

PROC MCMCによるShared Parameter Modelの実装

宮岡 悅良／東京理科大学

稲葉 洋介／新日本科学PPD

Implementing Shared Parameter Model using PROC MCMC

Etsuo Miyaoka / Tokyo University of Technology

Yosuke Inaba / PPD-SNBL



要旨：

PROC MCMCによりShared Parameter Modelを実装し、
実際のデータに適用した。

キーワード : missing data, Shared Parameter Model,
PROC MCMC, SPM

アジェンダ

- SHARED PARAMETER MODEL(SPM)の概要
- PROC MCMCについて
- 実装
- シミュレーション
- 感度分析

SHARED PARAMETER MODEL(SPM)の概要

欠測過程の分類

MCAR	probability of missing is independent of both observed and unobserved data
MAR	probability of missing is independent of only the unobserved data
NMAR	probability of missingness depends on the unobserved data

欠測過程の分類

- “It is well known that naive methods may result in biased inferences under an NMAR mechanism.”
- “The use of shared (random) parameter models has been one approach to accounting for non-random missingness. ”

(Albert et al)

SPMとは

- a model for the longitudinal response measurements is linked with a model for the missing-data mechanism through a set of **random effects** that are shared between the two process. (Albert et al)
- They are special case of “Selection model”(Little)

歴史

- 1988···Wu and Carroll
 - “Informative missing data”
- 1995···Follmann and Wu
 - Shared Parameter Modelを提案
- 2004···Guo and Carlin
 - “Joint Model”(Survival analysis + longitudinal analysis)
 - SAS(NLMIXED)とWINBUGS(MCMC)による実装

定義

$\mathbf{Y}_i = {}^t(Y_{i1}, Y_{i2}, \dots, Y_{iJ})$: vector of longitudinal outcomes for the i th subject
 $(i = 1, 2, \dots, I)$

observed on J occasions t_1, t_2, \dots, t_J

$\mathbf{R}_i = {}^t(R_{i1}, R_{i2}, \dots, R_{iJ})$: vector of missing-data status
 $(R_{ij} = 0 : \text{unobserved}, R_{ij} = 1 : \text{observed})$

$\mathbf{b}_i = {}^t(b_{i1}, b_{i2}, \dots, b_{iL})$: L-vector of random effects
(shared between outcomes and missing-data mechanism)

The joint distribution of $\mathbf{Y}_i, \mathbf{R}_i, \mathbf{b}_i$ can be written as

$$f(\mathbf{Y}_i, \mathbf{R}_i, \mathbf{b}_i) = g(\mathbf{Y}_i | \mathbf{b}_i, \mathbf{R}_i)m(\mathbf{R}_i | \mathbf{b}_i)h(\mathbf{b}_i)$$

We make assumption that, conditional on the random effects, the response do not depend on the missing-data status.

$$g(\mathbf{Y}_i | \mathbf{b}_i, \mathbf{R}_i) = g(\mathbf{Y}_i | \mathbf{b}_i)$$

※本発表はmonotone missingのみ対象

観測データの密度関数

\mathbf{Y}_i^o : observed data, \mathbf{Y}_i^m : missing data

$$\begin{aligned} f(\mathbf{Y}_i^o, \mathbf{R}_i) &= \int_{\mathbf{Y}_i^m} \int_{\mathbf{b}} f(\mathbf{Y}_i^o, \mathbf{Y}_i^m, \mathbf{R}_i) \\ &= \int_{\mathbf{Y}_i^m} \int_{\mathbf{b}} g(\mathbf{Y}_i^o \mid \mathbf{b}_i) g(\mathbf{Y}_i^m \mid \mathbf{b}_i) m(\mathbf{R}_i \mid \mathbf{b}_i) h(\mathbf{b}_i) d\mathbf{b} d\mathbf{Y}_i^m \\ &= \int_{\mathbf{b}} g(\mathbf{Y}_i^o \mid \mathbf{b}_i) m(\mathbf{R}_i \mid \mathbf{b}_i) h(\mathbf{b}_i) \left\{ \int_{\mathbf{Y}_i^m} g(\mathbf{Y}_i^m \mid \mathbf{b}_i) d\mathbf{Y}_i^m \right\} d\mathbf{b} \\ &= \int_{\mathbf{b}} g(\mathbf{Y}_i^o \mid \mathbf{b}_i) m(\mathbf{R}_i \mid \mathbf{b}_i) h(\mathbf{b}_i) d\mathbf{b} \end{aligned}$$

