The Future of Missing Data

Herbert Thijs

Biometrics on the lake (Taupo) 2009

General Settings

- Pain clinical trials
- Diary Data
- Side effects
- Dropout due to side effects = treatment failure
- Justification for BOCF (?)
- **Issue:** What is the **true treatment effect**?

The problem

What is the big problem with missing data?

The problem

What is the big problem with missing data?

They are missing!!!!

Analyzing Incomplete Data

- **Frequently used methods:**
  - Complete Case Analysis
  - Last Observation Carried Forward
  - Baseline Observation Carried Forward
  - Worst Observation Carried Forward
- **More recent methods:**
  - Direct likelihood (Ignoring missing data)
  - Weighted GEE
  - Multiple Imputation

Analyzing Incomplete Data

- **Modeling frameworks:**
  - Selection models

$$f(Y, D|X, \theta, \psi) = f(D|Y, X, \theta, \psi) f(Y|X, \theta, \psi)$$

  - Pattern-mixture models

$$f(Y, D|X, \theta, \psi) = f(Y|D, X, \theta, \psi) f(D|X, \theta, \psi)$$

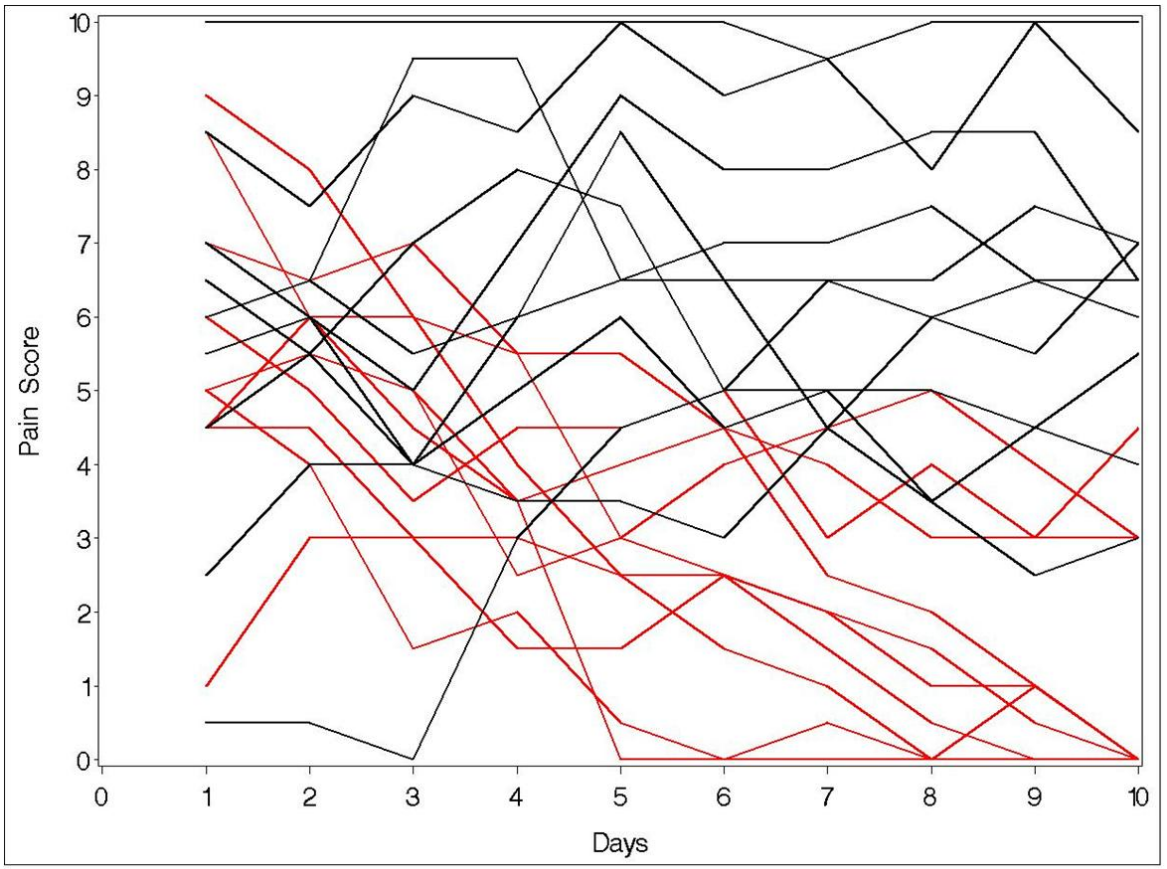
  - Shared parameter models

$$f(Y, |X, \theta, \psi, b_i) \qquad f(D|X, \theta, \psi, b_i)$$

independent given b_i

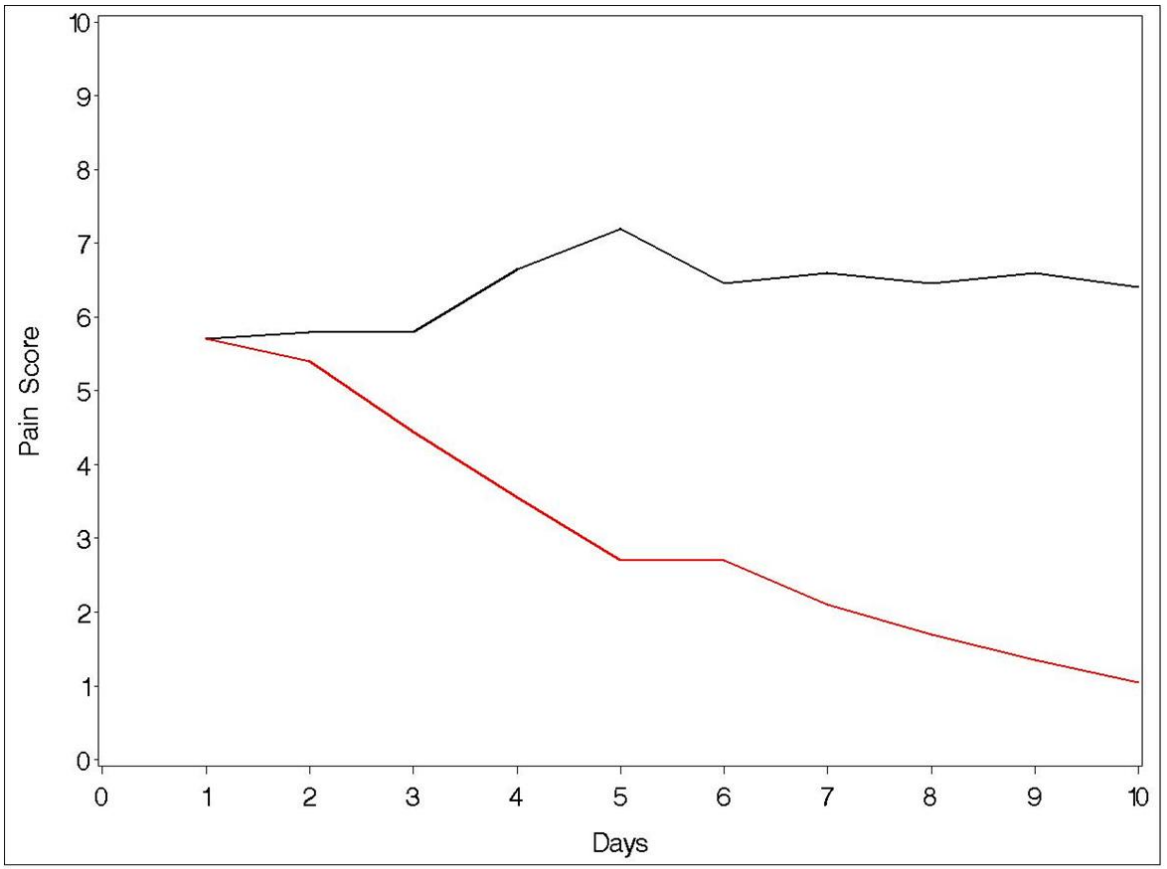
Simulation Settings

Individual pain profiles of some patients



Simulation Settings

Mean pain profile for treatment and control group



Simulation Results

Sample Size	Corr	Effect	Complete Data	Direct Likelihood	LOCF	BOCF
