PROC IMLなしで実装するPattern-Mixture Multiple Imputation

○ビービ 彩加[1]、室永 遼太郎[1]

([1]シミック（株）データサイエンス事業本部）

Implementation of Pattern-Mixture Multiple Imputation Without PROC IML

Ayaka Beebe, Ryotaro Muronaga
Data Science Div., CMIC Co., Ltd.

要旨

本発表では、反復測定データの欠測値に対し、欠測が初めて生じた時点とそれ以降に欠測が生じた時点で異なる仮定を置く NFMV＋CCMV 法による多重代入法を、SAS の標準プロシジャのみを用いて実装する手順を紹介する。
キーワード: 反復測定データ、欠測データ、単調脱落、多重代入法、Pattern-Mixture Model、NFMV＋CCMV

1　はじめに

反復測定データの解析では、観測期間中に欠測データが生じることは少なくない。特に臨床試験のように、被験者を複数の時点で繰り返し観測するデータでは、試験途中での脱落や被験者が来院しないなどにより、計画された時点の観測値が得られない場合がある。このような欠測データの取り扱いは解析結果に影響を及ぼす可能性があり、適切な解析手法を事前に検討しておくことが重要である。

Estimand の考え方では、まず何を推定したいのかを明確にしたうえで、欠測データをどのように扱うかについての仮定を整理することが重要である。主解析で置いた仮定が結果にどの程度影響するかを確認するためには、欠測データについて異なる仮定を置いた感度分析を行い、結果の頑健性を評価する必要がある。したがって、欠測データの扱いに配慮した解析手法を、実務にて使える形へ整備しておくことには大きな意義がある。

特に欠測データへの対応方法の一つとして、多重代入法は広く用いられている。中でも Pattern-Mixture Model に基づく多重代入は、脱落後の未観測データに対して具体的な仮定を置くことができるため、感度分析の枠組みとして有用である。しかし、Pattern-Mixture Multiple Imputation の実装は複雑になりやすく、特に初回の欠測とそれ以降の欠測で異なる仮定を適用する場合には、データの整形、脱落パターンの識別、参照する被験者集団の設定、欠測補完後データの再構成など、複数の処理を組み合わせる必要がある。そのため、実務上では、実装の難しさが大きな負担となることがある。

SAS は臨床試験データに対する解析を含む多くの実務領域で広く利用されているが、Pattern-Mixture Multiple Imputation の実装例としては、PROC IML を前提としたものも少なくない。PROC IML は柔軟な行列演算が可能である一方、SAS/STAT には含まれておらず、SAS/IML が追加導入されている環境が必要となる。よっ

て、SAS/IML が無くとも SAS の標準プロシジャのみを用いて、再利用しやすく見通しのよい形で実装する方法を示すことは、実務上有用と考えられる。
そこで本論文では、2 群 4 時点の単調な欠測データを例に、初回の欠測には NFMV（Non Future Missing Values）、それ以降の欠測には CCMV（Complete Case Missing Values）を適用する Pattern-Mixture Multiple Imputation を、PROC IML を用いずに PROC MI、PROC MIXED、PROC MIANALYZE のみで実装する方法を示す。NFMV＋CCMV を取り上げたのは、初回欠測時点とその後の欠測について仮定を変えるためである。NFMV＋CCMV を用いることで、欠測後データに対する考え方を段階的に反映でき、感度分析としても扱いやすい。また、今回のような単調脱落データであれば、SAS の標準プロシジャのみで実装できる点も実務上の利点である。あわせて、データの並べ替え、脱落時点ごとの処理、多重補代入の解析用データの作成、MMRM の実行、および結果統合までの一連の流れを整理し、SAS の標準機能を組み合わせて、分かりやすく再利用しやすい実装とするための工夫について述べる。

