

4.2.1 尤度に基づく推測

前節で述べた議論が成立するのは推測方式が尤度に基づく場合である。ベイズ推測に対しても上記と同じ結論が得られるが (Rubin, 1976), 頻度論の枠組みで推測を行う場合には, 欠測メカニズムが無視できるのはMCARの場合だけである。これは, MARのもとでは, 観察データ Y_i^o と脱落 D_i が独立ではないことに起因する。すなわち, $f(y_i^o | d_i; \theta) \neq f(y_i^o; \theta)$ であるため, 観察データのみから計算された単純な平均値などは不偏推定値にならないからである。このことは, 経時観察データ解析においてよく用いられる尤度に基づかない推測方式, 例えば, 一般化推定方程式 (generalized estimating equations; GEE) (Liang and Zeger, 1986) などは, 欠測メカニズムがMCARでない限り妥当な推測を導かないことを意味する。

4.2.2 周辺尤度の特定

たとえ欠測メカニズムを無視できた (MCAR, あるいはMAR) としても, バイアスのない結果を得るためには測定過程に対するモデル $f(y_i^o; \theta)$ が正しいことが前提である。この問題は, 欠測データ解析に限ったことではないが, 結果変数の経時変化, 共変量効果などのモデル化を慎重に行わなければならないことを意味する。また, 式(4)のように, 脱落するかどうかにかかわらず共変量が関係している場合には, そのような変数を考慮した解析を行う必要があることも意味する (Little, 1995)。

4.2.3 最尤推定量の標本分布

尤度に基づく推測を行ったとしても, 得られた最尤推定量の標本分布の特徴は欠測過程に依存する。つまり, 欠測メカニズムがMARの場合には, 4.2.1でも述べたように, 観察データ Y_i^o と脱落 D_i が独立ではないので, D_i を与えたもとの Y_i^o の条件付き期待値は欠測過程に依存する。したがって, θ の最尤推定量の漸近分散は通常の期待情報行列ではなく, 観察情報行列を用いて計算しなければならない (Kenward and Molenberghs, 1998; Laird, 1988; Little and Rubin, 2001)。

4.2.4 パラメータの分離条件

観察データの周辺尤度のみに基づいて推測を行うためには, 測定過程を表すパラメータ θ と欠測過程を表すパラメータ γ に共通のパラメータがないことが必要である。この仮定はパラメータの分離条件 (separability condition) と呼ばれる (Little and Rubin, 2001; Rubin, 1976)。この分離条件は, 多くの臨床・疫学研究において必ずしも成立するとは限らない。例えば, 脱落するかどうかで治療群間でことなる場合には, 治療変数は θ と γ の両方に含まれることになる。分離条件を満たしていない場合に, 欠測メカニズムを無視して観察データの周辺尤度のみに基づいて θ に関する推測を行うと, 一般に精度が減少する (Diggle et al., 2001)。

4.2.5 データから検証不能な仮定

式(3)あるいは式(4)のMCARの仮定に関しては, ある時点で脱落するかどうかをそれまでの観察データ Y_i^o , 共変量 X_i の関数としてモデル化し, それらの間の関連を見ることで, その妥当性を検討することができる。この検討は, 脱落者の数が少ないと一般に検出力が低い, Y_i^o と脱落との間に関連が見られた場合には, 少なくとも欠測メカニズムはMCARでないことがわかる。しかしながら, このような検討から欠測メカニズムがMARであるとはいえない。なぜなら, MARは観察していないデータ Y_i^m に基づく定義, すなわち, データからは検証不能な仮定であ

り、MNARの可能性は常に否定できないからである。この意味において、欠測データ解析におけるMARの仮定は、因果推論において因果効果をバイアスなく推定するために必要な仮定「調整された共変量レベル内では交絡は存在しない」によく似ている (Greenland and Finkle, 1995)。脱落メカニズムに関して、あるシナリオより別のシナリオのほうがもっともらしいということはあるかもしれないが、一般には、脱落メカニズムは複雑で未知であるのが通常であり、安易にMARを仮定することは危険である。

5. 欠測データに対する単純な解析方法

5.1 Complete case 解析

欠測データに対する最も単純な対処方法は、予定されたすべての測定がなされた対象者のみを解析対象とする方法、すなわち complete case 解析である。表1のデータでは、時点4で結果変数が観察されている714人が解析対象で、途中で1回でも結果が観察されなかった437人は解析から除外される。

予定されていたすべての測定がなされた対象者だけに解析対象を限定してどのような解析手法を用いるかは研究目的によるが、どのような解析手法を用いたとしても、その結果が妥当であるためには、欠測メカニズムがMCARであることが必要である。多くの臨床・疫学研究で、最後まで観察された対象者がもともとの集団からのランダムサンプルであることはまれなので、complete-case 解析の結果には一般にバイアスがあると思われる。また、たとえ欠測メカニズムにMCARを仮定することが合理的であったとしても、complete-case 解析は解析に寄与する対象者数が減るので、検出力の観点からも好ましい解析方法ではない。

5.2 Available data 解析

欠測データに対するもう一つの対処方法は、観察された利用可能なすべてのデータを用いて解析する方法、すなわち available data 解析である。表1のデータでは、試験途中で脱落したかどうかにかかわらず、観察されている時点までのデータを解析に利用する方法である (表1の合計と記されたところの人数が各時点で寄与する解析対象者となる)。この解析方法は、脱落した対象者も解析に含めるので、一般に complete case 解析よりも効率が良い。

この解析方法も、解析対象集団は一意に定まるものの、どのような解析方法を用いるかは研究目的による。4節で述べたように、欠測メカニズムにMARを仮定することができれば、尤度に基づいた available data 解析は欠測メカニズムを考慮することなく妥当な結果を導く。しかしながら、MARのもとでは、観察データはもともとのデータからのランダムサンプルではないので、観察データのみから計算された単純な平均値などは不偏推定値にならない。したがって、尤度に基づかない available data 解析の結果が妥当であるためには、欠測メカニズムがMCARであることが必要である。

5.3 LOCF 解析

欠測データに対してよく用いられるもう一つの対処方法は欠測データを何らかの値で埋める方法 (imputation method) である。欠測データの補完方法にはいくつかの方法が考えられるが (Little and Rubin, 2001)、表1のような経時観察データの解析においてよく用いられるのが、最後に観

察された値で補完する方法, すなわち LOCF 解析 (last observation carried forward analysis) である (Gillings and Koch, 1991). 例えば, 表 2 のデータで, 結果変数の観察パターンが (○, ○, ×, ×) であった 155 人の対象者の観察されていない 3 時点目と 4 時点目の結果変数の値は, 各個人の 2 時点目で観察された値でそれぞれを置き換える方法である.

この解析方法は, 予定していたすべてのデータを解析に用いることができ, 欠測が存在しない場合のデータ解析となるので, 非常に単純であるという利点がある. しかしながら, この方法が妥当であるためには, 「脱落後の結果変数の推移は最後に観察された値のまま変化しない」という非常に強い, しばしば非現実的な仮定を必要とする. 例えば, 症状が改善, あるいは治癒したことを理由に脱落した場合のように LOCF 解析の仮定が合理的と思える状況も存在するかもしれないが, そのような仮定を保証するだけの生物学的根拠は存在しないのが通常である.

LOCF 解析とよく似た方法で, 比較目的の相対評価の場合には, 各個人の最悪値で欠測値を補完するという方法もよく用いられる (Gillings and Koch, 1991). 表 1 のデータでは, 欠測値はすべて無月経「なし」とみなす方法である. また, この方法の延長として, 「試験薬に不利になるように補完する」という方法も考えられる. これらの対処方法は, 治療効果の比較に関して保守的な結果を導くことを期待して用いられることが多い. しかしながら, 上で述べた LOCF 解析と同様の非常に厳しい仮定を前提としており, 薬剤の有効率の推定などの絶対評価にはバイアスをともなう. さらに, LOCF 解析を含むこのような欠測値のある値で 1 回だけ置き換える単純な補完方法は, 補完された値と実際に観察された値を同等に扱っており, 推定精度を過小評価するという問題がある (Diggle et al., 2001; Fitzmaurice et al., 2004; Little and Rubin, 2001; Ware, 2003).