50	0.1	Intercept	6.49(0.23)	6.42(0.32)	6.18(0.30)	5.64(0.26)
50	0.1	DAYS	-0.52(0.04)	-0.54(0.08)	-0.44(0.05)	-0.09(0.04)
50	0.1	TREAT (Control)	0.40(0.35)	0.29(0.44)	0.65(0.45)	0.97(0.43)
50	0.1	DAYS*TREAT (Control)	0.42(0.06)	0.46(0.09)	0.23(0.06)	0.06(0.06)
50	0.9	Intercept	6.67(0.34)	6.77(0.35)	6.56(0.34)	6.24(0.33)
50	0.9	DAYS	-0.52(0.04)	-0.58(0.06)	-0.40(0.05)	-0.21(0.04)
50	0.9	TREAT (Control)	0.19(0.51)	0.10(0.51)	0.29(0.50)	0.56(0.49)
50	0.9	DAYS*TREAT (Control)	0.42(0.06)	0.47(0.08)	0.30(0.06)	0.15(0.06)
500	0.1	Intercept	6.50(0.08)	6.42(0.09)	6.21(0.09)	5.65(0.09)
500	0.1	DAYS	-0.52(0.01)	-0.54(0.02)	-0.44(0.01)	-0.09(0.01)
500	0.1	TREAT (Control)	0.36(0.11)	0.25(0.12)	0.59(0.13)	0.92(0.12)
500	0.1	DAYS*TREAT (Control)	0.42(0.02)	0.45(0.03)	0.24(0.02)	0.06(0.02)
500	0.9	Intercept	6.63(0.12)	6.74(0.12)	6.53(0.12)	6.20(0.12)
500	0.9	DAYS	-0.52(0.01)	-0.58(0.02)	-0.41(0.01)	-0.21(0.02)
500	0.9	TREAT (Control)	0.22(0.13)	0.13(0.13)	0.32(0.13)	0.59(0.13)
500	0.9	DAYS*TREAT (Control)	0.42(0.02)	0.47(0.02)	0.31(0.02)	0.16(0.02)

Some Facts

Dropouts $t_{ij} = 0$
Probability p_0
Treatment indicator $T_i = 0, 1$
$E(Y_{ij}) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$

Completers $t_{ij} = 0, 1$
Probability $p_1 = 1 - p_0$
Treatment indicator $T_i = 0, 1$
$E(Y_{ij}) = \gamma_0 + \gamma_1 T_i + \gamma_2 t_{ij} + \gamma_3 T_i t_{ij}$

	complete Case	LOCF
MCAR	0	$(p_1 - p_0)\beta_2 - (1 - p_1)\beta_3$