2 方法

本論文では、2 群比較の反復測定データを対象として、単調な欠測がある場合の Pattern-Mixture Multiple Imputation の実装方法を示す。ここで単調な欠測とは、ある時点以降の観測値が連続して欠測となることを意味する。このような欠測パターンは、試験の途中中止や来院の中断が生じるようなデータでみられることが多い。
今回の実装では、欠測データに対する仮定として、初回の欠測には NFMV（Non Future Missing Values）、それ以降の欠測には CCMV（Complete Case Missing Values）を適用する。すなわち、脱落直後の未観測値については対象被験者より 1 時点先まで観測されている被験者を参照し、残りの欠測部分については試験を最後まで完了した被験者を参照する構成とした。上記の仮定に基づき、欠測後のデータの扱いを段階的に設定した Pattern-Mixture 型の多重代入を実装する。
なお、本手法は Estimand 自体を定義するものではなく、脱落後の未観測データにどのような仮定を置くかを具体化する方法である。本論文では、その仮定を SAS の標準プロシジャのみで実装し、代入後に MMRM を適用して結果を統合するまでの一連の手順を示す。

2.1 想定するデータの構造

本論文で扱うデータは、各被験者について、ベースライン値、投与群および複数の評価時点における反応変数が収集された反復測定データである。被験者を i=1, …, N(=200)、投与群を $g \in \{0,1\}$、予定された評価時点を j=1, …, K(=4)とする。i 番目の被験者について、時点 j における反応変数を Y_{ij} とし、観測されている最後の評価時点を脱落時点（dropvisit）として定義する。たとえば dropvisit が $j=s$ の被験者は、時点 1〜s までの値は観測され、時点 s+1 以降の値は欠測である。一方、dropvisit が $j=K$ である被験者は、全評価時点で観測を完了した完遂例（completer）である。表 1 にて具体的なデータ構造を示す。なお、sim_long は実装例においてシミュレーションにより作成したデータである。

表 1. 解析開始時の反復測定データ（sim_long）の変数一覧

変数名	型	内容	格納値

id	numeric	被験者 ID	1〜200
trt	numeric	投与群	0: 対照群, 1: 投与群
base	numeric	ベースライン値	正規分布から生成
visit	numeric	評価時点	1〜4
y_obs	numeric	各時点の観測値	脱落後は欠測値

多重代入の実装にあたっては、long 形式データを wide 形式へ変換して用いる。実装全体の流れは 3 章および図 2 に示す。

2.2 NFMV＋CCMV に基づく多重代入アルゴリズム

Pattern-Mixture Model は、欠測パターンごとにデータを捉え、脱落後の未観測データに対して仮定を置く方法である。本論文で扱う NFMV＋CCMV による多重代入は、そのような Pattern-Mixture Model に基づく実装の一例である。

2.1 で定義した脱落時点（dropvisit）は、各被験者で観測された最後の評価時点を表す。したがって、dropvisit が s の被験者は、評価時点 1〜s までは観測され、評価時点 t=s+1 以降が欠測である。ここで s は脱落時点を表す一般の添字であり、1, ..., K−1 の値をとる。また、最初の欠測時点を t=s+1 と定義する。本論文では、各投与群にて欠測パターンごとに以下の 2 段階の代入を行う。具体的には、初回の欠測時点に対して NFMV、それ以降の欠測に対して CCMV の仮定をおく。適用関係を図 1 に示す。

図 1. 単調な欠測データに対する NFMV+CCMV の適用

例: dropvisit=2 の被験者

	Visit 1	Visit 2	Visit 3	Visit 4
対象被験者（dropvisit=2）	観測	観測	欠測	欠測
NFMV donor（dropvisit=3）	観測	観測	観測	欠測
CCMV donor（dropvisit=4）	観測	観測	観測	観測

Step 1: 初回の欠測時点に対する NFMV

最初の段階では、2.1 で定義した dropvisit を用い、dropvisit が s の被験者を target、dropvisit が t の被験者を donor とする。ここで t=s+1 である。すなわち、target と donor は同一投与群内にあり、donor は target より 1 時点先まで観測されている被験者である。この donor の情報を用いて、target の最初の欠測時点 t に値を代入する。

