原 著

中止・脱落の理由を考慮したIPCW法による 臨床試験データの解析

Analysis of Clinical Trials with Missing Data by Modeling Causes of Drop-Outs

嘉田晃子^{*1,*2}・松山 裕^{*1}・佐藤俊哉^{*1}
Akiko Kada^{*1,*2}, Yutaka Matsuyama^{*1}, and Tosiya Sato^{*1}

^{*1}京都大学大学院医学研究科社会健康医学系専攻医療統計学

^{*2}住友製薬株式会社開発本部

^{*1}Department of Biostatistics, Kyoto University School of Public Health

^{*2}Clinical Development Division, Sumitomo Pharmaceuticals Co.Ltd.

e-mail^{*3}:kada@hsp.ncvc.go.jp

In clinical trials, some patients are dropped-out of the trials by the different causes and their outcomes may happen to be missing. In such a case, analyses ignoring the missing mechanisms lead a biased estimator. One of the methods for taking them into account is the IPCW (Inverse Probability of Censoring Weighted) method, which accounts for the observed past histories of time-dependent factors that are predictors of drop-outs and are correlated with the outcomes. In this method, the probability of being censored is usually modeled via a logistic regression without considering the causes of drop-outs. We developed the IPCW method including the causes of drop-outs such as improvement or aggravation. As an example, we analyzed the data from a randomized clinical trial of a drug for osteoporosis. To evaluate the efficacy, the difference of the rate of increasing in the lumbar vertebral mineral density at 48 weeks from baseline in the two dose groups were estimated. We compared the results of the proposed IPCW methods with those of the usual IPCW methods, complete-case analysis, and LOCF method. Although the results of the two IPCW estimators did not change much, the missing mechanisms could be modeled reasonably and could be interpreted clinically by the proposed method compared with the usual IPCW method.

Key words: Causes of drop-outs, Missing at random, Missing data, Randomized clinical trial, Weighted estimating equations.

1. は じ め に

臨床試験では、しばしば観察期間の途中で被験者が中止・脱落することがある。中止・脱落は、有害事象の発現や症状の悪化、改善、あるいは来院しなくなったことなどによって起こり、その

Received April 2002. Revised November 2002. Accepted March 2003.

^{*3} 現在の所属：国立循環器病センター。

結果, エンドポイントである結果変数が欠測してしまう。臨床試験のための統計的原則(厚生省医薬安全局, 1998)のなかに、「欠測値は、臨床試験において偏りを起こし得る代表的な原因である。したがって、データ収集及びデータマネジメントに関しては、治験実施計画書が要求する事項について、すべてを満たすようあらゆる努力をすべきである。」とあるように、中止・脱落をできるだけ少なくするよう、臨床試験を実施する段階で工夫をすべきである。しかし、薬剤の副作用が起こった場合や症状が悪化した場合、薬剤の投与は中止される。このような中止・脱落によりエンドポイントに欠測が起こっている状況において、どのようにして欠測が起こったかを考えずに解析を行うと、一般に治療効果の推定にバイアスが生じることが知られている (Greenland and Finkle, 1995)。

欠測の状況と解析の関係を明確にするため、欠測を欠測メカニズムから分類すると、次の3つに分けられる (Rubin, 1976; Little and Rubin, 1987)。欠測するかどうかは、1) MCAR (Missing Completely at Random) : 観察されたデータにも観察されなかったデータにも依存しない場合、2) MAR (Missing at Random) : 観察されたデータにのみ依存する場合、3) non-ignorable : 観察されたデータと観察されなかったデータ両方に依存する場合である。1) の MCAR の場合は、解析対象集団はランダム化された元の集団にランダムに欠測が起こったものと考えられるので、欠測メカニズムを無視して解析しても効率の損失はあるものの治療効果の推定にバイアスは生じない。一方、2) や 3) の場合には、被験者特有の(既知、あるいは未知の)特性に応じた選択的な欠測が生じているため、その欠測メカニズムを無視して解析すると治療効果の推定にバイアスが生じる (Rubin, 1976; Little and Rubin, 1987; Greenland and Finkle, 1995)。