MAR	$-\sigma[(1 - p_1)(\beta_0 + \beta_1 - \gamma_0 - \gamma 1)$ $- (1 - p_0)(\beta_0 - \gamma 0)]$	$p_1(\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3) + (1 - p_1)(\beta_0 + \beta_1)$ $-p_1(\gamma_0 + \gamma_2) - (1 - p_0)\beta_0 - \gamma_1 - \gamma_3$ $-\sigma[(1 - p_1)(\beta_0 + \beta_1 - \gamma_0 - \gamma_1) - (1 - p_0)(\beta_0 - \gamma_0)]$

(Molenberghs et al.)

Some Facts

- Both CC and LOCF yield biased estimates

  (Shown theoretically)

- Conservative or liberal?

  No general statement

New Carrying Forward Strategy

- Previous .OCF methods not valid
- NOCF: a new method
- Really easy to conduct
- **VALID under ALL assumptions**

New Carrying Forward Strategy

- Previous .OCF methods not valid
- NOCF: a new method
- Really easy to conduct
- **VALID under ALL assumptions**

- **No** Observations Carried Forward

Conclusion

- NOCF: No Observations Carried Forward
- PAIN trials
  - Side effects
  - Treatment failure
  - BOCF $\rightarrow$ **WRONG!!**
- Estimate **true treatment effect** using direct likelihood
- Joint model for time to treatment failure and efficacy

(Dimitris Rizopoulos: afternoon 14.00h)

Guidelines for the future

- Clinical trials
- Regulatory authorities
- **HELP!**
- What do we need to do
- Some simple methods (valid under MAR)
  - Direct likelihood (always good starting point)
  - Multiple imputation
  - Weighted GEE
- Sensitivity Analysis (reasonable)
- **BE CAREFUL!**