実装では、target と donor を同じデータセットにまとめ、ベースライン値を説明変数、時点 1 から t までの反応変数を目的変数として PROC MI の monotone regression（単調回帰法）法を適用する。ここで、変数リストを Y_1 から Y_t までに限定することにより、初回欠測時点の値のみを補完し、それ以降の欠測部分は未補完のまま残す。このとき、donor の時点 tにおける観測値は、target の時点 tの欠測値を推定するための情報として利用されるのであり、donor の観測値をそのまま target に代入するわけではない。したがって、この処理により、初回欠測時点について、より近い脱落パターンを持つ被験者の情報を反映した代入値を得ることができる。

Step 2: 残りの欠測部分に対する CCMV

Step 1 で初回欠測時点の値を補完した後、残りの欠測部分、すなわち時点 t+1 から K までの未観測値に対して、CCMV を適用する。この段階では、同一投与群内の完遂例、すなわち dropvisit が K の被験者を donor とする。

具体的には、Step 1 により初回欠測時点が補完された target と、同一群内の完遂例を結合し、ベースライン値および全時点の反応変数 $Y_1, \ldots, Y_K$ を用いて PROC MI の monotone regression 法を再度適用する。このとき、Step 1 で得られた代入データごとに Step 2 を実施することで、初回欠測に対する NFMV の結果を保持したまま、残りの欠測部分に対する CCMV を実現する。

なお、脱落時点が K−1 の被験者については、初回欠測時点が最終時点 K に一致するため、Step 1 の時点で欠測部分がすべて補完される。この場合、Step 2 は不要であり、省略する。

群別・脱落パターン別の繰り返し

上記の Step 1 および Step 2 は、各投与群に対して独立に実施する。また、dropvisit が $1, \ldots, K-1$ の各脱落パターンについて同様の処理を繰り返す。最終的に、各群・各脱落パターン・各代入回数に対して得られたデータを結合し、全被験者を含む多重代入後データセットを作成する。

このように、本手法は「同一群内で donor を選択すること」「初回欠測とその後の欠測で異なる donor 集団を用いること」「脱落パターンごとに処理を分けること」を特徴とする。これにより、欠測後データに対する仮定を反映しつつ、SAS の標準機能で実装可能な多重代入手順を構成できる。

2.3 代入後データの解析

多重代入後に得られた wide 形式データは、各時点の反応変数を縦持ちに戻し、long 形式データとして解析に用いる。各代入データに対して、ベースライン値、投与群、時点、および投与群と時点の交互作用を固定効果とする MMRM を適用する。被験者内相関については、時点を反復要因として非構造型共分散構造を仮定する。

各代入データで得られた推定値および標準誤差は、PROC MIANALYZE を用いて Rubin の方法に基づき統合する。本論文の実行例では、最終時点における群間差を対象として、各代入データで得られた推定結果を統合することで、Pattern-Mixture Multiple Imputation の下での最終的な推定値を得る。