5.4 解析結果

これまで述べた 3 つの単純なアプローチを表 1 のデータに対して適用する. 解析方法は, 個人内で結果変数が繰り返し測定されることによる相関を考慮した一般化推定方程式 (GEE) による方法 (Liang and Zeger, 1986) で, そのモデルから算出される時点ごとの無月経割合の推定値, およびその群間差を求めることを目的とする.

GEE による解析では, 対象者 i に対する時点 t ($t = 1, \dots, 4$) における無月経の有無を Y_{it} (有なら $Y_{it} = 1$, 無なら $Y_{it} = 0$) とし, その周辺平均に対して, 以下のようなロジスティック回帰モデルを仮定した (Fitzmaurice et al., 2004).

$$\text{logit}(\mu_{it}) = \beta_1 + \beta_2 t_{it} + \beta_3 t_{it}^2 + \beta_4 \text{dose}_i + \beta_5 (t_{it} \times \text{dose}_i) + \beta_6 (t_{it}^2 \times \text{dose}_i) \quad (8)$$

式 (8) で, $\mu_{it} = \Pr(y_{it} = 1)$, 150 mg 群なら $\text{dose} = 1$, 100 mg 群なら $\text{dose} = 0$ である.

個人内の結果変数の相関構造はオッズ比で定量化し, 以下のような時点ごとの各ペアに対する対数オッズ比を仮定した (Cary et al., 1993).

$$\log OR(y_{ij}, y_{ik}) = \alpha_{jk} \quad (9)$$

ただし, j と k は時点を表し, $OR(y_{ij}, y_{ik}) = \frac{\Pr(y_{ij}=1, y_{ik}=1)\Pr(y_{ij}=0, y_{ik}=0)}{\Pr(y_{ij}=1, y_{ik}=0)\Pr(y_{ij}=0, y_{ik}=1)}$ である. この GEE 解析は, 統計パッケージ SAS の GENMOD プロシジャで実行可能である.

解析結果を表 4 に示す. 表中の解析方法で, CC は complete case 解析, AD は available data 解析を意味する. Complete case 解析と available data 解析の結果について考える. もしこのデー

表 4. 時点ごとの無月経割合の推定値とその群間差の推定値 (単純な解析方法)

解析方法	時点	150 mg	100 mg	差	差の標準誤差	Z 値	p 値
CC	1	0.155	0.176	-0.022	0.027	-0.79	0.43
	2	0.317	0.258	0.059	0.028	2.07	0.04
	3	0.463	0.368	0.094	0.033	2.83	0.005
	4	0.540	0.502	0.038	0.037	1.03	0.30
AD	1	0.201	0.184	0.017	0.023	0.73	0.46
	2	0.363	0.274	0.089	0.025	3.55	0.0004
	3	0.499	0.388	0.111	0.030	3.68	0.0002
	4	0.572	0.517	0.055	0.036	1.52	0.13
LOCF	1	0.201	0.184	0.017	0.023	0.75	0.45
	2	0.344	0.263	0.081	0.024	3.43	0.0006
	3	0.453	0.350	0.103	0.027	3.78	0.0002
	4	0.498	0.437	0.061	0.029	2.09	0.04

CC: complete case 解析 AD: available data 解析

タの脱落メカニズムがMCARであれば、上記のGEE解析から得られる時点ごとの無月経割合の推定値はいずれのアプローチでも妥当である。Available data解析では、時点2と時点3において、無月経割合が高用量(150mg)群のほうが有意に高い結果となっている。時点4では有意な群間差は存在しない。一方、complete case解析では、同様の傾向がみられているものの、対象者数の減少にともない標準誤差がわずかに大きくなっている。また、他の方法と比べていずれの時点においても群間差が小さな値となっている。

LOCFに基づくGEE解析の結果では、available data解析の結果と比較して、時点2と時点3の無月経割合の推定値が小さく見積もられている。ただし、それらの時点での群間差の結果に関しては大きな違いは見られない。しかしながら、LOCF解析の結果は、他の方法よりも群間差の標準誤差が小さな値となっている。その結果、LOCF解析でのみ時点4において有意な治療効果が見られている(p値=0.04)。

6. 欠測メカニズムを考慮した解析方法

前節で示したGEEによる解析結果が妥当であるためには、complete case解析とavailable data解析については、脱落メカニズムがMCARであることが必要であり、LOCF解析では、「脱落後の結果変数の推移は最後に観察された値のまま変化しない」という非常に強い仮定が必要である。4.2.5でも述べたように、MCARの仮定が妥当かどうかはデータからチェックすることができる。各時点において脱落するかどうかにかかわらずにそれまでに観察された結果変数が依存するかどうかを調べるために、以下のようなロジスティック回帰モデルのあてはめを考える。

$$\log \frac{\Pr(d_i = t \mid d_i \geq t, y_{i1}, \dots, y_{it-1})}{\Pr(d_i > t \mid d_i \geq t, y_{i1}, \dots, y_{it-1})} = \gamma_0 + \gamma_1 \text{dose}_i + \gamma_2 y_{i1} + \dots + \gamma_t y_{it-1} \quad (10)$$

式(10)で、 γ は回帰パラメータで、表1のデータでは $t = 2, \dots, 4$ である。

このモデルをあてはめた結果を表5に示す。いずれの時点においても、脱落するかどうかにか
Jpn J Biomet Vol. 25, No. 2, 2004

表 5. 各時点の脱落に影響する要因の検討

脱落時点	変数	推定値	標準誤差	p 値
時点 2	dose	-0.001	0.157	0.95
	Y_{i1}	0.540	0.181	0.003
時点 3	dose	0.252	0.178	0.16
	Y_{i1}	0.600	0.195	0.002
	Y_{i2}	-0.046	0.236	0.85
時点 4	dose	-0.327	0.236	0.17
	Y_{i1}	0.718	0.259	0.006
	Y_{i2}	-0.331	0.295	0.26
	Y_{i3}	0.186	0.310	0.55

点 1 (最初の 3ヶ月) における結果変数の値が影響している (無月経ありの対象者のほうが脱落しやすい) ことがわかる. したがって, 表 1 のデータの場合, 脱落メカニズムは少なくとも MCAR でないことがわかる. 以下の 6.1 節と 6.2 節では, データが欠測した理由を解析で考慮することで, 欠測メカニズムが MAR の場合でも妥当な結果を導く方法を紹介する.

6.1 重み付き解析

欠測メカニズムを考慮した解析方法の一つが, 観察データを何らかの適当な値で重み付ける方法である (Heyting et al., 1992; Robins et al., 1995; Rotnitzky and Robins, 1995). この重み付き解析では, 観察されたデータには選択バイアスがあり, その選択確率を考慮してバイアスを修正することを目的としている. ある対象者がある時点で観察されるかどうかの選択確率は, それまでの結果変数の値や対象者の特性を表す共変量などの欠測に影響すると思われる要因から推定される.