他のモデルとの比較

- Selection Model(SM)

$$f(\mathbf{Y}_i, \mathbf{R}_i) = g(\mathbf{Y}_i^o, \mathbf{Y}_i^m)m(\mathbf{R}_i \mid \mathbf{Y}_i^o, \mathbf{Y}_i^m)$$

- Pattern Mixtre Model(PMM)

$$f(\mathbf{Y}_i, \mathbf{R}_i) = g(\mathbf{Y}_i^o, \mathbf{Y}_i^m \mid \mathbf{R}_i)m(\mathbf{R}_i)$$

他のモデルとの比較

- SM、PMM
 - 尤度関数が欠測したデータに依存
- SPM
 - 尤度関数が欠測したデータに依存しない

→欠測データに対する感度分析は不要
(感度分析自体は必要)

PROC MCMCについて

SPMのパラメータ推定方法

- DIAが公開しているマクロ(PROC NLMIXED)
- PROC GLIMMIX
- PROC MCMC

PROC MCMCとは

- “The MCMC procedure is a general purpose Markov chain Monte Carlo (MCMC) simulation procedure that is designed to fit **Bayesian models.**”

(SAS Reference)

PROC MCMCによるSPMの実装

```
proc mcmc data=XXXXX nmc=XXXXXX seed=XXX  
  outpost=RESULT;  
  
parms <パラメータの宣言>;  
prior <パラメータの事前分布>;  
random <変量効果の宣言>;  
llike = <定量的解析部分の対数尤度関数>;  
slike = <欠測過程の対数尤度関数>;  
model general(llike + slike);  
  
run;
```

PROC MCMCによる実装の PROS/CONS

- PROS
 - 各パラメータに事前分布を設定できる
 - 変量効果の分布を正規分布以外に設定できる
 - パラメータの漸近正規性の仮定が不要
- CONS
 - 実行時間が長い
 - (当然)検定は無い
 - 臨床試験で主解析とするには難しい(?)
 - 臨床試験でのベイズ指標の取扱いはKawasaki et al参照

DIAマクロのPROS/CONS

- PROS
 - 出力される図表が多い
 - コーディングが楽
 - 実行時間が速い
- CONS
 - マクロのコードが古いので読み辛い
 - 設定の自由度が狭い
 - 欠測過程のモデル
 - 変量効果の指定
 - 感度分析が難しい