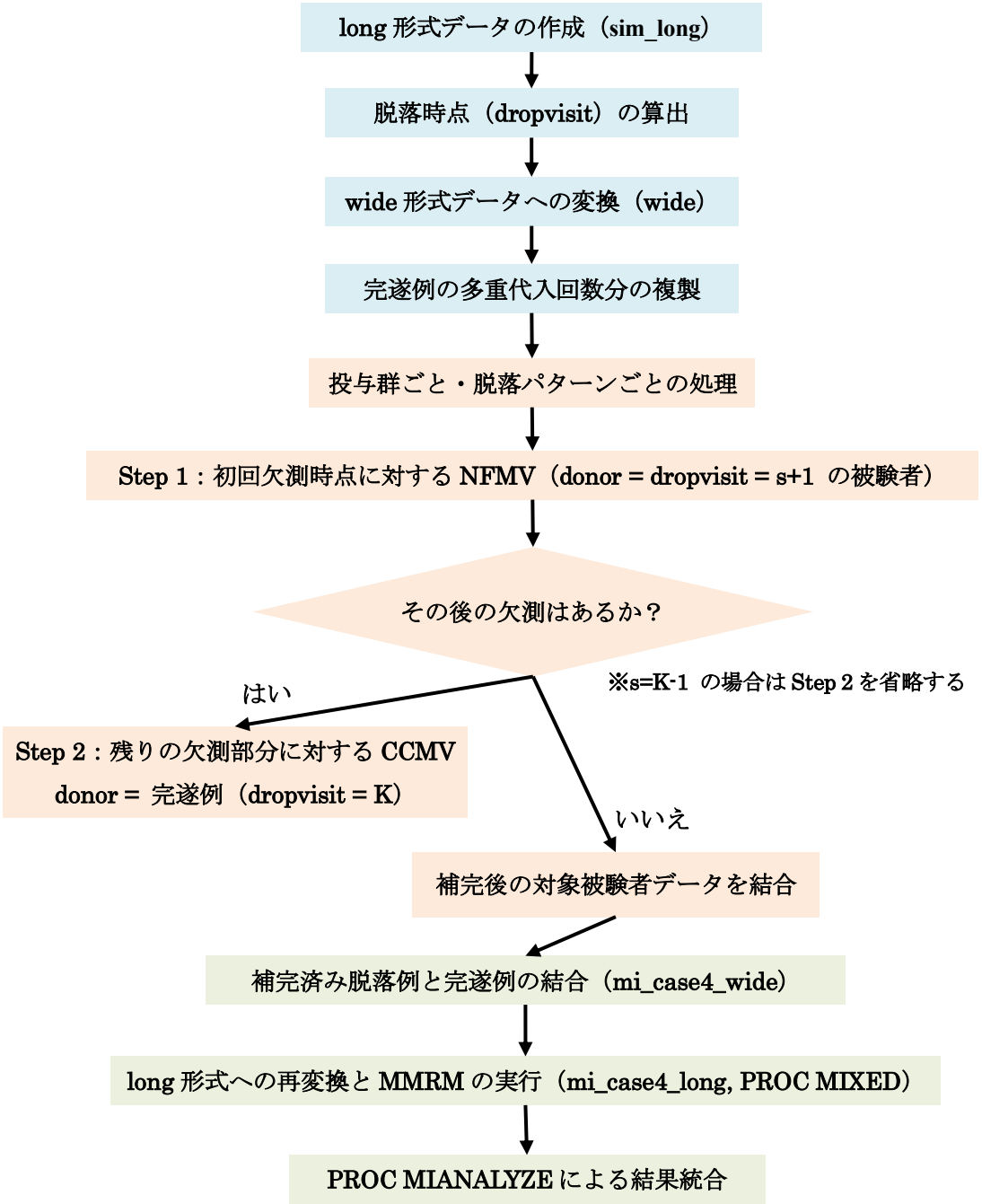
3 実装

本論文では、前節で述べた NFMV＋CCMV に基づく Pattern-Mixture Multiple Imputation を、PROC IML を用いずに SAS の標準プロシジャおよびデータステップのみで実装した。実装の全体像は、(1) long 形式データの準備、(2) 脱落時点の算出と wide 形式への変換、(3) 群別・脱落パターン別の多重代入、(4) 代入後データの再構成、(5) MMRM の実行と結果統合、の 5 段階から構成される。

本実装では、初回欠測時点に対する NFMV と、その後の欠測部分に対する CCMV を明確に分離して処理するため、投与群ごと、欠測脱落パターンごと、さらに代入回数ごとにデータを整理する構成とした。これにより、欠測後の仮定を段階的に反映した代入データを、SAS の標準機能のみで構築できるようにしている。