先に述べた臨床試験における中止・脱落においては、欠測メカニズムとして MCAR を想定できることは一般には稀であり、MAR または non-ignorable の状況が生じているのが通常である。したがって、欠測をともなう臨床試験データの解析を行う際には欠測メカニズムを何らかの形で考慮する必要がある。

欠測値が存在する場合の解析手法として、これまでいくつかの方法が提案されてきている (Little and Rubin, 1987; Wu and Carroll, 1988; Wu and Bailey, 1989; Little, 1993; Diggle and Kenward, 1994; Greenland and Finkle, 1995; Little, 1995; Fitzmaurice, et al., 1996)。欠測メカニズムが MAR の場合に、バイアスのない治療効果の推定を行う方法のひとつとして、欠測するまでの個人ごとの履歴を考慮して欠測確率をモデル化する IPCW (Inverse Probability of Censoring Weighted) 法と呼ばれる解析方法が近年提案されている (Robins, et al., 1995)。この IPCW 法は、被験者が観察される確率の逆数で被験者を重み付けた解析であり、欠測値がある場合の解析として通常よく用いられる complete-case 解析を MAR の状況でも妥当な推測を導くことができるように拡張した方法と捉えることができる。

通常の IPCW 法では欠測理由の違いは無視して欠測するかどうかのみがモデル化されるが (Robins, et al., 1995; Rotnitzky and Robins, 1995)、多くの臨床試験において、中止・脱落の理由によって、欠測までの過程が異なる場合がある。本論文では、中止・脱落の理由までも考慮した IPCW 法を提案し、その方法を用いて骨粗鬆症治療薬の臨床試験データ(藤田ら, 1993)を解析し、通常の IPCW 法、欠測値を最終のデータで代用する方法 (Last Observation Carried Forward :

LOCF), complete-case 解析と比較する。

2 節では骨粗鬆症治療薬の臨床試験の概略と、発生した中止・脱落の様子を示す。3 節では中止・脱落の理由を考慮した IPCW 法の統計モデルを示し、どのように骨粗鬆症治療薬の治療効果を推定するかについて説明する。4 節では、中止・脱落の理由を考慮した IPCW 法の解析結果を、通常の IPCW 法、LOCF 解析、complete-case 解析とともに示し、最後に 5 節でそれらを比較し、考察する。

2. 骨粗鬆症治療薬の臨床試験

この臨床試験は、1990 年 7 月から 1992 年 6 月までの期間に行われた骨粗鬆症治療薬のランダム化二重盲検比較試験であり、低用量群、高用量群および対照薬群において骨塩量の増加を比較した(藤田ら, 1993)。観察期間は 48 週間で、骨塩量や臨床検査は 12 週ごとに、疼痛は 4 週ごとに観察された。骨塩量は腰椎に 2 種類の X 線をあて評価する DXA 法で行われた。本論文では、対照薬群を除いた低用量群および高用量群の 2 群比較に着目した。治療効果の指標は、48 週時点での低用量群と高用量群の骨塩量変化率 $((48 \text{ 週時点の骨塩量} - \text{開始時骨塩量}) / \text{開始時骨塩量} \times 100(\%))$ の群間差である。

開始時の被験者数は、低用量群および高用量群とも 137 名であった。そのうち、36 週以上観察され、第 2 腰椎から第 4 腰椎のうち 3 椎体、または隣り合わせた 2 椎体で骨塩量の測定が可能であった被験者(低用量群 83 名、高用量群 85 名)が、本来の臨床試験における骨塩量に関する解析対象者であった。本論文では、観察期間が 36 週未満のため解析から除外された被験者を加え、低用量群 101 名、高用量群 104 名を骨塩量の解析対象とした。