表 1 の時点 2 のデータについてのみ考えてみる. 各群における無月経割合の真値をそれぞれ β^{150} , β^{100} とすると, 単純な推定量 $\sum R_{i2}Y_{i2}/\sum R_{i2}$ (150 mg 群では 160/476, 100 mg 群では 125/477) は, 欠測メカニズムが MCAR でないので, β^{150} と β^{100} の不偏推定値ではない. ここで, 欠測メカニズムが MAR であれば, 時点 2 で観察されるかどうかはそれまでの観察履歴で予測可能なので, ある対象者が時点 2 で観察される確率に関して, 以下の式が成立する.

$$\pi_i = \Pr(r_{i2} = 1 \mid y_{i1}, y_{i2}, dose_i, x_i) = \Pr(Y_{i2} = 1 \mid y_{i1}, dose_i, x_i) \quad (11)$$

ここで, 式 (11) の各対象者に対する観察確率が既知とすると,

$$\begin{aligned} & E\{\pi_i^{-1} r_{i2} \{y_{i2} - \beta^{150} dose_i - \beta^{100} (1 - dose_i)\}\} \\ &= E\{E(r_{i2} \mid y_{i1}, y_{i2}, dose_i, x_i) \pi_i^{-1} \{y_{i2} - \beta^{150} dose_i - \beta^{100} (1 - dose_i)\}\} \\ &= E\{y_{i2} - \beta^{150} dose_i - \beta^{100} (1 - dose_i)\} = 0 \end{aligned}$$

となるので, 時点 2 で観察された対象者をその個人の観察確率の逆数で重み付けた推定量 (inverse probability weighted estimator) $\sum \pi_i^{-1} R_{i2} Y_{i2} / \sum \pi_i^{-1} R_{i2}$ は不偏推定量となる.

この重み付き解析の背後にある考え方は、観察確率が例えば0.25であれば、そのような対象者は期待的には3人観察されないことになるので、その3人分のデータも解析で考慮しようというものである。つまり、観察対象者の解析への貢献度を $1/\pi_i$ 倍することで、その対象者自身と、その時点で脱落したよく似た（それまでの結果変数の履歴や共変量が同じ）対象者 $(1/\pi_i - 1)$ 人も解析で考慮することになる。

観察確率の逆数で重み付ける推定量は、標本調査の分野でも古くから提案されており、Horvitz-Thompson 推定量として知られている (Horvitz and Thompson, 1952)。標本調査の分野では、調査デザインにもとづいて重み（サンプリング確率の逆数）が既知であるが、表1のような多くの臨床試験データでは重みは未知であるのが通常である。したがって、重みはデータから推定しなければならない。例えば、式(10)のような時点ごとにロジスティック回帰モデルをあてはめることで重みは推定可能である。したがって、データから重みを推定さえできれば、結果変数の型によらず、観察されたすべてのデータに対して標準的な解析手法の重み付き解析を実行すればよい。ただし、この解析方法が妥当であるためには、重みが正しく推定されていることが前提なので、脱落に影響すると思われる多くの要因を重みの推定のためのモデルに取り込むことが重要である (Robins et al., 1995)。

重み付き推定量の分散に関しては、重みの推定誤差を考慮した分散が提案されている (Robins et al., 1995)。直感的には、重みの推定誤差を考慮したほうが分散が大きくなるように思えるが、データから重みを推定したほうが精度が上昇することが知られている。付録Aにその分散の推定量を示す。

最後に、表1のような経時観察データの場合、各時点での重みは累積確率の逆数となることに注意が必要である。例えば、時点2での重みは、 $1/\hat{\pi}_{i2}$ であるが、時点3、時点4では、式(10)のようなモデルから求まる各時点での条件付き存在確率 $\hat{\pi}_{i3}$ 、 $\hat{\pi}_{i4}$ を用いて、それぞれ $(1/\hat{\pi}_{i2}) \times (1/\hat{\pi}_{i3})$ 、 $(1/\hat{\pi}_{i2}) \times (1/\hat{\pi}_{i3}) \times (1/\hat{\pi}_{i4})$ となる。

6.2 多重補完法

データが欠測した理由を解析で考慮することで、欠測メカニズムがMARの場合でも妥当な結果を導くもう一つの方法は、多重補完法 (multiple imputation method) である (Rubin, 1987, 1991; Schafer, 1997, 1999)。LOCF解析や欠測値を観察平均値で置き換える単純な補完法は、得られる結果のバイアスの問題だけでなく、本来は観察されていない値をあたかも観察されたかのように扱うので、欠測値のもつ不確実性を適切に考慮していない。多重補完法では、欠測値をことなる値で複数回 (M 回) 補完することで、この問題を回避する。

多重補完法では、 M 個の欠測のない完全なデータセットが作成されるので、 M 個のことなるパラメータ推定値とそれぞれの標準誤差が得られる。最終的な結果は、それらを併合して1つの結果にまとめられる。 m 個目 ($m = 1, \dots, M$) の欠測のないデータセットから得られる推定量を $\hat{\theta}^{(m)}$ 、 $\hat{\theta}^{(m)}$ の分散を $\hat{v}^{(m)}$ とすると、多重補完法による θ の推定量は、

$$\hat{\theta} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}^{(m)} \quad (12)$$

であり、その分散は、

$$\text{Var}(\hat{\theta}) = \bar{U} + (1 + M^{-1})B \quad (13)$$

ただし、 $\bar{U} = \frac{1}{M} \sum_{m=1}^M \hat{U}^{(m)}$, $B = \frac{1}{M-1} \sum_{m=1}^M (\hat{\theta}^{(m)} - \hat{\theta}) - (\hat{\theta}^{(m)} - \hat{\theta})^T$ である。分散の推定量は複雑な式にみえるが、2つのバラツキの要素の和で表現されている。式(13)の第1項の \bar{U} が補完内分散で、第2項の B が補完間分散である。

θ に関する仮説検定や区間推定は、式(12)と式(13)を用いて、

$$(\theta - \hat{\theta}) / \sqrt{\text{Var}(\hat{\theta})}$$

が自由度 v の t 分布に従うことから行うことができる。ただし、 $v = (M-1)(1+r^{-1})^2$, $r = (1+M^{-1})B/\bar{U}$ である。なお、欠測のない完全なデータセットの自由度 v_{comp} が小さく、欠測データの数が少ない場合には、上式から計算された自由度 v が v_{comp} よりも大きな値となる場合があるので、以下のような調整した自由度 v_s を用いることが提案されている (Barnerd and Rubin, 1999)。

$$v_s = (1/v + 1/\hat{v}_{obs})^{-1} \quad (14)$$

ただし、 $\hat{v}_{obs} = (1+r)^{-1}v_{comp}(v_{comp}+1)/(v_{comp}+3)$ である。

実際のデータ解析において、補完回数 M をいくつに設定するかは、5から10程度を採用している論文が多いが、補完回数を増やしても推定値が安定していることを確認しておくべきである (Horton and Lipsitz, 2001)。

多重補完法の基本的な考え方は単純であるが、補完するためのデータの生成にはいくつかの方法が提案されている (Rubin, 1987; Schafer, 1997)。

6.2.1 回帰モデルにもとづく方法

この方法は欠測値を予測するための回帰モデルを作成し、そのモデルから補完するためのデータを発生させる方法である。例えば、ある時点 t の結果変数 Y_{it} が一部の対象者に対しては欠測しているが、治療群を含む共変量 X_i は全員に対して測定がなされている状況では、以下のような回帰モデルを結果変数 Y_{it} が観察されている対象者に対してあてはめる。

$$E(y_{it}) = \beta_0 + \beta_1 x_i + \beta_2 y_{i1} + \dots + \beta_t y_{it-1} \quad (15)$$

式(15)の回帰パラメータの推定値 $\hat{\beta}$ と誤差分散の推定値 $\hat{\sigma}^2$ が得られれば、以下のステップに従い、補完するためのデータを生成する。

- (1) β と σ^2 を推定する際の不確実性を考慮するために、 β と σ^2 の新しい値 β^* と σ^{2*} を、それぞれ $\hat{\beta}$ と $\hat{\sigma}^2$ の分布から発生させる。
- (2) β^* と σ^{2*} を用いて、予測式 $y_{it} = \beta_0^* + \beta_1^* x_i + \beta_2^* y_{i1} + \dots + \beta_t^* y_{it-1} + \sigma^* e_i$ から補完するためのデータ(予測値)を求める。ただし、 X_i や Y_{ij} ($j=1, \dots, t-1$)の値は、結果変数 Y_{it} が欠測している対象者のデータを用い、 e_i は標準正規分布からの乱数である。

多重補完法では、上記のステップを独立に M 回繰り返すことで、完全なデータセットを M 個作成する。この回帰モデルにもとづく方法の変法として、予測平均マッチング法(predictive mean

matching method) と呼ばれる方法もある (Heitjan and Little, 1991). この方法は, 上記のステップ (2) で得られた予測値 Y_{it}^* 自体を補完のために使うのではなく, Y_{it}^* の値と予測値が最も近い観察データの値を補完に用いる方法である. 通常の回帰モデルにもとづく方法では, 極端な値, あるいは範囲外の値をとる予測値 Y_{it}^* がたまたま生成される可能性があるが, 予測平均マッチング法ではそのようなことが起こりにくいという特徴がある.