本実装の全体的な流れを図 2 に示す。本実装に用いた SAS コード全体を付録に示す。SAS での実装は、SAS Version 9.4 を使用した。以下では、各処理の内容を順に説明する。

図 2. PROC IML を使用しない Pattern-Mixture Multiple Imputation の SAS 実装フロー
long 形式データから dropvisit を算出し、wide 形式への変換、完遂例の複製、投与群別・脱落パターン別の NFMV および CCMV、多重代入データの結合、long 形式への再構成、MMRM、および MIANALYZE による結果統合までの流れを示す。



3.1 データ準備と脱落時点の算出

実装の起点となるのは、被験者 ID、投与群、ベースライン値、評価時点、観測値を含む long 形式データである。まず各被験者について、観測されている最終時点を dropvisit として算出する。この処理には PROC SQL を用い、各被験者について欠測でない時点の最大値を取得することで脱落時点を定義した。

次に、この dropvisit 情報を元の long 形式データに結合し、PROC TRANSPOSE により各評価時点の観測値を wide 形式に変換した。wide 形式では、各被験者につき 1 行のレコードに Y_1 から Y_K までの変数が並ぶため、後続の多重代入処理において、脱落パターン別の抽出や対象時点の制御が容易になる。また、dropvisit が K に一致する被験者については、完遂例であることを示す補助変数を付与し、CCMV における donor 抽出に利用した。

```
/* 各被験者の最終観測時点を dropvisit として定義 */
proc sql;
   create table dropinfo as
   select id, max(case when y_obs is not null then visit end) as dropvisit
   from sim_long
   group by id;
quit;
```

3.2 NFMV の実装

NFMV の実装では、脱落時点が s の被験者を target、脱落時点が $t=s+1$ の被験者を donor とし、両者を同一投与群内で結合したデータセットを作成する。ここで重要なのは、初回欠測時点 t の値のみを補完対象とし、それ以降の欠測はこの段階では補完しないことである。

この目的のため、PROC MI の monotone regression 法を用いる際の変数リストを、ベースライン値および Y_1 から Y_t までに限定した。すなわち、PROC MI が参照する変数範囲を初回欠測時点までに制御することで、初回欠測までを補完し、その後の時点は欠測のまま保持する構成とした。これにより、NFMV の仮定に従って、target の初回欠測値に対し、1 時点先まで観測されている target の情報を反映させることができる。

得られた出力データからは、Imputation が 0 より大きいレコード、かつ dropvisit が s の target 被験者のみを抽出した。さらに、後続の CCMV 処理で wide 形式の構造を保つため、未作成の時点変数については欠測値を設定し、各代入データ間で変数構成が一致するように調整した。

```
/* 初回欠測時点までを補完 */
data nfmv_in_arm&arm._s&s.;
   set wide(where=(trt=&arm and dropvisit in (&s., &t.)));
   keep id trt base dropvisit y1-y&t.;
run;
proc mi data=nfmv_in_arm&arm._s&s.
            out=nfmv_out_arm&arm._s&s.
            nimpute=&nimp.
            seed=%eval(&seed. + 1000*&arm. + 10*&s.);
   monotone reg;
   var base y1-y&t.;
run;
```

3.3 CCMV の実装

CCMV の実装では、Step 1 で初回欠測時点が補完された target と、同一投与群内の完遂例を結合して、新たな代入用データセットを作成する。ここでは完遂例が donor となり、初回欠測以降に残る未観測値を補完する。この処理は、Step 1 で生成された各代入データに個別に実施した。すなわち、Imputation ごとに target を取り出し、それに完遂例を加えた上で PROC MI の monotone regression 法を再度適用した。変数リストにはベースライン値および Y_1 から Y_K までの全時点を含めることで、Step 1 後に残された欠測部分を補完した。これにより、初回欠測値には NFMV の仮定を保持しつつ、それ以降の欠測部分には CCMV の仮定を反映した多重代入データが得られる。