この試験においては、医師の判断による試験の中止を「中止」、患者の自己都合によって来院しなくなった場合を「脱落」とし、それぞれ理由が調査された。具体的には、副作用発現、合併症併発・悪化、服薬拒否、来院や服薬の不規則などの理由による中止と、自覚症状改善、効果なし、合併症併発・悪化、家事・仕事多忙、患者の非協力、転医、転院などの理由による脱落があった。これらの中止と脱落の理由を、被験者の状態の改善によるのか、あるいは悪化によるのかという観点から、1) 自覚症状改善、2) 副作用発現、合併症併発・悪化、効果なし、その他の 2 つに分類した。以降、それぞれ理由 1)、理由 2) と表す。その他の理由については、中止・脱落に影響があると考えられる変数と中止・脱落の関係を検討したところ、悪化による中止・脱落と類似していることが示唆されたため、理由 2) に含めた。なお、本論文における中止・脱落とは、「骨塩量の測定が一度欠測すれば、それ以降の測定も欠測する」という単調な欠測を意味する。単調でない骨塩量の欠測がある被験者は低用量群と高用量群あわせて 45 名であったが、経時的に骨塩量の大幅な増減は認められなかったことから、欠測時点の前に測定された値で欠測データを補完した。

表 1 に、各投与量群におけるそれぞれの理由で中止・脱落した被験者数を示す。低用量群では、試験開始時に 101 名であったが、12 週までに理由 1) で 5 名、理由 2) で 7 名、12 週から 24 週間にそれぞれ 1 名、2 名、24 週から 36 週間にそれぞれ 1 名、2 名、36 週から 48 週間にそれぞれ 0 名、2 名の中止・脱落があり、最終的に予定通り 48 週間観察された被験者は 81 名であった。一方、高用量群では試験開始時に 104 名であったが、12 週までに理由 1) で 2 名、理由 2) で

表 1. 投与量群ごとのそれぞれの理由による中止・脱落数

		観察週				
		0	12	24	36	48
低用量群	理由 1) で中止・脱落	0	5	1	1	0
	理由 2) で中止・脱落	0	7	2	2	2
	観察人数	101	89	86	83	81
高用量群	理由 1) で中止・脱落	0	2	1	2	0
	理由 2) で中止・脱落	0	9	2	3	1
	観察人数	104	93	90	85	84

9 名, 12 週から 24 週の間それぞれ 1 名, 2 名, 24 週から 36 週の間それぞれ 2 名, 3 名, 36 週から 48 週の間それぞれ 0 名, 1 名の中止・脱落があり, 最終的に予定通り 48 週間観察された被験者は 84 名であった.

3. 中止・脱落の理由を考慮した IPCW 法

被験者 $i(i = 1, \dots, N)$ に対する観察時点 t 週 ($t = 0, 12, 24, 36, 48$) における骨塩量変化率を Y_{it} , 被験者 i に対する投与量群を含むベースライン共変量ベクトルを $\mathbf{X}_i$, t 週における疼痛の有無などの時間依存性共変量ベクトルを $\mathbf{V}_{it}$ とし, $\mathbf{W}_{it} = (\mathbf{V}_{it}^T, Y_{it})^T$ ($t = 12, 24, 36, 48$), $\mathbf{W}_{i0} = (\mathbf{X}_i, \mathbf{V}_{i0}, Y_{i0})$ とする. また, 被験者 i に対する投与量群を表す指示変数を Z_i とし, 低用量群の場合に $Z_i = 1$, 高用量群の場合に $Z_i = 0$ とする.

48 週目での骨塩量変化率 Y_{i48} の条件付期待値に対して, 次のモデルを想定する.

$$E(Y_{i48} | Z_i) = \beta_0 + \beta_1 Z_i \quad (1)$$

(1) 式において, β_1 は被験者全員が 48 週目まで観察された場合の低用量群と高用量群の骨塩量変化率の差を表す. 解析の目的は, 一部の被験者において Y_{i48} が観察されていないときに, 以下で説明する IPCW 法により治療効果 β_1 の大きさを推定することである.

中止・脱落の理由を考慮するために, 中止・脱落の状態を表す変数 R_{it} を以下のように定義する. 被験者 i が t 週目で観察されていれば $R_{it} = 0$, 理由 1) で中止・脱落すれば $R_{it} = 1$, 理由 2) で中止・脱落すれば $R_{it} = 2$ とする. ただし, すべての被験者に対して $R_{i0} = 0$ とし, 各時点で $j = 0, 1, 2$ のいずれかに分類されるとする.