シミュレーション

使用したデータ

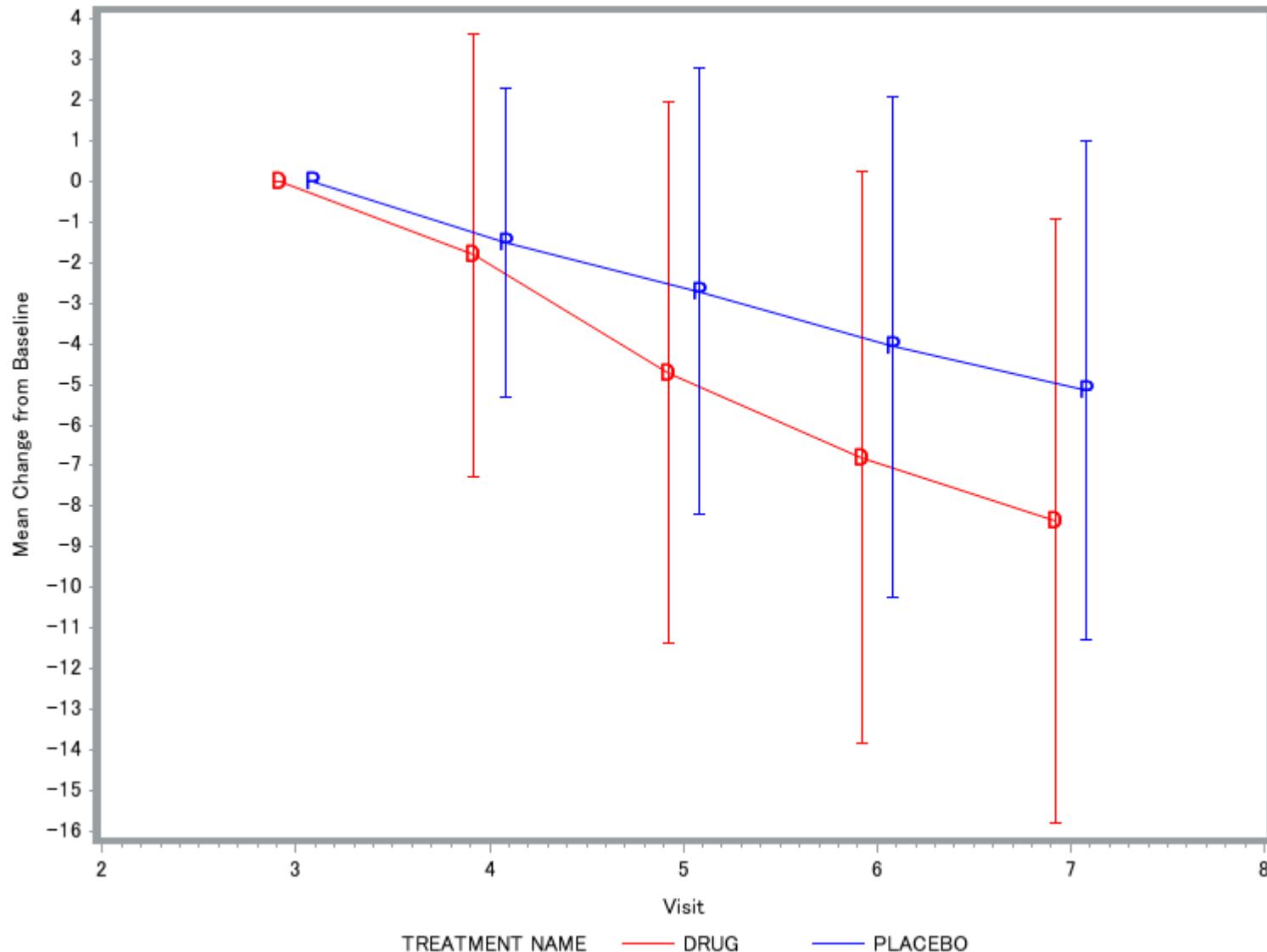
- 二重盲検プラセボ対照試験
 - 対象疾患:うつ病
 - 主要評価項目:HAM-Dスコア(低下で改善)
 - 1群100例
 - monotone missingのみ
 - <http://missingdata.lshtm.ac.uk/>より入手