なお、dropvisit が $K-1$ の被験者では、初回欠測時点が最終時点 K に相当するため、Step 1 で欠測部分がすべて補完される。この場合、CCMV による追加の補完は不要であり、PROC MI による再代入は行わず、そのまま最終的な代入データとして採用した。このように処理を分けることで、不要な処理を避けつつ、欠測構造に応じた効率的な実装としている。

```
/* Step 1 の各 _Imputation_ ごとに Step 2 を実施 */
data work_ccmv_arm&arm._s&s._m&m;
   set wide(where=(trt=&arm. and dropvisit=&K))
         targets_step1_arm&arm._s&s.(where=(_Imputation_=&m.));
   drop _Imputation_ completer;
run;
proc mi data=work_ccmv_arm&arm._s&s._m&m
            out=ccmv_out_arm&arm._s&s._m&m
            nimpute=1
            seed=%eval(&seed + 2000*&arm + 100*&s + &m);
   monotone reg;
   var base y1-y&K.;
run;
```

3.4 代入データの結合と解析用データの作成

群別、脱落パターン別、および代入回数別に得られた補完済み target データは、PROC APPEND を用いて順次結合した。また、完遂例については各 Imputation に共通して存在する必要があるため、あらかじめ代入回数分だけ複製し、補完済み dropout 被験者と合わせて最終的な wide 形式の多重代入データセットを構成した。その後、この wide 形式データを再び long 形式に変換し、各被験者について各時点の観測値を 1 行ずつ持つ解析用データを作成した。このデータセットは、PROC MIXED による MMRM に直接入力可能な形式となっている。実装は、各時点の変数を順に取り出し、評価時点と観測値を対応付けながら出力することで、簡潔に long 形式へ戻した。

3.5 MMRM の実行と結果統合

作成した長形式の多重代入データに対して、Imputation ごとに PROC MIXED を実行した。モデルには、ベースライン値、投与群、評価時点および投与群と評価時点の交互作用を固定効果として含め、被験者内の繰り返し測定に対しては非構造型共分散構造を仮定した。これにより、各代入データについて MMRM に基づく推定結果を得た。

本論文の実行例では、最終時点における投与群間差を関心対象とし、PROC MIXED の LS-means 差の出力から当該推定値と標準誤差を抽出した。最後に、PROC MIANALYZE を用いて各代入データの推定結果を Rubin の規則に基づいて統合し、最終的な推定値とその不確実性を評価した。これにより、多重代入から MMRM、推定結果の統合までの一連の解析フローを、PROC IML を用いずに完結させることができた。Imputation ごとに MMRM を実行し、その結果を PROC MIANALYZE で統合する SAS コードを以下に示す。

```
/* 各 Imputation ごとに MMRM を実施 */
proc mixed data=mi_case4_long;
   by _Imputation_;
   class id trt visit;
   model y_obs = base trt visit trt*visit / solution;
   repeated visit / subject=id type=un;
   lsmeans trt*visit / diff;
   ods output diffs=diffs;
run;
/* Rubin の規則に基づいて統合 */
proc mianalyze data=diffs_final;
   modeleffects Estimate;
   stderr StdErr;
run;
```

4 実行例

本章では、前章で示した実装方法に基づき、2 群 4 時点の単調脱落データに対する実行例を示す。ここでは、群別・脱落パターン別に NFMV＋CCMV を適用した Pattern-Mixture Multiple Imputation を行い、その後に MMRM および結果統合を実施する。

4.1 シミュレーションデータの作成

実行例では、被験者数 *N*=200、予定評価時点数 *K*=4、代入回数 *NIMP*=20 とした。各被験者には投与群、ベースライン値、および被験者ごとのランダム効果を与え、各時点の連続反応変数を正規分布から生成した。さらに、観測値に依存する形で途中脱落を発生させることにより、単調脱落構造を持つデータを作成した。これにより、実務で想定される反復測定データに近い状況を再現した。

