsymbol{U}_i(\boldsymbol{Y}_i^o) \\
\boldsymbol{U}_{\text{IPWCC}} &= \sum_{i=1}^{N} \frac{R_i}{\pi_i} \boldsymbol{U}_i(\boldsymbol{Y}_i) \\
\boldsymbol{U}_{\text{IPWAC}} &= \sum_{i=1}^{N} \frac{R_i'}{\pi_i'} \cdot E_{Y^m|y^o} \boldsymbol{U}(Y_i) \\
\boldsymbol{U}_{\text{IPWAC,seq}} &= \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{R_{ij}}{\pi_{ij}} \cdot \boldsymbol{U}(Y_{ij} | \boldsymbol{U}_{i\bar{j}})
\end{aligned}
$$

$$\boldsymbol{U}_{\mathsf{IPWCC,dr}} = \sum_{i=1}^{N} \left[\frac{R_i}{\pi_i} U_i(\boldsymbol{Y}_i) + \left(1 - \frac{R_i}{\pi_i}\right) E_{Y_i^m | y_i^o} \boldsymbol{U}_i(\boldsymbol{Y}_i) \right]$$

$$\boldsymbol{U}_{\mathsf{IPWAC,dr}} = \sum_{i=1}^{N} \left\{ \sum_{j=1}^{n_i} \left[\frac{R_{ij}}{\pi_{ij}} \cdot \boldsymbol{U}(Y_{ij} | \boldsymbol{U}_{i\bar{j}}) + \left(1 - \frac{R_{ij}}{\pi_{ij}}\right) \cdot E_{Y^m | y^o} \boldsymbol{U}(Y_{ij} | \boldsymbol{U}_{i\bar{j}}) \right] \right\}$$

2.10 Precision Estimation

$$\boldsymbol{S}_i = (\boldsymbol{V}_i', \boldsymbol{W}_i')'$$

$$\widehat{\mathsf{Var}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) = \widehat{I_0}^{-1} \widehat{I_1} \widehat{I_0}^{-1}$$

$$I_0 = \sum_{i=1}^{N} \begin{pmatrix} \frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\theta}} & \frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\psi}} \\ 0 & \frac{\partial \boldsymbol{W}_i}{\partial \boldsymbol{\psi}} \end{pmatrix}$$

$$I_1 = \sum_{i=1}^{N} \boldsymbol{S}_i(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) \boldsymbol{S}_i'(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}})$$

2.11 Pseudo-likelihood (Naive)

$$\boldsymbol{U}_{\text{naive, CC}} = \sum_{i=1}^{N} R_i \sum_{s \in S} \delta_s \boldsymbol{U}_s(\boldsymbol{y}_i^{(s)})$$

$$\boldsymbol{U}_{\text{naive, CS}} = \sum_{i=1}^{N} \sum_{s \in S} R_{i,s} \delta_s \boldsymbol{U}_s(\boldsymbol{y}_i^{(s)})$$

$$\boldsymbol{U}_{\text{naive, AC}} = \sum_{i=1}^{N} \sum_{s \in S} \delta_s E_{Y^m|y^o} \boldsymbol{U}_s(\boldsymbol{y}_i^{(s)})$$

2.12 Pseudo-likelihood (IPW $\equiv$ Singly Robust)

$$\boldsymbol{U}_{\text{IPWCC}} = \sum_{i=1}^{N} \frac{R_i}{\pi_i} \cdot \sum_{s \in S} \delta_s \boldsymbol{U}_s(\boldsymbol{y}_i^{(s)})$$

$$\boldsymbol{U}_{\text{IPWCS}} = \sum_{i=1}^{N} \sum_{s \in S} \frac{R_{i,s}}{\pi_{i,s}} \cdot \delta_s \boldsymbol{U}_s(\boldsymbol{y}_i^{(s)})$$

$$\boldsymbol{U}_{\text{IPWAC}} = \sum_{i=1}^{N} \sum_{s \in S} \delta_s \sum_{j=1}^{n_i} I(j \in s) \frac{R_{ij}}{\pi_{ij}} \cdot \boldsymbol{U}_s(y_{ij} \boldsymbol{y}_{i\bar{j}}^{(s)})$$