表 1 のように結果変数が 2 値データの場合も上記と同様の考え方で補完のためのデータを生成できる. ただし, 欠測値を予測するためのモデルを作成する際には, 結果変数が観察されている対象者において, 例えば, ロジスティック回帰モデルをあてはめることになる. そして, 上記のステップ (2) では, 結果変数が欠測している対象者の予測反応確率を求め, それを一様乱数と比較するという操作が追加される.

6.2.2 傾向スコアにもとづく方法

傾向スコア (propensity score) とは, もともとは共変量を与えたもとである特定の曝露 (治療法) を受ける条件付き確率のことである (Rosembaum and Rubin, 1983). 式 (10) のようなモデルから計算される個人ごとの欠測確率を同様に傾向スコアと呼ぶとすると, 傾向スコアにもとづく方法とは, 各対象者の傾向スコアの値が同じくらいの (結果変数が観察された) 対象者のデータで欠測値を補完する方法である (Lavori et al., 1995; Rubin, 1991). 具体的には, 以下のステップに従って補完するためのデータを生成する.

- (1) データが欠測するかどうかを結果変数とした式 (10) のようなモデルをあてはめ, そのモデルから予測される個人ごとの傾向スコアの推定値を計算する.
- (2) 傾向スコアの推定値の大きさにより対象者をいくつかのグループに分ける.
- (3) 各グループにおいて, 結果変数が観察されている対象者, 欠測している対象者の数をそれぞれ N^O, N^M とする. 結果変数が観察されている N^O 個のデータからランダムに復元抽出で N^O 個の補完の候補となるデータをサンプリングする (この操作は, 回帰モデルにもとづく方法のステップ (1) と同様に, 傾向スコアを推定する際の不確実性を考慮するためである). この N^O 個のデータからランダムに復元抽出で N^M 個の補完のためのデータをサンプリングする.

表 1 のような経時データの場合, 上記の 3 つのステップを時点 2 から順番に実行し, それらの結果を併合する. なお, 上記のステップ (3) の操作は, approximate Bayesian bootstrap imputation と呼ばれる (Rubin, 1987, 1991).

6.2.3 MCMC 法にもとづく方法

多重補完法はベイズ流の枠組みから正当化できる (Rubin, 1987). θ の事後分布を $f(\theta | y^o)$ と表すと, 多重補完法の手続きは, 以下のように数式表現できる.

$$f(\theta | y^o) = \int f(\theta | y^o, y^m) f(y^m | y^o) dy^m$$

$$f(y^m | y^o) = \int f(y^m | y^o, \gamma) f(\gamma | y^o) d\gamma \tag{16}$$

ここで, γ は欠測値の予測分布を規定するパラメータで, 式 (15) のモデルにおいては, $\gamma = (\beta, \sigma^2)$ である. 式 (16) の 2 式目からわかるように, 観察データ Y^o が与えられたもとでの γ の事後分布

$f(\gamma|y^0)$ から γ を抽出し、次に抽出した γ と観察データが与えられたもとの Y^m の条件付き分布から欠測値を予測することになる。予測値が得られれば、式(16)の1式目にしたがって、 θ に対する複数の結果を併合する。

MCMC (Markov Chain Monte Carlo) 法とは、ベイズ統計学において、事後分布や事後確率の計算を行なうためのマルコフ連鎖に基づいたシミュレーションによる方法である (Gelfand and Smith, 1990; Gilks et al., 1996)。MCMC にもとづく方法とは、式(16)の2つの式からそれぞれのサンプルを繰り返し発生させることによって、ターゲットである θ の事後分布を推定する方法である (Schafer, 1997)。具体的には、 θ に対する適当な値を与えたもとの Y^m の条件付き分布 $f(y^m|y^0)$ から欠測値 Y^m をサンプリングし、その値で欠測値を補完した完全なデータセットから θ の値をサンプリングする。サンプリングされた θ の値を用いて同様のサンプリングを繰り返すという手順である。この繰り返しサンプリングからマルコフ連鎖 $(Y^{m(1)}, \theta^{(1)})$, $(Y^{m(2)}, \theta^{(2)})$, ..., $(Y^{m(K)}, \theta^{(K)})$ が形成され、繰り返し数 K が十分大きければ、 K 回目以降のサンプルは $f(y^m, \theta|y^0)$ からの独立なサンプルとみなすことができる。

6.3 解析結果

表1のデータに対して重み付き解析と多重補完法による解析を適用した結果を示す。重み付き解析に関しては、試験途中で脱落したかどうかにかかわらず、観察されている時点までのすべてのデータを利用した available data 解析であり、どちらの解析方法も最終的な解析モデルは、式(8)と式(9)に基づく GEE 解析である。

重み付き解析と多重補完法は、どちらも欠測値を何らかの値で代用するという点でよく似た解析方法である。前者は、観察データをその観察確率の逆数にしたがって膨らますことで、観察データで欠測値を代用する方法であり、後者は、欠測データを観察データから予測し、その予測値で欠測値を代用する方法である。しかしながら、非常に単純な状況では2つの方法が同じ点推定値を与える場合もあるが、一般には2つの方法はことなる結果を導く (Fleiss et al., 2003)。

重み付き GEE 解析について考える。式(10)を用いて各時点の観察確率を推定した (表5参照)。群ごとの重みの推定値の分布を表6に示す。重みの推定値、バラツキともに時点4になるにつれ大きくなっているが、極端に大きな重みを示す対象者は存在しないようである。

表6. 重みの推定値の分布

群	時点	対象者数	平均値	標準偏差	最小値	最大値
150 mg	2	476	1.207	0.051	1.183	1.314
	3	389	1.475	0.113	1.396	1.726
	4	353	1.624	0.142	1.498	1.950
100 mg	2	477	1.208	0.050	1.185	1.317
	3	409	1.411	0.098	1.350	1.638
	4	361	1.600	0.145	1.488	1.932

重み付き GEE 解析の結果を表7に示す。表4に示した単純な (重みなしの) available data 解析とほぼ同じ結果になっている。脱落メカニズムが MAR であったとしても、時点2と時点3において、無月経割合が高用量 (150 mg) 群のほうが有意に高い結果となっている。また、時点4

表 7. 時点ごとの無月経割合の推定値とその群間差の推定値 (重み付き解析と多重補完法)

解析方法	時点	150 mg	100 mg	差	差の標準誤差	Z 値	p 値
重み付き解析	1	0.199	0.184	0.015	0.023	0.67	0.51
	2	0.368	0.279	0.089	0.024	3.71	0.0002
	3	0.508	0.398	0.110	0.030	3.67	0.0002
	4	0.580	0.529	0.052	0.035	1.49	0.14
多重補完法	1	0.200	0.183	0.016	0.027	0.62	0.54
	2	0.364	0.272	0.092	0.024	3.89	0.0001
	3	0.500	0.387	0.113	0.027	4.24	0.0001
	4	0.570	0.520	0.049	0.034	1.45	0.15

