

多重補完法におけるPattern-Mixtureモデルに基づく感度分析

伊藤 陽一¹ 西本 尚樹²

1 北海道大学大学院医学研究科医学統計学分野

2 北海道科学大学保健医療学部

Sensitivity analysis with the pattern-mixture model approach to multiple imputation

Yoichi M. Ito¹, Naoki Nishimoto²

1 Department of Biostatistics, University of Hokkaido Graduate School of Medicine

2 Department of Radiological Technology, Faculty of Health Sciences, Hokkaido University of Science

要旨

近年、臨床試験の主要評価項目の解析において、欠測値を考慮した解析が推奨されている。多重補完法(multiple imputation methods)は、欠測値を補完するモデルを用いて欠測値の補完を行い、欠測値が埋められた完全データを複数作成し、完全データに対して予定された解析を行い、得られた複数の結果を併合する手法である。平均値補完など、単一の値を埋めて解析する手法と比較して、バラツキの過少評価を防ぐことができるため、欠測値を伴ったデータに対する手法として主要な手法の一つとなっている。しかし、多重補完法では、欠測の補完にあたっては、対象者が観測された値に従って、欠測が起こっていると仮定する Missing at Random(MAR)を仮定しており、対象者が欠測値そのものの値に従って欠測が起こっている(Missing Not at Random; MNAR)場合には結果にバイアスが入ることが知られている。このMNARの状況下では、バイアスのない推定値を求めることはできないため、結果変数が欠測にどの程度の影響を受けているかを調べる感度解析が行われることが多い。Pattern-Mixture モデルは、この感度解析を行うために提案されているモデルである。SAS9.4 から MI プロシジャに MNAR ステートメントが追加され、Pattern-Mixture モデルに基づいた感度解析が実行可能となった。本論文では、MI プロシジャにおける MNAR ステートメントの使用方法について解説を行う。

キーワード：多重補完法, Pattern-Mixture モデル, MI プロシジャ, MNAR ステートメント

はじめに

近年、臨床試験の主要評価項目の解析においては、欠測データが臨床試験の結果に大きく影響を与えることが指摘され、欠測値を考慮した解析が推奨されている¹⁾。米国食品医薬品局 (Food and Drug Administration; FDA) が、2008年に全米研究評議会 (National Research Council; NRC) の専門家パネルに対して依頼して作成された“The prevention and treatment of missing data in clinical trials”という報告書においても、欠測データの発生に関する仮定に対する感度解析を行うことが推奨されている²⁾。

欠測の仮定

欠測の仮定は、欠測が共変量(X)、補助変数(V)、結果変数(Y)のいずれにも依存しないMCAR (Missing Completely at Random)、観測された値のみに依存するMAR (Missing at Random)、欠測した値にも依存するMNAR (Missing Not at Random)に分類される。条件付分布の記法を用いると、MNARに関する分布がMARやMCARのときに、下記のように簡略化されることが分かる。ここで、 V_{obs} は観測された補助変数、 V_{mis} は欠測した補助変数、 Y_{obs} は観測された結果変数、 Y_{mis} は欠測した結果変数を表す。

$$\text{MNAR: } [M|X, V_{obs}, V_{mis}, Y_{obs}, Y_{mis}]$$

$$\text{MAR: } [M|X, V_{obs}, V_{mis}, Y_{obs}, Y_{mis}] = [M|X, V_{obs}, Y_{obs}]$$

$$\text{MCAR: } [M|X, V_{obs}, V_{mis}, Y_{obs}, Y_{mis}] = [M]$$

欠測値を埋める解析方法 (Single imputation, Multiple imputation)

欠測値に対して単一の値で補完する Single imputation のうち経時的な測定を伴う臨床試験で、広く用いられているのが、LOCF (Last Observation Carry Forward)法である。LOCF 法は、経時的に観察される結果変数のうち途中で打ち切りになった症例については、最終観察値をもって、試験終了時点の観察値とするものである。アルツハイマー病などの QoL Score の自然悪化を伴う疾患に対する薬剤の治療目的は、悪化をできる限り遅らせることになるが、ある薬剤がプラセボと比較して早期に治療打ち切りが起こるとすると、LOCF では、薬剤の方がプラセボよりも、結果が良くなるというバイアスが入る (図 1)³⁾。