2.13 Pseudo-likelihood (Doubly Robust)

$$\boldsymbol{U}_{\text{IPWCC,dr}} = \sum_{i=1}^{N} \left\{ \frac{R_i}{\pi_i} \left[\sum_{s \in S} \delta_s \boldsymbol{U}_s(\boldsymbol{y}_i^{(s)}) \right] + \left(1 - \frac{R_i}{\pi_i}\right) E_{\boldsymbol{Y}_i^m | \boldsymbol{y}_i^o} \left[\sum_{s \in S} \delta_s \boldsymbol{U}_s(\boldsymbol{y}_i^{(s)}) \right] \right\}$$

$$\boldsymbol{U}_{\text{IPWCS,dr}} = \sum_{i=1}^{N} \sum_{s \in S} \left\{ \frac{R_{i,s}}{\pi_{i,s}} \cdot \delta_s \boldsymbol{U}_s(\boldsymbol{y}_i^{(s)o}) + \left(1 - \frac{R_{i,s}}{\pi_{i,s}}\right) \cdot \delta_s E_{\boldsymbol{Y}_i^m | \boldsymbol{y}_i^o} \boldsymbol{U}_s(\boldsymbol{y}_i^{(s)}) \right\}$$

$$\boldsymbol{U}_{\text{IPWAC,dr}} = \sum_{i=1}^{N} \sum_{s \in S} \delta_s \sum_{j=1}^{n_i} I(j \in s) \times$$

$$\times \left[\frac{R_{ij}}{\pi_{ij}} \cdot \boldsymbol{U}_s(y_{ij} \boldsymbol{y}_{i\bar{j}}^{(s)}) + \left(1 - \frac{R_{ij}}{\pi_{ij}}\right) \cdot E_{\boldsymbol{Y}_i^m | \boldsymbol{y}_i^o} \boldsymbol{U}_s(y_{ij} | \boldsymbol{Y}_{i\bar{j}}^{(s)}) \right]$$

2.14 Exchangeable Version

$$\boldsymbol{U}_{\text{IPW, exch}} = \sum_{i=1}^{N} \sum_{s \in S} \delta_s \boldsymbol{U}_s \left(\boldsymbol{y}_i^{(s)o} \right)$$

2.15 Pairwise Likelihood (Naive)

$$\boldsymbol{U}_{\text{naive, CC}} = \sum_{i=1}^{N} R_i \sum_{j<k} \boldsymbol{U}(y_{ij}, y_{ik})$$

$$\boldsymbol{U}_{\text{naive, CP}} = \sum_{i=1}^{N} \sum_{j<k<d_i} \boldsymbol{U}(y_{ij}, y_{ik})$$

$$\boldsymbol{U}_{\text{naive, AC}} = \sum_{i=1}^{N} \left[\sum_{j<k<d_i} \boldsymbol{U}(y_{ij}, y_{ik}) + \sum_{j=1}^{d_i-1} (n_i - d_i + 1) U(y_{ij}) \right]$$

2.16 Pairwise Likelihood (Singly Robust)

$$\boldsymbol{U}_{\text{IPWCC}} = \sum_{i=1}^{N} \frac{R_i}{\pi_i} \left[\sum_{j<k} \boldsymbol{U}(y_{ij}, y_{ik}) \right]$$

$$\boldsymbol{U}_{\text{IPWCP}} = \sum_{i=1}^{N} \sum_{j<k<d_i} \frac{R_{ijk}}{\pi_{ijk}} \cdot U(y_{ij}, y_{ik})$$

$$\boldsymbol{U}_{\text{IPWAC}} = \sum_{i=1}^{N} \sum_{j<k} \left[\frac{R_{ij}}{\pi_{ij}} \cdot U(y_{ij}) + \frac{R_{ik}}{\pi_{ik}} \cdot U(y_{ik}|y_{ij}) \right]$$