生成された long 形式データについて、各被験者の最終観測時点を dropvisit として算出し、wide 形式データへ変換した。wide 形式では、完遂例と脱落例が明確に区別され、後続の群別・脱落パターン別処理に利用できる形となった。

4.2 多重代入の実施

各投与群について、dropvisit が 1 から *K*−1 までの各脱落パターンに対し、まず NFMV により初回欠測時点を補完した。続いて、初回欠測後に残る欠測部分に対して、完遂例を donor とする CCMV を適用した。これらの処理を代入回数 *NIMP* 回分繰り返し、最終的に Imputation ごとに全被験者を含む wide 形式の多重代入データセットを得た。

代入後のデータについて、各 Imputation に含まれる被験者数を確認したところ、すべての代入データで元の被験者数と一致しており、完遂例と補完済み脱落例が適切に結合されていることが確認できた（表 2)。この

点は、実装の妥当性を確認する上で基本的かつ重要な確認項目である。

表 2. 各多重代入データにおける被験者数

多重代入番号（Imputation）	被験者数
1	200
2	200
3	200
4	200
5	200
6	200
7	200
8	200
9	200
10	200
11	200
12	200
13	200
14	200
15	200
16	200
17	200
18	200
19	200
20	200

4.3 MMRM による解析と結果統合

多重代入後の wide 形式データを long 形式へ変換した後、Imputation ごとに PROC MIXED を用いて MMRM を実行した。関心対象は最終時点における投与群間差とし、LS-means 差から推定値および標準誤差を抽出した。これらを PROC MIANALYZE に入力することで、代入間のばらつきを考慮した最終的な推定結果を得た。統合後の推定値、標準誤差および 95%信頼区間を表 3 に示す。

表 3. 最終時点における投与群間差の統合推定結果

評価時点	推定値	標準誤差	95%信頼区間下限	95%信頼区間上限
Visit 4	4.5001	1.8589	0.7537	8.2465

この実行例により、NFMV＋CCMV に基づく Pattern-Mixture Multiple Imputation を、SAS の標準プロシジャのみを用いて実行可能であることが確認できた。また、データ整形から多重代入、MMRM、結果統合までの一連の処理を一貫したフローとして構成できることから、実務における再利用しやすい実装例となることが示

された。
なお、参考のため、観測値のみを用いた MMRM による結果との比較を表 4 に示す。本論文の主目的は実装方法の提示であるため、ここでは結果の差異の有無を確認するにとどめる。

表 4. 観測値のみを用いた MMRM と Pattern-Mixture Multiple Imputation 後の MMRM の比較

解析方法	推定値	標準誤差	95%信頼区間下限	95%信頼区間上限	p 値
Pattern-Mixture Multiple Imputation 後の MMRM	4.5001	1.8589	0.7537	8.2465	0.0197
観測値のみを用いた MMRM	3.9481	1.5525	0.8865	7.0097	0.0118