欠測メカニズムについては次の 2 つの仮定をおく. ある時点における中止・脱落は, 投与量群を含むベースライン共変量とその時点直前までの時間依存性共変量, および結果変数(骨塩量変化率)で説明できること(ランダムな中止・脱落)と, 一度中止・脱落すると再び試験に参加することはないこと(単調な中止・脱落)である. 後者の単調な中止・脱落の仮定は, $R_{it} = j$ ならば $R_{i(t+12)} = j$ ($j = 1, 2; t = 12, 24, 36$) となることを意味する. また, 前者のランダムな中止・脱落の仮定は次のように表すことができる.

$$\Pr(R_{it} = j | R_{i(t-12)} = 0, \overline{\mathbf{W}}_{it}, Y_{it}, \dots, Y_{i48}) = \Pr(R_{it} = j | R_{i(t-12)} = 0, \overline{\mathbf{W}}_{it}) \quad (2)$$

ただし, $j = 1, 2$, $t = 12, 24, 36, 48$, $\bar{\mathbf{W}}_{it} = (\mathbf{W}_{i0}^T, \dots, \mathbf{W}_{i(t-12)}^T)^T$ である. ここで, $\bar{\mathbf{W}}_{it}$ は被験者 i に対する $(t-12)$ 週目までのすべての観察履歴(投与量群, ベースライン共変量, 時間依存性共変量, および結果変数)を表す.

(2) 式の仮定は, t 週直前まで観察されている被験者 i が t 週目で観察されるか, 理由 1) で中止・脱落するか, 理由 2) で中止・脱落するかは, 直前までの各被験者の履歴で決まり, t 週目以降の結果変数には依存しないことを意味する. Robins and Finkelstein (2000) は, この仮定を sequential ignorability of censoring と呼んでいる. このランダムな中止・脱落の仮定は, 欠測メカニズムが Rubin (1976) の MAR であれば満たされる.

(2) 式の被験者ごとの各時点における観察確率(あるいは, それぞれの理由による中止・脱落確率)は通常未知なので, データから推定しなければならない. IPCW 法では, (2) 式の右辺を λ_{ijt} とし, 次式のような関数 $h(\cdot | \cdot)$ が存在すると仮定する. ただし, $j = 1, 2$ であり, $h(\cdot | \cdot)$ は $\bar{\mathbf{W}}_{it}$ の関数, α は未知パラメータベクトルである.

$$\log \frac{\lambda_{ijt}(\alpha)}{1 - \lambda_{i1t}(\alpha) - \lambda_{i2t}(\alpha)} = h(\bar{\mathbf{W}}_{it} | \alpha) \quad (3)$$

これまで提案されてきている IPCW 法では, (3) 式の代わりにロジスティックモデルを想定しているが (Robins, et al., 1995; Rotnitzky and Robins, 1995; Rotnitzky et al., 1998), 本論文では, 観察されるか, 中止・脱落するか, 理由 1) で中止・脱落するか, 理由 2) で中止・脱落するか, 理由 3) で中止・脱落するかの 3 値をモデル化するため, polytomous ロジスティックモデル (Rothman and Greenland, 1998) を用いた. 具体的に 2 節で述べた骨粗鬆症治療薬の臨床試験データの解析においては, 次のようなモデルを仮定した.

$$\begin{aligned} \log \frac{\lambda_{ijt}(\alpha)}{1 - \lambda_{i1t}(\alpha) - \lambda_{i2t}(\alpha)} = & \alpha_{j1} + \alpha_{j2}t + \alpha_{j3}t^2 + \alpha_{j4}\text{投与量群} + \alpha_{j5}\text{年齢}_{i0} \\ & + \alpha_{j6}\text{骨折数}_{i0} + \alpha_{j7}\text{骨塩量}_{i0} + \alpha_{j8}\text{疼痛の有無}_{i(t-12)} \\ & + \alpha_{j9}\text{疼痛変化}_{i(t-12)} + \alpha_{j10}\text{骨塩量変化率}_{i(t-12)} \end{aligned} \quad (4)$$