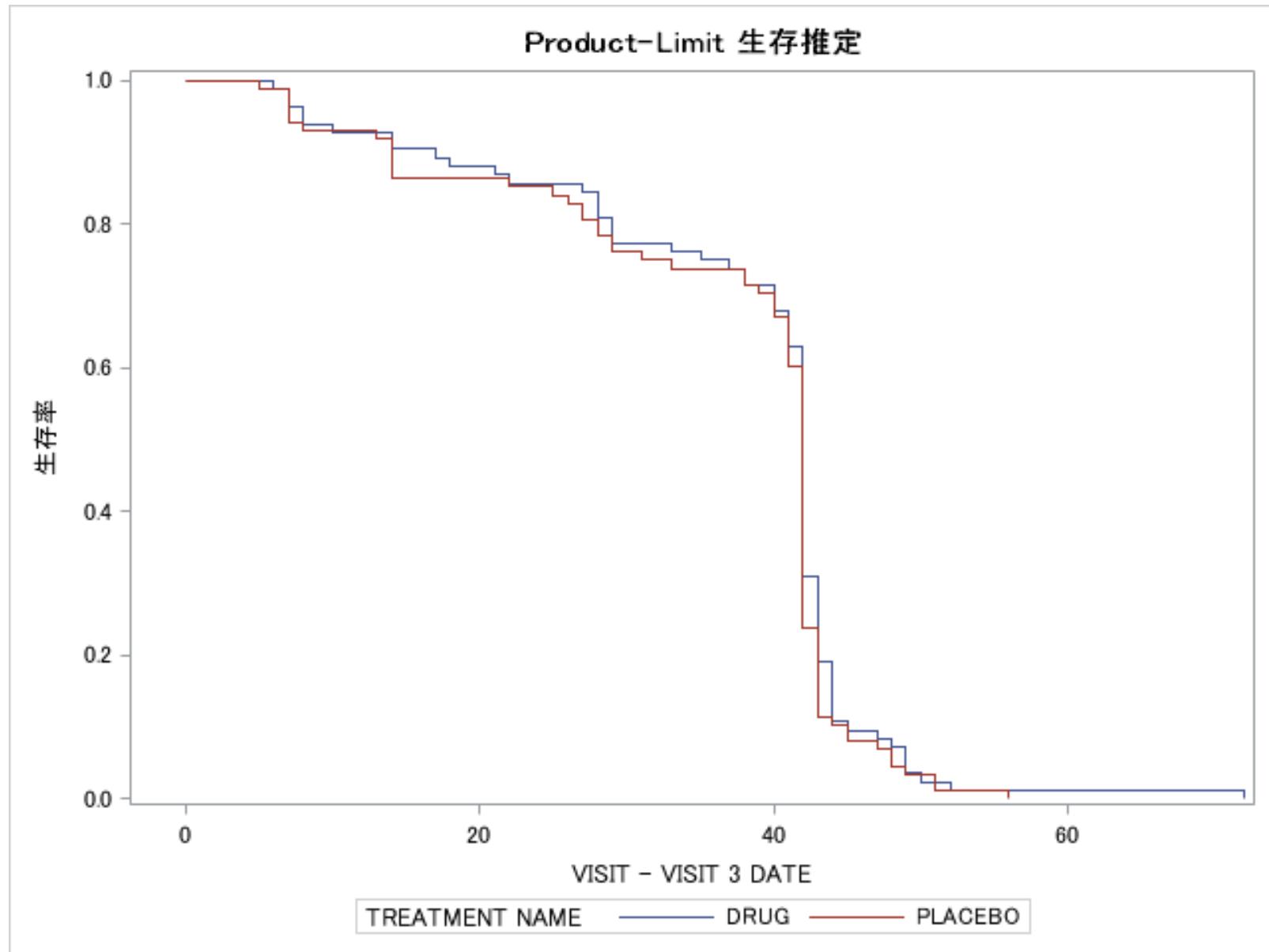
データセット

VIEWTABLE: Datalib.Chapter15_example

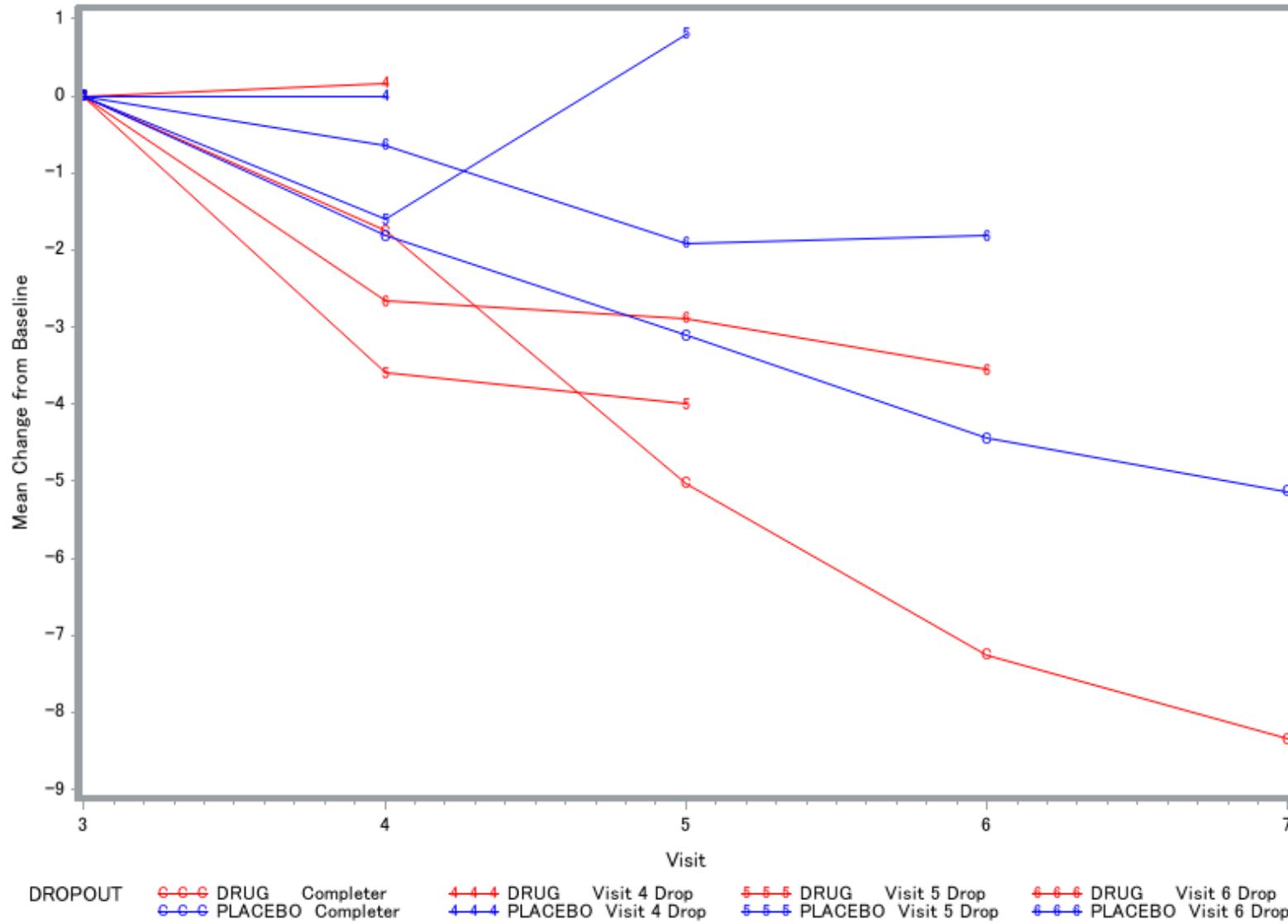
	PATIENT NUMBER	HAMA TOTAL	PGI IMPROVEMENT	VISIT - VISIT 3 DATE	VISIT NUMBER	TREATMENT NAME	PATIENT SEX	POOLED INVESTIGATOR	basval	HAMD17 TOTAL	change
1	1503	21	2		7	4 DRUG	F	006		32	-11
2	1503	19	2		14	5 DRUG	F	006		32	-12
3	1503	21	3		28	6 DRUG	F	006		32	-13
4	1503	17	4		42	7 DRUG	F	006		32	-15
5	1507	18	3		7	4 PLACEBO	F	006		14	-3
6	1507	18	2		15	5 PLACEBO	F	006		14	0
7	1507	14	3		29	6 PLACEBO	F	006		14	-5
8	1507	8	2		42	7 PLACEBO	F	006		14	-9
9	1509	18	3		7	4 DRUG	F	006		21	-1
10	1509	17	3		14	5 DRUG	F	006		21	-3
11	1509	12	3		28	6 DRUG	F	006		21	-5
12	1509	9	3		44	7 DRUG	F	006		21	-8
13	1511	15	3		5	4 PLACEBO	F	006		21	-5