では有意な群間差は存在しない。

多重補完法の結果を表 7 に示す。補完のためのデータは回帰モデルにもとづく方法で生成した。補完モデルは、各時点での反応確率に対するロジスティック回帰モデルで、予測因子は、治療群とそれまでの結果変数の値である。補完回数は $M = 20$ とし、解析は統計パッケージ SAS の MI プロシジャと MIANALYZE プロシジャで行った。その結果は、時点 1 を除けば、各時点での無月経割合がわずかに小さくなっているものの、いずれの時点においても群間差に関しては重み付き解析とほぼ同じ結果である。

6.4 欠測メカニズムが MNAR の場合の解析

脱落するかどうか欠測した結果変数自体に影響を受ける場合、すなわち欠測メカニズムが MNAR である場合には、これまで述べてきた解析方法を含むほとんどすべての標準的な解析方法ではバイアスのない群間比較を行えない。バイアスのない推定値を得るためには、結果変数の測定過程と欠測過程の同時分布に対するモデルを考える必要がある。その方法は、同時分布 $f(y_i, d_i; \theta, \gamma)$ をどのように分解するかによって、セレクションモデル (selection models) とパターン混合モデル (pattern mixture models) の 2 つのアプローチにわけることができる (Little, 1995; Little and Rubin, 2001)。ここでは、セレクションモデルについて考える (パターン混合モデルについては、6.5 で述べる)。

セレクションモデルは、式 (1) の同時分布の分解にもとづいたアプローチである。例えば、表 1 のデータにおいて、ある時点 t ($t \geq 2$) で脱落するかどうかその時点での (一部の対象者では欠測している) 結果変数 Y_{it} も影響すると仮定した以下のようなモデルを考える。

$$\log \frac{\Pr(d_i = t \mid d_i \geq t, y_{i1}, \dots, y_{it-1}, y_{it})}{\Pr(d_i > t \mid d_i \geq t, y_{i1}, \dots, y_{it-1}, y_{it})} = \gamma_0 + \gamma_1 \text{dose}_i + \gamma_2 y_{i1} + \dots + \gamma_t y_{it-1} + \gamma_{t+1} y_{it} \quad (17)$$

式 (17) で、 $\gamma_2 = \dots = \gamma_{t+1} = 0$ であれば、脱落メカニズムが MCAR に相当し、 γ_2 から γ_t のいずれかがゼロでなく、 $\gamma_{t+1} = 0$ であれば、脱落メカニズムが MAR に相当する。 $\gamma_{t+1} \neq 0$ であれば、欠測メカニズムは MNAR である。

4.1 で述べたように、 $\gamma_{t+1} \neq 0$ の場合には、観察データに対する同時分布 $f(y_i^o, d_i; \theta, \gamma)$ は、測定過程に対する周辺分布 $f(y_i^o; \theta)$ と欠測過程に対する分布 $f(d_i \mid y_i^o; \gamma)$ の積に分解することはできない。すなわち、式 (6) を直接評価しなければならない。脱落メカニズムが MNAR の場合を

考慮するために、そのような同時モデルの提案がいくつかなされている (Diggle and Kenward, 1994; Hogan and Laird, 1997; Verbeke and Molenberghs, 2000). それらのモデルのパラメータ推定には、シンプレックス法やEMアルゴリズムなどが利用されている. しかしながら、それらのモデルは、既存の統計ソフトで容易に実行できないという実際上の問題や式(6)に含まれる積分計算が複雑という技術的な問題だけでなく、式(17)のパラメータ、特に γ_{t+1} の識別可能性という本質的な問題が存在する (Fitzmaurice et al., 1996). また、そのような同時モデルから得られる結果が妥当であるためには、測定過程に対するモデル $f(y_i^o; \theta)$ だけでなく、式(17)で表現される欠測過程に対するモデルも正しいことを前提としている. 一般に、欠測理由に関する十分な外部情報がない限り、式(17)のような観察されていないデータにもとづいたモデルの妥当性の評価は困難である. したがって、データから検証不能な仮定に統計的推測が強く依存している以上、得られる結果の解釈は慎重にならざるを得ない. 脱落過程に対するいくつかのモデルのもとで最終的な結果がどのように変化するかを検討する感度解析 (sensitivity analysis) を行うことが必要である (Copas and Li, 1997; Laird, 1994; Verbeke and Molenberghs, 2000).

6.1 で述べた重み付き解析に関しても、そのモデルをMNARの場合に拡張した方法が提案されている (Rotnizky et al., 1998; Scharfstein et al., 1999). この方法は、上述の同時分布にもとづく方法が完全なパラメトリックなアプローチであるのに対して、同時分布を特定しないセミパラメトリックなアプローチである. その方法は、欠測した結果変数による選択バイアスの大きさを表す式(17)のパラメータ γ_{t+1} の値をデータから推定するのではなく、その値を適当な範囲で設定し、それぞれのもとでの重みを推定し、結果の感度解析を行うアプローチである.

6.5 パターン混合モデル

セレクションモデルとはことなる以下のような同時分布 $f(y_i, d_i)$ の分解に基づいて、欠測メカニズムを考慮するアプローチがパターン混合モデルである (Little, 1993).

$$f(y_i, d_i) = f(y_i | d_i) f(d_i) \quad (18)$$

式(18)の右辺の第1項は脱落パターン (例えば、表2で示した4つのパターン) ごとの結果変数の分布であり、 Y_i の周辺分布は、それぞれのパターンが右辺の第2項を重みとして混在したものととらえることができる.

セレクションモデルとパターン混合モデルは、同じ同時分布に対する分解方法がことなるだけなので、理論的には、一方のモデルを他方のモデルで表現することが可能である. しかしながら、解析を簡略化、あるいはパラメータを識別可能にするための仮定は、2つの分解方式でことなるものが導かれるので、一般にこの2つのアプローチはことなつた解析となる. ただし、欠測メカニズムがMCARであれば、式(18)は式(1)のセレクションモデルと同じ形 $f(y_i) f(d_i)$ となり、どちらも同じモデルとなる.

式(18)のもとでは、観察データに対する同時分布は、以下のように表現される.

$$\begin{aligned} f(y_i^o, d_i) &= \int f(y_i, d_i) dy_i^m = f(d_i) \int f(y_i^o, y_i^m | d_i) dy_i^m \\ &= f(d_i) \int f(y_i^o | d_i) f(y_i^m | y_i^o, d_i) dy_i^m \end{aligned} \quad (19)$$

脱落した対象者 (脱落パターン $d_i < T+1$) に対しては, $f(y_i^m | y_i^o, d_i)$ に関する情報が観察データに含まれていないので, パターン混合モデルによって解析を行うためには, いくつかのパラメータを識別可能にするための何らかの制約が必要となる.

表 1 のデータに対して, 時点ごとの無月経割合の単純な推定をパターン混合モデルで行う (ただし, 150 mg 群のみを考える). 脱落パターンごとに推定すべきパラメータ $\theta_t^{(d)}$ ($t=1, \dots, 4$) の推定値を表 8 に示す. 表中の \times は推定不能なパラメータを意味し, 1 番最後の行が関心のある時点ごとの無月経割合の (脱落パターンに関して平均した) 推定値 $\sum_d p^{(d)} \hat{\theta}_t^{(d)}$ (ただし, $p^{(d)}$ は各脱落パターンの観察割合) である.