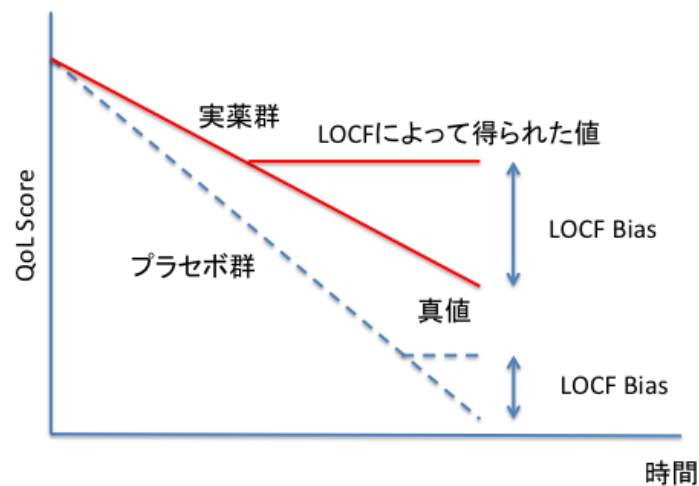


図 1 LOCF Bias の概念図

このように、LOCF は、MCAR の下であったとしても、バイアスを生じ得る。また、欠測を単一の値で補完するため、欠測に伴う不確実性について考慮できていない。したがって、LOCF は欠測に対する解析方法として、全く推奨できない。

LOCF 以外の Single imputation の方法としては、BOCF (Baseline Observation Carry Forward)法、regression imputation、hot deck imputation などがある。BOCF 法は、最終観察値ではなく、試験開始時点の値をもって、

試験終了時点の値する方法であるが、LOCF法と同様に、バイアスが入り得る。Regression imputation法は、観測されている値を用いて、欠測値を予測する回帰モデルを構築し、予測値によって欠測値を補完する方法である。Hot deck imputation法は、欠測している結果変数以外の変数が類似している症例を選択し、その症例の結果変数をもって、結果変数の欠測を補完する方法である。どちらの方法も、欠測値を予測するプロセスが妥当であれば、LOCF法やBOCF法で生じるようなバイアスは減少させることができるが、欠測を単一の値で補完することによるバラツキの過少評価が生じる。

このバラツキの過少評価を改善する方法が、Multiple imputation法である。Multiple imputation法では、欠測値の予測分布からサンプリングを行い、欠測値を補完する。補完された完全データに対して通常の解析を実行する。その操作を複数回実行し、解析結果の平均を取る。この際に、欠測を補完したことに伴うバラツキについても考慮される。Multiple imputation法の重要な利点は、最終解析に用いられない補助変数を補完モデルに用いることができる点である。多重補完法における予測分布は、MARを仮定すると、ベイズ法における事後分布となる。

感度解析の原理と方法

これまでに紹介した解析はMARを前提としてきた。主要な解析はMARを前提とした解析を行い、観測データのMARからの乖離は、MNARを前提とした解析方法による感度解析で検討されるべきだと考える。解析の枠組みとして、2つの仮定を考える。(i)欠測データの分布に関する検証不能な仮定と(ii)観測データの分布に関する検証可能な仮定である。完全データの分布は以下のように分割できる。

$$[Y_{obs}, Y_{mis}, M|X] = [Y_{obs}, M|X] \times [Y_{mis}|Y_{obs}, M, X]$$

$[Y_{mis}|Y_{obs}, M, X]$ が仮定(i)に関するもの、 $[Y_{obs}, M|X]$ が仮定(ii)に関するものである。

感度解析手法としては、Pattern Mixture Modelアプローチと Selection Modelアプローチが提案されている。SAS9.4で追加されたMIプロシジャにおけるMNARステートメントによって、Pattern Mixture Modelアプローチに基づく感度解析を行うことができるようになった。Pattern Mixture Modelの考え方は、欠測のパターンが先にあり、そのパターンごとに Y_{mis} の分布が異なるというものである。条件付き分布の記法を用いると以下のように記述できる。

Pattern Mixture Model

$$[Y_{obs}, Y_{mis}, M|X] = [Y_{obs}, Y_{mis}|M, X] \times [M|X]$$

$$[Y_{obs}, Y_{mis}, M|X] = [Y_{mis}|Y_{obs}, M, X] \times [Y_{obs}|M, X] \times [M|X]$$

例として、結果変数が1つで、補助変数がない場合について示す。

Pattern Mixture Modelの場合、観測できるかできないかで、結果変数 Y の期待値が異なる。

$$E(Y|R=0) = E(Y|R=1) + \Delta$$

$$\mu_0 = \mu_1 + \Delta$$

この期待値の違いを表す Δ が感度解析パラメータである。より一般的には、関数 $g()$ を用いて、下記のように書ける。

$$\mu_0 = g^{-1}(g(\mu_1) + \Delta)$$

完全データにおける結果変数の期待値 μ は、欠測確率を π とすると以下のようになる。

$$\mu = \pi\mu_1 + (1 - \pi)g^{-1}(g(\mu_1) + \Delta)$$

この式において、 μ_1 と π を観測データから得られた $\hat{\mu}_1$ と $\hat{\pi}$ に置き替え、 Δ ごとに μ を求めることによって、感度解析を行う。