2.17 Pairwise Likelihood (Doubly Robust)

$$\boldsymbol{U}_{\text{IPWCC,dr}} = \sum_{i=1}^{N} \left\{ \frac{R_i}{\pi_i} \left[\sum_{j<k} \boldsymbol{U}(y_{ij}, y_{ik}) \right] + \left(1 - \frac{R_i}{\pi_i}\right) E_{\boldsymbol{Y}_i^m | \boldsymbol{y}_i^o} \left[\sum_{j<k} \boldsymbol{U}(y_{ij}, y_{ik}) \right] \right\}$$

$$\boldsymbol{U}_{\text{IPWCP,dr}} = \sum_{i=1}^{N} \sum_{j<k<n_i} \left\{ \frac{R'_{ijk}}{\pi'_{ijk}} \cdot \boldsymbol{U}(y_{ij}, y_{ik}) + \left(1 - \frac{R'_{ijk}}{\pi'_{ijk}}\right) \cdot E_{\boldsymbol{Y}_i^m | \boldsymbol{y}_i^o} \boldsymbol{U}(y_{ij}, y_{ik}) \right\}$$

$$\begin{aligned} \boldsymbol{U}_{\text{IPWAC,dr}} = \sum_{i=1}^{N} \sum_{j<k} & \left[\frac{R_{ij}}{\pi_{ij}} \cdot U(y_{ij}) + \frac{R_{ik}}{\pi_{ik}} \cdot U(y_{ik}|y_{ij}) \right. \\ & \left. + \left(1 - \frac{R_{ij}}{\pi_{ij}}\right) \cdot E_{\boldsymbol{Y}_i^m | \boldsymbol{y}_i^o} \boldsymbol{U}(y_{ij}) + \left(1 - \frac{R'_{ik}}{\pi'_{ik}}\right) \cdot E_{\boldsymbol{Y}_i^m | \boldsymbol{y}_i^o} \boldsymbol{U}(y_{ik}|y_{ij}) \right] \end{aligned}$$

$$= \sum_{i=1}^{N} \left\{ \sum_{j<k<d_i} \boldsymbol{U}_i(y_{ij}, y_{ik}) + \sum_{j=1}^{d_i-1} (n_i - d_i + 1) \cdot \boldsymbol{U}_i(y_{ij}) \right.$$

$$\left. \sum_{j<d_i\leq k} E[\boldsymbol{U}_i(y_{ik}|y_{ij})] + \sum_{d_i\neq j<k} E[\boldsymbol{U}_i(y_{ij}, y_{ik})] \right\}$$

2.18 Exchangeable Version

$$\boldsymbol{U}_{\text{IPW, exch}} = \boldsymbol{U}_{\text{naive, AC}} = \sum_{i=1}^{N} \left[\sum_{j<k<d_i} \boldsymbol{U}(y_{ij}, y_{ik}) + \sum_{j=1}^{d_i-1} (n_i - d_i + 1) U(y_{ij}) \right]$$

2.19 The Onychomycosis Trial

- **T**oenail **D**ermatophyte **O**nychomycosis: Common toenail infection, difficult to treat, affecting more than 2% of population

- Treatments A (new) and B (standard): 2×189 patients randomized, 36 centers

- Measurements at months: treatment: 0, 1, 2, 3, follow up: 6, 9, 12

- Are both treatments equally effective for the treatment of TDO?