表 8. 脱落パターンごとの識別可能なパラメータ (無月経割合の単純平均)

パターン	人数 (割合)	時点			
		1	2	3	4
1 ($d=2$)	99 (17.22%)	0.313	\times	\times	\times
2 ($d=3$)	87 (14.13%)	0.241	0.483	\times	\times
3 ($d=4$)	36 (6.26%)	0.250	0.361	0.639	\times
4 ($d=5$)	353 (61.39%)	0.162	0.298	0.479	0.535
		0.205	\times	\times	\times

\times は推定不能なパラメータを示す

Little (1993) は, 以下のような CCMV (complete case missing variable) と呼ばれる制約を提案している. ただし, $t \geq d_i, d_i < T+1$ である.

$$f(y_{it} | y_{i1}, \dots, y_{it-1}, d_i) = f(y_{it} | y_{i1}, \dots, y_{it-1}, d_i = T+1) \tag{20}$$

この制約は, ある脱落パターンの集団における脱落時点以降の (未観測の) 結果変数の条件付き分布が, 最後まで脱落なく観察された集団 ($d_i = T+1$) の条件付き分布と同じであることを意味している. この制約のもとでは, 表 8 の推定不能なパラメータは, 脱落パターン 4 ($d=5$) のデータを利用して推定 (予測) することになる. 例えば, 脱落パターン 1 ($d=2$) の時点 2 での無月経割合の推定値 $\hat{\theta}_2^{(2)}$ は, 式 (20) の仮定 $f(y_{i2}^{(2)} | y_{i1}^{(2)}, d_i = 2) = f(y_{i2}^{(5)} | y_{i1}^{(5)}, d_i = 5)$ より,

$$\hat{\theta}_2^{(2)} = E(y_{i2}^{(2)}) = \alpha_0^{(2)} + \alpha_1^{(2)} E(y_{i1}^{(2)}) = \hat{\alpha}_0^{(5)} + \hat{\alpha}_1^{(5)} E(y_{i1}^{(2)}) = 0.352$$

と計算される. ただし, $\alpha_0^{(d)}, \alpha_1^{(d)}$ は結果変数 $Y_{i2}^{(d)}$ の $Y_{i1}^{(d)}$ の上への回帰モデルにおけるパラメータである. 最終的に, 脱落パターンに関して平均した時点 2 での無月経割合の推定値は,

$$\sum_d p^{(d)} \hat{\theta}_2^{(d)} = (17.22 \times 0.352 + 14.13 \times 0.483 + 6.26 \times 0.361 + 61.39 \times 0.298) / 100 = 0.334$$

となる. 同様にして残りの $\theta_t^{(d)}$ を求めると, $\hat{\theta}_3^{(2)} = 0.522, \hat{\theta}_4^{(2)} = 0.566, \hat{\theta}_3^{(3)} = 0.566, \hat{\theta}_3^{(4)} = 0.601, \hat{\theta}_4^{(4)} = 0.601$ となり, 脱落パターンに関して平均した 3 時点目と 4 時点目の無月経割合の推定値は, それぞれ 0.504 と 0.548 となる. なお, この時点ごとの周辺平均 $\sum_d p^{(d)} \hat{\theta}_t^{(d)}$ の分散は, デルタ法を用いれば計算可能である. 付録 B に漸近分散の推定量を示す.

パラメータを識別可能にするための制約は、CCMV 制約以外にもいくつかのものが提案されている。以下の ACMV (available case missing variable) 制約がその一つである。

$$f(y_{it} | y_{i1}, \dots, y_{it-1}, d_i) = f(y_{it} | y_{i1}, \dots, y_{it-1}, d_i > t)$$

この制約は CCMV 制約とはことなり、識別不能なパラメータを予測するためにそれまでの時点で見られたすべてのデータを利用する。例えば、脱落パターン 1 ($d=2$) の時点 2 での無月経割合 $\theta_2^{(2)}$ の推定には、パターン 2、パターン 3、パターン 4 において観察された時点 1、時点 2 のデータを利用する。この制約は、欠測パターンが脱落の場合には、セレクションモデルのもとでの MAR の仮定に相当する (Molenberghs et al., 1998)。また、Little (1994) は、ある時点での欠測値がその時点での値にも依存することを許した、すなわちセレクションモデルにおける MNAR の状況に対応する制約を提案している。

Fitzmaurice and Laird (2000) は、式 (8) のロジスティックモデルにもとづいた時点ごとの無月経割合の推定をパターン混合モデルで行う方法を提案している。脱落パターン 1 (2 時点目で脱落) のグループでは、2 時点目以降のデータが存在しないので、そのグループにおける結果変数の経時変化を表すパラメータは推定不能である。したがって、いくつかの回帰パラメータが、脱落パターン間を通して共通であるという制約を入れなければならない。例えば、脱落パターンを示すダミー変数を D_{i1}, D_{i2}, D_{i3} (パターン 4 の最後まで観察された対象者の場合、 $D_{i1} = D_{i2} = D_{i3} = 0$) として、以下のようなモデルが考えられる。

$$\log \left(\frac{\Pr(Y_{it} = 1 | D_i)}{\Pr(Y_{it} = 0 | D_i)} \right) = \beta_1 + \beta_2 D_{i1} + \beta_3 D_{i2} + \beta_4 D_{i3} \\ + \beta_5 t_{it} + \beta_6 t_{it}^2 + \beta_7 dose_i + \beta_8 (t_{it} \times dose_i) + \beta_9 (t_{it}^2 \times dose_i) \quad (21)$$

式 (21) のモデルでは、脱落パターンの違いはベースラインでの反応割合に影響を与えるが、その経時変化には影響を与えないことを仮定したモデルである。式 (21) の回帰パラメータ β は GEE をあてはめることができる標準的な統計ソフトを用いれば求めることができる。式 (9) の相関構造を仮定した GEE モデルをあてはめた結果を表 9 に示す。 $\beta_2, \beta_3, \beta_4$ のパラメータ推定値から、ベースラインでの無月経割合は、最後まで観察されたグループ (パターン 4) よりも途中で脱落した集団のほうが高い傾向があり、その傾向は脱落時点が早期ほど顕著であることがわかる。したがって、表 1 のデータでは、状態の良い (反応あり) 対象者ほど早期に脱落しているといえる。この結果は、時点 1 で反応ありの対象者ほど時点 2 以降で脱落しやすい傾向があることを示した表 5 の結果と矛盾しない。

式 (21) の回帰パラメータ β の推定値と、群 (あるいは、共変量) ごとの各脱落パターンの観察割合が求まれば、最終的に、各群・時点における無月経割合の周辺平均を求めることができる。その結果を表 10 に示す。表 7 の重み付き解析や多重補完法の結果と比べて、いずれの群・時点においても無月経割合の推定値は高めに推定されている。これは、脱落した対象者ほど反応「あり」の傾向があることに起因していると思われる。しかしながら、各時点における群間差に関しては、表 7 の結果とほぼ同じである。

表 9. 式 (21) のモデルを表 1 のデータにあてはめた結果

変数	推定値	標準誤差	95%信頼区間
切片	-2.344	0.270	-2.874, -1.814
D_{i1} (パターン 1 v.s. 4)	0.689	0.186	0.325, 1.053
D_{i2} (パターン 2 v.s. 4)	0.453	0.158	0.143, 0.764
D_{i3} (パターン 3 v.s. 4)	0.387	0.185	0.025, 0.749
時点	0.638	0.219	0.209, 1.068
時点 ²	-0.014	0.041	-0.095, 0.067
治療群	-0.490	0.358	-1.190, 0.211
治療群 × 時点	0.726	0.304	0.130, 1.321
治療群 × 時点 ²	-0.137	0.058	-0.251, -0.024

表 10. 時点ごとの無月経割合の推定値とその群間差の推定値 (パターン混合モデル)

解析方法	時点	150 mg	100 mg	差	差の標準誤差	Z 値	p 値
モデル (21)	1	0.200	0.184	0.016	0.023	0.70	0.49
	2	0.380	0.288	0.091	0.027	3.37	0.001
	3	0.526	0.415	0.111	0.031	3.58	0.0003
	4	0.600	0.547	0.053	0.036	1.47	0.14
モデル (22)	1	0.201	0.185	0.016	0.023	0.70	0.49
	2	0.382	0.282	0.100	0.029	3.45	0.0006
	3	0.563	0.457	0.106	0.042	2.52	0.01
	4	0.674	0.648	0.026	0.057	0.46	0.65