14	1511	17	3		12	5 PLACEBO	F	006		21	-3
15	1511	12	3		28	6 PLACEBO	F	006		21	-3
16	1511	9	2		41	7 PLACEBO	F	006		21	-9
17	1513	25	6		7	4 DRUG	M	006		19	5
18	1514	25	5		7	4 PLACEBO	F	006		21	2
19	1516	15	3		7	4 PLACEBO	M	006		23	-8
20	1516	15	4		15	5 PLACEBO	M	006		23	-2
21	1516	13	4		28	6 PLACEBO	M	006		23	-3
22	1516	8	4		44	7 PLACEBO	M	006		23	-8
23	1517	26	6		8	4 DRUG	F	006		19	5
24	1521	4	3		9	4 DRUG	M	006		13	-9
25	1521	5	4		16	5 DRUG	M	006		13	-8
26	1521	2	4		30	6 DRUG	M	006		13	-10
27	1521	4	3		44	7 DRUG	M	006		13	-10
28	1526	7	3		6	4 PLACEBO	F	006		20	-5
29	1526	6	3		13	5 PLACEBO	F	006		20	-10
30	1526	8	5		27	6 PLACEBO	F	006		20	-8
31	1526	10	4		42	7 PLACEBO	F	006		20	-3
32	1802	10	3		7	4 PLACEBO	F	009		10	7
33	1802	12	3		16	5 PLACEBO	F	009		10	7