ここで, $j = 1, 2$, $t = 12, 24, 36, 48$, $\alpha = (\alpha_{11}, \dots, \alpha_{1,10}, \alpha_{21}, \dots, \alpha_{2,10})$ である. 中止・脱落に影響を与える変数 $\bar{\mathbf{W}}_{it}$ としては, 時間に関連する変数と投与量群の他に, ベースライン共変量として年齢, 開始時骨折数, 開始時骨塩量を, 時間依存性共変量として疼痛の有無, 疼痛変化, 骨塩量変化率を用いた. なお, 疼痛については 4 週ごとに疼痛重症度が高度, 中等度, 軽度, なしの 4 段階で評価されたため, 疼痛の有無 $_{i(t-12)}$ と疼痛変化 $_{i(t-12)}$ を次のように定義した. すなわち, 疼痛の有無 $_{i(t-12)}$ は, $(t-12)$ 週から $(t-4)$ 週の間に疼痛があれば 1, なければ 0 をとる変数とし, 疼痛変化 $_{i(t-12)}$ は $(t-12)$ 週から $(t-4)$ 週における疼痛重症度の変化が悪化の場合は -1, 不変の場合は 0, 改善の場合は 1 をとる変数とし, いずれも $\mathbf{W}_{i(t-12)}$ に含めた. これらの説明変数は, 中止・脱落理由の内容や単変数による中止・脱落との関係から, 影響を及ぼす可能性が考えられるものを選択した. また, これらの中には観察途中で欠測データがみられたが, 欠測時点の前に得られているデータで補完し, 説明変数において欠測はないものとした. 以降, (4) 式を欠測理由を考慮したモデルと呼ぶ.

(4) 式から推定される α の最尤推定量を $\hat{\alpha}$, 被験者ごとの各時点での観察確率の推定値を $\lambda_{i0t} = 1 - \lambda_{i1t}(\hat{\alpha}) - \lambda_{i2t}(\hat{\alpha})$, 48 週間のその累積確率を $\pi_{i48}(\hat{\alpha}) = \prod \lambda_{i0t}(\hat{\alpha})$ とすると, すべての時点で観察された被験者(完了例)に対する次式の重み付き推定方程式を解くことにより, $\beta = (\beta_0, \beta_1)^T$ を推定する.

$$\sum_{i=1}^N \frac{I(R_{i48}=0)}{\pi_{i48}(\hat{\alpha})} \{Y_{i48} - (\beta_0 + \beta_1 Z_i)\} = 0 \quad (5)$$

ただし, $I(R_{i48}=0)$ は被験者 i が 48 週目に観察されているかどうかを表す指示変数で, 観察されていれば 1, されていなければ 0 となる. Robins, et al. (1995) は, $j=1$ の場合にモデル (1), (2), (3) のもとで, (5) 式の重み付き推定方程式を解くことでバイアスのない治療効果が推定できることを示している.

(5) 式の重み付き推定方程式では, 最終時点まで観察された被験者の通常の推定方程式に対する寄与が, その被験者の観察確率の逆数倍されている. すなわち, IPCW 法では, 最後まで観察された被験者 i の観察確率が $\pi_{i48}(\hat{\alpha}) = 0.25$ とすると, 48 週直前までの履歴 $\overline{W}_{i48}$ が同じで, 48 週直前に中止・脱落となった被験者がもう 3 人存在すると考え, 中止・脱落がなければ観察されていたはずの擬似集団内で通常の推定を行うことに相当する. このような被験者の観察確率の逆数で重み付けした推定量は, 標本調査の分野では Horvitz-Thompson 推定量として知られている (Horvitz and Thompson, 1952).

β_1 に対する 95%IPCW ワルド信頼区間は, 次式から求めることができる.

$$\hat{\beta}_1 \pm 1.96 \sqrt{\text{var}(\hat{\beta}_1)} \quad (6)$$

ただし, $\text{var}(\hat{\beta}_1)$ は $\hat{\beta}_1$ に対するロバスト(サンドウィッチ)分散 (Liang and Zeger, 1986) である. すべての計算は SAS[®] Release 8.02 (SAS Institute, 1999) で行った.