SAS の MI プロシジャにおける Pattern-Mixture Model

SAS での簡単な解析例を示す。臨床試験において、ベースライン共変量 Y_0 と割付群 Trt (1: 実薬 0: プラセボ)には欠測がないものとする。主要評価項目 Y_1 に対して、下記の解析を行うことを考える。

$$Y_1 = \mu + \beta_1 Trt + \beta_2 Y_0$$

このとき、 β_1 が薬剤の効果を推定していることになる。臨床試験においては、薬剤に効果がないという仮定のもと解析を行うことが自然なので、 Y_1 の欠測に対しては、プラセボ群のデータのみを用いて埋めることが妥当である。これはプラセボ群のデータのみを用いて、Pattern Mixture Model における $\Delta = 0$ の下での補完を考えていくことになる。欠測データの構造としては、以下のようになる。

Obs	Trt	y0	y1
1	0	10.5212	11.3604
2	0	8.5871	8.5178
3	0	9.3274	.
4	0	9.7519	.
5	0	9.3495	9.4369
6	1	11.5192	13.2344
7	1	10.7841	.
8	1	9.7717	10.9407
9	1	10.1455	10.8279
10	1	8.2463	9.6844

このとき、下記のプログラムによって、プラセボ群のみを用いた多重補完を行うことができる。

```
proc mi data=Mono1 seed=14823 nimpute=10 out=outex15;
  class Trt;
  monotone reg (/details);
  mnar model( y1 / modelobs= (Trt='0'));
  var y0 y1;
run;
```

MNAR ステートメントでは、単調な欠測か、欠測が他の変数の条件付き分布として特定できることを前提としているので、MONTONE ステートメントか FCS ステートメントを同時に指定する必要がある。MNAR ステートメントの model オプション内の modelobs によって、どの対象者を用いて補完するかを指定することができる。ここでは $Trt = 0$ のプラセボ群のみを用いて補完するよう指定している。また、補完に用いる変数は Y_0 と Y_1 のみである。

Method として、Monotone が指定されており、Monotone ステートメントで reg オプションが指定されているので、MI プロシジャは回帰モデルに基づいて Y_1 補完を行っている。

The MI Procedure

Model Information	
Data Set	WORK.MONO1
Method	Monotone
Number of Imputations	10
Seed for random number generator	14823

Monotone Model Specification	
Method	Imputed Variables
Regression	y1

Missing data pattern テーブルでは、それぞれの missing pattern の頻度が集計される。下記の例では、25%の欠測が生じていることが分かる。

Missing Data Patterns						
Group	y0	y1	Freq	Percent	Group Means	
					y0	y1
1	X	X	75	75.00	9.996993	10.709706
2	X	.	25	25.00	10.181488	.

MNAR ステートメントで model オプションを指定しているため、以下の出力がなされる。

Observations Used for Imputation	
Models Under MNAR Assumption	
Imputed	Observations
Variable	
y1	Trt = 0

Details オプションを指定しているため、補完に用いた回帰モデルの回帰係数が出力されている。

Regression Models for Monotone Method												
ImputedEffect	Obs-Data	Imputation										
Variable		1	2	3	4	5	6	7	8	9	10	
y1	Intercept	-0.30169	-0.174265	-0.280404	-0.275183	0.090601	-0.457480	-0.241909	-0.501351	-0.058460	-0.436650	-0.509949
y1	y0	0.69364	0.641733	0.629970	0.507776	0.752283	0.831001	0.970075	0.724584	0.623638	0.563499	0.621280

補完された結果は以下のようになり、MIANALYZE プロシジャによって解析可能なデータセットが作られる。

Obs	Imputation	Trt	y0	y1
1		1	0	10.5212
2		1	0	8.5871
3		1	0	9.3274
4		1	0	9.7519
5		1	0	9.3495
6		1	1	11.5192
7		1	1	10.7841
8		1	1	9.7717
9		1	1	10.1455
10		1	1	8.2463

$\Delta \neq 0$ の下での補完例については、講演にて紹介する。

参考文献

1. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med.* 2012 Oct 4;367(14):1355-60. PubMed PMID: 23034025. Epub 2012/10/05. eng.
2. National Research Council (U.S.). Panel on Handling Missing Data in Clinical Trials., National Research Council (U.S.). Committee on National Statistics., National Academies Press (U.S.). The prevention and treatment of missing data in clinical trials. Washington, D.C.: National Academies Press; 2010. xv, 144 p. p.
3. O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clinical pharmacology and therapeutics.* 2012 Mar;91(3):550-4. PubMed PMID: 22318615.