Mean Change with Standard Error Bars From Baseline at Each Visit by Therapy





Visitwise Mean Change From Baseline for Completers and Dropouts at Each Visit by Therapy



モデル

- HAM-Dスコア変化量のモデル

$$Y_i = \beta_0 + \beta_1 * \text{DRUG} + \beta_2 * \sqrt{\text{WEEK}} + \beta_3 * \text{DRUG} * \sqrt{\text{WEEK}} + b_0 + b_1 * \sqrt{\text{WEEK}} + \epsilon_{ij}$$

$$(b_0, b_1) \sim N(\mathbf{0}, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

$$\epsilon_{ij} \sim N(0, s)$$

- 欠測過程のモデル

$$z_i = a_1 * \text{DRUG} + a_{\text{Int}} * b_0 + a_{\text{Slp}} * b_1$$

$$Pr(R_{ij} = 1 \mid R_{i1}, \dots, R_{i(j-1)} = 0) = 1 - exp(-exp(a_{0j} + z_i))$$

実行結果

Posterior Summaries							
	Parameter	N	Mean	Standard Deviation	Percentiles		
					25%	50%	75%
Random Effect	σ_{11}	50000	14.8047	4.9042	11.1694	14.3941	17.8885
	σ_{12}	50000	-3.3124	2.7596	-5.0382	-3.0600	-1.3039
	σ_{22}	50000	9.3590	2.0179	7.9265	9.1521	10.5766
Longitudinal Mesurement	s	50000	3.0465	0.1268	2.9572	3.0454	3.1315
	b0	50000	5.4631	0.6625	5.0123	5.4342	5.9034
	bBASE	50000	-0.2876	0.0373	-0.3146	-0.2857	-0.2608
	bDRUG	50000	0.8032	0.6799	0.3465	0.8197	1.2680
	bDRUGSWK	50000	-1.7798	0.4776	-2.1065	-1.7819	-1.4611
	bSWK	50000	-1.9797	0.3780	-2.2253	-1.9787	-1.7399
	I1	50000	-2.4393	0.2586	-2.6100	-2.4301	-2.2617
Missing-data Mechanism (C-loglog link)	I2	50000	-1.8641	0.2048	-2.0006	-1.8597	-1.7217
	I3	50000	-1.6606	0.1895	-1.7856	-1.6584	-1.5291
	I4	50000	-1.1350	0.1621	-1.2423	-1.1317	-1.0260
	I5	50000	-0.9298	0.1530	-1.0299	-0.9275	-0.8259
	I6	50000	1.0021	0.1447	0.9058	0.9982	1.0929
	I7	50000	1.7835	0.2353	1.6215	1.7683	1.9320
	aDRUG	50000	-0.0764	0.1782	-0.2007	-0.0767	0.0440
Endpoint	aInt	50000	0.0382	0.0467	0.0111	0.0390	0.0666
	aSlp	50000	0.0899	0.0495	0.0577	0.0868	0.1189
	trt_diff(*)	50000	-4.1804	1.1615	-4.9664	-4.1923	-3.4079