式 (21) とはことなるパターン混合モデルも考えることができる。例えば、以下のようなモデルである (Fitzmaurice and Laird, 2000)。ただし、変数 D_i^* は、追跡途中で脱落したかどうかを表す指示変数 (脱落パターン 1, 2, 3 なら 1, パターン 4 なら 0 をとる変数) である。

$$\begin{aligned} \log \left(\frac{\Pr(Y_{it} = 1 | D_i)}{\Pr(Y_{it} = 0 | D_i)} \right) = & \beta_1 + \beta_2 D_{i1} + \beta_3 D_{i2} + \beta_4 D_{i3} \\ & + \beta_5 t_{it} + \beta_6 t_{it}^2 + \beta_7 dose_i + \beta_8 (t_{it} \times dose_i) + \beta_9 (t_{it}^2 \times dose_i) \\ & + \beta_{10} (D_i^* \times t_{it}) + \beta_{11} (D_i^* \times t_{it}^2) + \beta_{12} (D_i^* \times dose_i) \\ & + \beta_{13} (D_i^* \times t_{it} \times dose_i) + \beta_{14} (D_i^* \times t_{it}^2 \times dose_i) \end{aligned} \quad (22)$$

式 (22) のモデルは、脱落パターンの違いがベースラインでの反応割合に影響を与えるだけでなく、脱落者と最後まで観察された対象者で反応の経時変化が治療群ごとにことなることを許したモデルである。表 10 に式 (22) から求まる各群・時点における無月経割合の周辺平均の推定値を示す。式 (21) のモデルの結果と比べて、特に時点 3 と時点 4 における無月経割合の推定値がいずれの群においても高い値となっている。しかしながら、時点 4 においての群間差が小さくなっているのを除けば、モデル (22) を用いたとして群間差に関してはほぼ同じ結果となっている。

パターン混合モデルは、セレクションモデルとはことなり、欠測メカニズムに対する明示的な仮定をおくことなく、欠測メカニズムを考慮した解析 (non-ignorable analysis) を行うことがで

きるが、いくつかの問題もある。1つ目の問題は、欠測パターンの分類方法である。脱落時点が同じ対象者は何らかの意味でよく似ていると思われるので、表8で示したように脱落時点でパターン进行分类することが多いが、本来は追跡不能、無効、副作用による脱落などのように脱落理由ごとに分類してもかまわない。しかしながら、パターン混合モデルをあてはめるためには各パターンに十分な対象者がいることが前提なので、パターン数が多く、各パターン内での人数は少ない場合には、いくつかのパターンをまとめなければならず、誤分類の問題が生じる。2つ目の問題は、パターン混合モデルから得られる結果は、パラメータ推定のためにどのような制約を用いたかに完全に依存する点である。パラメータを識別可能とするための制約は、一般にデータからはその妥当性を検証することはできないので、表10に示したようにいくつかのモデルを仮定して得られる結果の頑健性を検討しておくことが重要である (Little, 1994; Verbeke and Molenberghs, 2000)。

6.6 Latent variable モデル

欠測メカニズムが MNAR の場合の解析方法として、結果変数の測定過程と欠測過程の両方に影響する潜在変数 (latent variable) を仮定したアプローチもある (Follmann and Wu, 1995; Wu and Bailey, 1988; Wu and Carroll, 1988)。このアプローチは、ある時点での欠測が (将来の) 結果変数の値自体ではなく、結果変数とも関連する個人の潜在的な特性に依存すると考えたモデルで、Little (1995) は non-ignorable random-coefficient-based drop-out と呼んでいる。

対象者 i に対する潜在変数を u_i とし、 u_i を与えたもとの、脱落は結果変数と独立であると仮定すると、観察データ (Y_i^o, D_i) に対する同時分布 $f(y_i^o, d_i)$ は、以下のように表現できる。

$$\begin{aligned} f(y_i^o, d_i) &= \iint f(y_i^o, y_i^m, d_i, u_i) dy_i^m du_i \\ &= \iint f(y_i | u_i) f(d_i | u_i) f(u_i) dy_i^m du_i \\ &= \int f(y_i^o | u_i) f(d_i | u_i) f(u_i) du_i \end{aligned} \quad (23)$$

式 (23) で、結果変数の測定過程に対するモデル $f(y_i^o | u_i)$ は、そのモデルに潜在変数 (変量効果) が含まれる以外はセレクションモデルと同じである。脱落過程 $f(d_i | u_i)$ に対しては、例えば、個人ごとの結果変数の経時変化の切片 (random intercept) と傾き (random slope) の値を u_i と考え、式 (10) の結果変数 y_{it}, \dots, y_{it-1} のかわりに u_i を用いたモデルなどが提案されている。この latent variable モデルにもとづくアプローチは、既存の統計ソフトで容易に実行できないという実際上の問題はあっても、結果変数の測定にかなりの測定誤差がともない、脱落するかどうか結果変数の値が直接依存すると考えるよりも、疾病の進行度合いのような潜在的な変数が関係していると思われる状況の解析に有用なアプローチである (Little, 1995)。

7. 共変量に欠測がある場合の解析

ほとんどの疫学研究では、交絡が常に存在すると考えられるので、交絡因子で調整した解析を行うことが多い。そのような場面では、結果変数の欠測だけでなく、共変量の欠測の問題も重要である。共変量に欠測がある場合の解析手法に関しては、さまざまな種類の共変量が欠測する可能性

があるため、これまでの章で述べた結果変数の欠測の場合と比べて、いくぶん問題の定式化が複雑になる。しかしながら、その基本的な考え方は、これまで述べてきたものと同じである (Fleiss et al., 2003; Greenland and Finkle, 1995; Little, 1992)。

共変量に欠測がある場合の標準的な解析手法 (あるいは、標準的な統計パッケージを用いた解析) は、すべての共変量が観察された対象者のみを解析対象とする complete case 解析である。この解析方法は、5.1 で述べたように、解析の効率が低下するだけでなく、欠測メカニズムが MCAR でない限り、バイアスのない結果を導かない。ある共変量の欠測がたまたま偶然に生じていると考えることができる場面もあるかもしれないが、ある共変量が欠測するかどうかに関係された共変量や結果変数が関係する場合 (MAR) や欠測した変数自体が関係する場合 (MNAR) には、complete case 解析は妥当な解析ではない。

Vach and Blettner (1991) は、ケース・コントロール研究において交絡因子が欠測している状況で、いくつかの欠測メカニズムのもとで、complete case 解析がオッズ比の推定に及ぼすバイアスの程度について議論している。欠測メカニズムが MAR の場合、6.1 で述べた重み付き解析 (Lipsitz et al., 1998; Robins et al., 1994)、6.2 で述べた多重補完法による解析 (Xie and Paik, 1997) が、共変量が欠測した場合にも提案されている。また、欠測メカニズムが MNAR の場合のセレクションモデルにもとづく解析方法も提案されている (Ibrahim et al., 1999; Lipsitz et al., 1999)。

8. 欠測データ解析におけるいくつかの注意点

本総説では、欠測データが存在する場合のいくつかの統計解析手法について述べたが、欠測データが存在したとしてもこれらの方法を用いれば、その結果が常に妥当であることを意味しない。どんな統計解析手法にも前提が存在し、その前提が成立していれば、得られる結果は理論どおり妥当であるかもしれないが、前提が大きく崩れるようであれば、適用した解析手法の頑健性が問題となる。欠測データに対する解析方法の問題は、解析のために必要な欠測メカニズムに関する前提条件の多くが、データからは検証不能な点である。データから検証不能な仮定に統計的推測が強く依存している以上、特に欠測データが多量の場合には、解析は確証的な結論を導かず、得られる結果の解釈は慎重にならざるを得ない。欠測データ解析を行うためには、どういう理由で欠測が生じたのか、データの欠測に至る経緯を代表する変数を可能な限り測定しておくことが大事であり、また、欠測メカニズムに関するいくつかのシナリオのもとで結果の感度解析を行うことが重要である。