5 考察

本論文では、単調脱落がある 2 群 4 時点の反復測定データを対象として、初回欠測時点に NFMV（Non Future Missing Values）、それ以降の欠測部分に CCMV（Complete Case Missing Values）を適用する Pattern-Mixture Multiple Imputation を、PROC IML を使用せずに SAS の標準プロシジャのみで実装する方法を示した。実装例を通じて、データの整形、欠測パターンの識別、多重代入、MMRM、結果統合までを一連の処理として構築できることを確認した。
本論文の主な貢献は、NFMV と CCMV を組み合わせた Pattern-Mixture Multiple Imputation を、PROC IML に依存せず、SAS の標準プロシジャのみで実装する方法を具体的に示した点にある。さらに、多重代入そのものにとどまらず、データの整形、解析用データセットの再構成、MMRM、結果統合までを含めた一連の解析フローとして整理したことで、実務で再利用しやすい実装例を提示した点にも意義がある。
本実装の主な利点の一つは、PROC MI、PROC MIXED、PROC MIANALYZE といった比較的利用頻度の高い標準プロシジャにより構成できる点にある。Pattern-Mixture Model に基づく多重代入は、欠測後データに対する仮定を明確に扱える一方で、実務上は実装の複雑さが導入の妨げとなることが少なくない。特に、初回欠測時点とそれ以降の欠測部分で異なる仮定を適用する場合、対象データの選択、donor 群の設定、欠測補完後データの再構成など、多段階の処理が必要となる。本論文で示した方法は、こうした処理を SAS の標準機能の範囲で整理し、比較的見通しのよい手順として実装できる点で有用と考えられる。
また、本手法では、同一投与群内で donor を選択し、脱落パターンごとに処理を分けることで、脱落後の未観測データに対する仮定を段階的に反映している。初回欠測時点では、対象被験者より 1 時点先まで観測されている被験者の情報を参照することで、脱落直後の挙動に対して比較的限定的な仮定を置いている。一方、それ以降の欠測部分では完遂例を参照することで、残りの未観測部分に対する補完を行っている。このように、NFMV と CCMV を組み合わせることで、脱落後の未観測データに対する仮定を単一の形に固定せず、時点に応じて切り分けて扱える点は、本手法の特徴である。
一方で、本論文の実装にはいくつかの前提と制約がある。第一に、対象とする欠測構造は単調脱落を前提としているため、間欠的欠測を多く含むデータに対しては、そのままでは適用しにくい。第二に、実行例は 2 群 4 時点の設定に基づいており、より多くの評価時点や複雑な群構成に拡張する場合には、マクロ処理や中間

データの管理を含めた追加の工夫が必要となる。第三に、NFMV および CCMV は欠測後データに対する仮定を具体化する一例であり、その妥当性は解析目的、対象疾患、脱落理由、観測スケジュールなどを踏まえて検討されるべきである。したがって、本手法はあくまで一つの感度分析の実装例として位置づけるのが適切である。

さらに、Estimand の観点からは、本論文で扱った手法は推定対象そのものを定義するものではなく、脱落後の未観測データをどのように扱うかという解析上の仮定を具体化する方法と理解する必要がある。すなわち、まず何を推定したいのかという Estimand を明確にし、その上で、その Estimand に対応した欠測後データの仮定として NFMV＋CCMV 型の補完を採用する、という順序で整理することが重要である。この点を明示することで、本手法を主解析を補完する感度分析として適切に位置づけることができる。

今後の課題としては、評価時点数や群数の増加に対する一般化、より柔軟な欠測構造への対応、解析目的に応じた donor 設定の検討、ならびに実務で再利用しやすいテンプレート化やモジュール化などが挙げられる。また、主解析との比較や、異なる欠測後仮定を置いた複数の Pattern-Mixture Multiple Imputation 手法との比較検討を行うことで、本手法の位置づけをさらに明確にできると考えられる。

6 まとめ

本論文では、単調脱落がある反復測定データに対して、初回欠測時点に NFMV、以降の欠測部分に CCMV を適用する Pattern-Mixture Multiple Imputation を、PROC IML を用いずに SAS の標準プロシジャのみで実装する方法を示した。具体的には、PROC MI、PROC MIXED、PROC MIANALYZE、ならびに PROC SQL、PROC TRANSPOSE、データステップ、マクロ処理を組み合わせることで、データ整形から多重代入、MMRM、結果統合までの一連の解析フローを構築した。

本実装により、脱落後の未観測データに対する仮定を反映した Pattern-Mixture 型の多重代入を、比較的見通しよく再利用しやすい形で SAS 上に構成できることを示した。特に、PROC IML を前提としない点は、実務上の導入可能性を高める上で意義があると考えられる。

反復測定データにおける欠測の扱いは解析結果に影響する重要な要素であり、Estimand の観点からも、欠測後データに対する仮定を明確にした上で感度分析を行うことが求められる。本論文で示した実装は、そのような感度分析を SAS の標準機能で実施するための一つの具体例として、実務において有用であると考えられる。