(*) trt_diff = bDRUG+bDRUGSWK*2.8

余談

- PROC MCMCでの欠損値のハンドリング
- 以下、SAS Referenceより抜粋
 - **MISS=keyword** specifies how missing values are handled (see the section [Handling of Missing Data](#) for more details). The default is **MISSING=COMPLETECASE**.
 - **ALLCASE | AC** gives you the option to model the missing values in an all-case analysis. You can use any techniques that you see fit, for example, **fully Bayesian** or **multiple imputation**.
 - **COMPLETECASE | CC** assumes a complete case analysis, so all observations with missing variable values are discarded prior to the simulation.

(MISS=ACと指定すれば自動で補完してくれる)

感度分析

SPMにおける感度分析

- 欠測過程のモデル選択
- 変量効果の設定
 - Copulaの利用… see Rizopoulos D et al

結果

	cLogLog		Logit		Weibull	
	Mean	Sd	Mean	Sd	Mean	Sd
Trt_diff	-4.1804	1.1615	-3.9554	1.0662	-3.9934	1.1013

まとめ

- PROC MCMCによりSPMを実装した。
- missing-data mechanismを変更して、エンドポイントの推定値の頑健性を確認する方法を示した。
- DIAマクロは便利だが、それだけでは検討に不十分さが残る。
- コピュラを用いた感度分析は今後の課題

参考文献

- 高橋文博, 藤原正和, 大浦智紀, 横山雄一, 欠測のあるデータに対する各種解析手法と欠測メカニズムに対する感度分析 (3) 解析手法の解説2, SASユーザー総会論文集2014.
- 高橋文博, 横溝孝明, 駒寄弘, 欠測のあるデータに対する各種解析手法と欠測メカニズムに対する感度分析 (4) 欠測メカニズムに対する感度分析, SASユーザー総会論文集2014.
- 土居正明, 欠測のあるデータに対する総合的な感度分析と主解析の選択, SASユーザー総会論文集2014

参考文献

- Roderick Little "Selection and pattern-mixture Models", Ch18, Longitudinal Data Analysis: A Handbook of Modern Statistical Methods, Chapman & Hall/CRC Press
- Albert P. S. & Follmann, D.A. "Shared Random Effects Models", Ch19, Longitudinal Data Analysis: A Handbook of Modern Statistical Methods, Chapman & Hall/CRC Press
- Wu M.C. & Carroll R.J, "Estimation and Comparison of Changes in the Presence of Informative Right Censoring: Modeling the Censoring Process," Biometrics, 1988; 44,175-188

参考文献

- Dean Follmann and Margaret Wu, "An Approximate Generalized Linear Model with Random Effects for Informative Missing Data", *Biometrics*, 1995: Vol. 51, No. 1, 151-168
- Xu Guo and Bradley P. CARLIN, "Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages", *The American Statistician*, 2004: vol. 58, 16-24
- Y.Kawasaki and E.Miyaoka, "A Bayesian inference of $P(\lambda_1 < \lambda_2)$ for two Poisson parameters", *Journal of Applied Statistics*, 2012, 39:10, 2131-2152

参考文献

- Fang Chen, "The RANDOM Statement and More: Moving On with PROC MCMC", SAS GLOBAL FORUM, 2011: Paper 334
- Fang Chen, "Missing No More: Using the MCMC Procedure to Model Missing Data", SAS GLOBAL FORUM, 2013: Paper 436
- Rizopoulos D and G Verbeke, "Shared Parameter Models under Random-effect Misspecification", IAP STATISTICS NETWORK