すべての臨床・疫学研究にあてはまることであるが、そもそも欠測データが存在しないように、研究を計画・実施する努力、研究に参加した対象者は研究終了まで追跡する義務を果たさない限り、どのような解析方法を用いたとしてもその結果を積極的に解釈することは危険である。特に、ランダム化臨床試験は、明確な仮説が存在するもとで、その仮説を検証するための実験である。仮説検証に関して確証的側面が強い第 III 相試験においては、脱落が重度、あるいは多量であることは、適用する解析手法の問題ではなく、その臨床試験の質の問題、あるいはその臨床試験が不成立になっていることを意味するかもしれない。

本総説で紹介した解析手法を含むほとんどの欠測データに対する解析方法は、「もし欠測が存在せず、すべてのデータが観察されていた場合の結果」を求めていることにも注意が必要である。例えば、重篤な疾患に対するQOL (quality of life) の測定や高齢者に対するADL (activity of daily living) の測定場面では、死亡による欠測が起こりえる。一般に、そのような状況での死亡は不可避であり、死亡という現象を制御可能でない限り、すべての対象者が生存していた場合の結果を求めることには意味がないかもしれない (Frangakis and Rubin, 2002)。また、綿密な研究計画を誠実に実施したとしても、薬剤の副作用による投薬中止、用法用量が守れないなどのプロトコル逸脱例は存在しえる。このような問題は、市販後にも同じことが起こると予想され、それ自体が薬効表示されるべきものである。少なくとも市販ということを考えて臨床試験であれば、ケースバイケースの議論になるかもしれないが、副作用による投薬中止などは薬剤の有効性に関しては「無効」と判断すべきである。あるいは、そのような予測ができるように、脱落や中止理由をこれまでの臨床試験データなどから十分吟味して、その情報を欠測メカニズムに反映させた解析を行う必要がある。

最後に、本総説で述べたいくつかの解析手法、特に、6章で述べた欠測メカニズムを考慮した解析手法をさまざまな種類の実際のデータにあてはめ、その挙動を比較検討していくことも重要である。どのような程度・種類・パターンの欠測データであれば、どの解析方針が適しているのかを実際のデータ解析を通して検討していくことが今後重要になってくると思われる。

謝 辞

本総説の一部は朝倉書店から2005年に刊行される『臨床試験ハンドブック (丹後俊郎・上坂浩之編)』, 22章「不完全データの解析」から抜粋した。掲載を許可いただいた朝倉書店編集部, 国立保健医療科学院の丹後俊郎先生に感謝いたします。また、本総説の一部は、科学研究費補助金 (基盤研究 (A)(1)) 課題番号 16200022 の補助を受けた。本総説の草稿を読んで貴重なコメントをいただいた東京大学大学院医学系研究科の山口拓洋先生, 土居主尚氏に深く感謝いたします。

参考文献

- Barnerd, J. and Rubin, D. B. (1999). Small sample degrees of freedom with multiple imputation. *Biometrika* 86, 948-955.
- Cary, V. C., Zeger, S. L. and Diggle, P. J. (1993). Modelling multivariate binary data with alternating logistic regressions *Biometrika* 80, 517-526.
- Copas, J. B. and Li, H. G. (1997). Inference from non-random samples (with discussion). *Journal of the Royal Statistical Society, Series B.* 59, 55-96.
- Diggle, P. J., Heagerty, P., Liang, K-Y. and Zeger, S. L. (2001). *Analysis of Longitudinal Data*, 2nd ed. Oxford University Press.
- Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* 43, 49-93.

- Fitzmaurice, G. M. and Laird, N. M. (2000). Generalized linear mixture models for handling nonignorable dropouts in longitudinal studies. *Biostatistics* 1, 141-156.
- Fitzmaurice, G. M., Laird, N. M. and Zahner, G. E. P. (1996). Multivariate logistic models for incomplete binary response. *Journal of the American Statistical Association* 91, 99-108.
- Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2004). *Applied Longitudinal Analysis*. John Wiley and Sons.
- Fleiss, J. L., Levin, B and Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*, 3rd ed. John Wiley and Sons.
- Follmann, D. and Wu, M. C. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* 51, 151-168.
- Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* 58, 21-29.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*. 85, 398-409.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall.
- Gillings, D. and Koch, G. (1991). The application of the principle of intention-to-treat to the analysis of clinical trials. *Drug Information Journal* 25, 411-424.
- Greenland, S and Finkle, W. D. (1995). A critical look at methods for handling covariates in epidemiologic regression analyses. *American Journal of Epidemiology* 142, 1255-1264.
- Heitjan, F. and Little, R. J. A. (1991). Multiple imputation for the fatal accident reporting system. *Applied Statistics* 40, 13-29.
- Heyting, A., Tolboom, J. T. B. M. and Essers, J. G. A. (1992). Statistical handling of dropouts in longitudinal clinical trials. *Statistics in Medicine* 11, 2043-2061.
- Hogan, J. W. and Laird, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* 16, 259-272.
- Hogan, J. W., Roy, J. and Korkontzelow, C. (2004). Tutorials in Biostatistics: Handling drop-out in longitudinal studies. *Statistics in Medicine* 23, 1455-1497.
- Horton, N. J. and Lipsitz, S. R. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician* 55, 244-254.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite population. *Journal of the American Statistical Association* 47, 663-685.

- Ibrahim, J. G., Lipsitz, S. R. and Chen, M-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is nonignorable. *Journal of the Royal Statistical Society, Series B.* **61**, 173-190.
- Kenward, M. G. and Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science* **13**, 236-247.
- 厚生省医薬安全局審査管理課長 (1998). 「臨床試験のための統計的原則」について (平成10年11月30日医薬審第1047号). <http://www.nihs.go.jp/dig/ich/eindex.html>.
- Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine* **7**, 305-315.
- Laird, N. M. (1994). Discussion to Diggle, P. J. and Kenward, M. G.: Informative drop-out in longitudinal data analysis. *Applied Statistics* **43**, 84.
- Lavori, R. W., Dawson, E. and Shera, D. (1995). A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in Medicine* **14**, 1913-1925.
- Liang, K-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.
- Lipsitz, S. R., Parzen, M. and Ewell, M. (1998). Inference using conditional logistic regression with missing covariates. *Biometrics* **54**, 295-303.
- Lipsitz, S. R., Ibrahim, J. G., Chen, M-H. and Peterson, H. (1999). Non-ignorable missing covariates in generalized linear models. *Statistics in Medicine* **18**, 2435-2448.
- Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association* **87**, 1227-1237.
- Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125-134.
- Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471-483.
- Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**, 1112-1121.
- Little, R. J. A. and Rubin, D. B. (2001). *Statistical Analysis with Missing Data*, 2nd ed. John Wiley and Sons.
- Machin, D., Farley, T., Busca, B., Campbell, M. and D'Archangues, C. (1988). Assessing changes in vaginal bleeding patterns in contraceptive women. *Contraception* **38**, 165-179.
- Molenberghs, G., Nichiels, B., Kenward, M. G. and Diggle, P. J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica* **52**, 153-161.

- Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846-866.
- Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106-121.
- Rosembaum P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41-55.
- Rothman, K. J. and Greenland, S. (1998). *Modern Epidemiology*, 2nd ed. Lippincott-Raven.
- Rotnitzky, A. and Robins, J. M. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika* **82**, 805-820.
- Rotniyzky, A., Robins, J. M. and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321-1339.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons.
- Rubin, D. B. (1991). Multiple imputation in health-care databases: an overview and some applications. *Statistics in Medicine* **10**, 585-598.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall.
- Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research* **8**, 3-15.
- Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association* **94**, 1096-1146.
- Vach, W. and Blettner, M. (1991). Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *American Journal of Epidemiology* **134**, 895-907.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag.
- Ware, J. H. (2003). Interpreting incomplete data in studies of diet and weight loss. *New England Journal of Medicine* **348**, 2136-2137.
- Wu, M. C. and Bailey, K. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine* **7**, 337-346.