- Complication: Dropout (24%)

2.20 Model for Toenail Data

- Consider model:

$$
\begin{aligned}
Y_{ij}|b_i &\sim N[b_i + \beta_0 \cdot I(T_i = 0) + \beta_1 \cdot I(T_i = 1)\beta_2 t_j \cdot I(T_i = 0) + \beta_3 t_j \cdot I(T_i = 1), \sigma^2] \\
b_i &\sim N(0, \tau^2)
\end{aligned}
$$

- Four forms of pseudo-likelihood:
  - ▷ full likelihood
  - ▷ complete cases
  - ▷ complete pairs
  - ▷ available cases

- Three ways of dealing with missingness:
  - ▷ naive
  - ▷ singly robust
  - ▷ doubly robust

Effect	Par.	$\boldsymbol{U}_{\text{full.lik.}}$	$\boldsymbol{U}_{\text{naive, CC}}$	$\boldsymbol{U}_{\text{naive, CP}}$	$\boldsymbol{U}_{\text{naive, AC}}$
Int.A	β_0	2.52(0.247;0.228)	2.77(0.086;0.272)	2.70(0.081;0.248)	2.56(0.075;0.231)
Int.B	β_1	2.77(0.243;0.249)	2.82(0.083;0.271)	2.81(0.078;0.254)	2.77(0.073;0.250)
Sl.A	β_2	0.56(0.023;0.045)	0.55(0.011;0.046)	0.56(0.011;0.045)	0.57(0.011;0.045)
Sl.B	β_3	0.61(0.022;0.043)	0.60(0.011;0.044)	0.61(0.011;0.043)	0.61(0.010;0.043)
R.I.v.	τ^2	6.49(0.628;0.633)	6.71(0.226;0.731)	6.67(0.213;0.680)	6.41(0.200;0.645)
Res.v.	σ^2	6.94(0.248;0.466)	7.31(0.150;0.520)	7.13(0.140;0.483)	7.05(0.137;0.472)
Effect	Par.	$\boldsymbol{U}_{\text{wt.lik.}}$	$\boldsymbol{U}_{\text{IPWCC}}$	$\boldsymbol{U}_{\text{IPWCP}}$	$\boldsymbol{U}_{\text{IPWAC}}$
Int.A	β_0	1.85(0.092;0.303)	2.71(0.074;0.266)	2.77(0.079;0.270)	2.59(0.069;0.237)
Int.B	β_1	2.65(0.089;0.517)	2.78(0.073;0.265)	2.82(0.077;0.269)	2.77(0.069;0.249)
Sl.A	β_2	0.68(0.014;0.068)	0.54(0.010;0.046)	0.53(0.010;0.044)	0.55(0.010;0.045)
Sl.B	β_3	0.73(0.013;0.101)	0.60(0.010;0.044)	0.59(0.010;0.044)	0.60(0.010;0.043)
R.I.v.	τ^2	6.21(0.235;1.032)	6.66(0.195;0.717)	6.72(0.209;0.753)	6.44(0.187;0.669)
Res.v.	σ^2	5.05(0.088;0.603)	7.29(0.130;0.513)	7.59(0.142;0.562)	7.35(0.130;0.514)
Effect	Par.			$\boldsymbol{U}_{\text{IPWCC,dr}} = \boldsymbol{U}_{\text{IPWCP,dr}} = \boldsymbol{U}_{\text{IPWAC,dr}}$	
Int.A	β_0			2.52(0.074;0.226)	
Int.B	β_1			2.77(0.072;0.247)	
Sl.A	β_2			0.56(0.011;0.046)	
Sl.B	β_3			0.61(0.011;0.044)	
R.I.v.	τ^2			6.23(0.197;0.636)	
Res.v.	σ^2			7.09(0.139;0.483)	

2.21 Considerations

- Computationally simple ways for dealing with expectation:
  - ▷ Paik's method
  - ▷ multiple imputation
- Careful evaluation of performance:
  - ▷ Compared to full likelihood
  - ▷ Compared to GEE
- **Also done:**
  - ▷ **Marginal model for binary data**
  - ▷ **Conditionally specified model**
- Other settings: survival data,. . .
- Appealing feature: simplicity under (near) exchangeability

2.22 Overview and (No Longer Premature) Conclusion

MCAR/simple	CC LOCF single imputation	biased inefficient not simpler than MAR methods
MAR	direct lik./Bayes IPW/d.r. multiple imputation	easy to conduct Gaussian & non-Gaussian
MNAR	variety of methods	strong, untestable assumptions most useful in sensitivity analysis