4. 骨粗鬆症治療薬の臨床試験の解析結果

2 節で述べた骨粗鬆症治療薬の臨床試験データにおいて, 完了例(48 週目まで観察)と中止・脱落例別に, 低用量群と高用量群ごとの骨塩量変化率の推移をみたところ, 完了例において高用量群の方が低用量群よりも若干高い推移を示した. 中止・脱落例においては, 両群とも骨塩量変化率の推移における変化は小さかった. これらのデータに対し, 欠測値がある場合の解析として通常よく行われている, 完了例のみを解析対象とした complete-case 解析と, 中止・脱落例も含めて最終のデータで欠測値を代用する LOCF 解析を実施した. 解析モデルは (1) 式と同様とした.

LOCF 解析は, 欠測値を何らかの値で補完する方法のうち, 最も単純な方法である. どちらの解析方法も, 欠測メカニズムとして MCAR を前提としている. また, LOCF 解析では, 中止・脱落以降, その時点の観察データと同じ値が最後まで続くことを前提としている. これらの前提がくずれた場合, 例えば一方の群で骨塩量が減少している人が多く中止・脱落している場合や, 逆に一方の群で骨塩量が増加している人が多く中止・脱落している場合, どちらの方法でも効果の推定にバイアスが入る.

表 2. Complete-case 解析(解析 1)と LOCF 解析(解析 2)の結果

解析方法	解析対象データ	解析対象者数	切片(高用量群)		差(低用量群 - 高用量群)				
			推定値	標準誤差	推定値	標準誤差	P 値	95%信頼区間 下限	上限
解析 1	48 週すべて観察	165	3.474	0.522	-1.062	0.745	0.156	-2.533	0.410
解析 2	36 週以降観察	168	3.419	0.516	-1.051	0.735	0.154	-2.502	0.399
	24 週以降観察	176	3.293	0.495	-0.954	0.708	0.180	-2.351	0.443
	12 週以降観察	182	3.105	0.490	-0.845	0.701	0.230	-2.228	0.539

表 2 に, complete-case 解析(解析 1)と LOCF 解析(解析 2)による解析結果を示す. 解析 2 においては, 最終観察データとして 36, 24, 12 週以降のそれぞれの場合を用いた 3 通りの解析を行った. 解析 1 では低用量群と高用量群の骨塩量変化率の差の推定値は, -1.062 , 95%信頼区間(95% C.I.) は, -2.533 から 0.410 であった. 解析 2 では 36 週以降観察された被験者の最終データを用いた場合, 差の推定値は, -1.051 , 95% C.I. は -2.502 から 0.399 であった. 24 あるいは 12 週以降観察された被験者を LOCF 解析の対象とした場合には, 被験者数の増加のため推定値の標準誤差が小さくなっているものの, 短い投与期間では骨塩量の増加が不十分な被験者が多いため, 24 週以降観察されたデータを用いた場合, 骨塩量変化率の差の推定値はそれぞれ -0.954 , および -0.845 と小さな値を示した.

次に, IPCW 法の結果を示す. 表 3 に (4) 式におけるパラメータの推定値を示す. 欠測理由のモデル化において, 開始時骨折数, 開始時骨塩量, 疼痛有無, 疼痛変化, および骨塩量変化率に対するパラメータは, (4) 式の $j=1$ の場合と $j=2$ の場合で異なる値を示し, 符号が反対のものも見られた. また, 欠測理由を考慮せずに, 中止・脱落の有無をモデル化した通常の IPCW 法も実施した. この場合, (4) 式は通常のロジスティックモデルとなり, 以降このモデルを, 欠測理由を考慮しないモデルと呼ぶ. 表 3 の右端に, 欠測理由を考慮しないモデルのパラメータの推定値

表 3. ロジスティックモデルのパラメータ推定値

	欠測理由を考慮したモデル				欠測理由を考慮しないモデル	
	理由 1 ($j=1$)		理由 2 ($j=2$)		推定値	標準誤差
	推定値	標準誤差	推定値	標準誤差		
切片	-0.568	5.482	-2.201	2.390	-1.810	2.119
t	0.745	2.574	-1.191	1.102	-0.518	0.979
t^2	-0.451	0.645	0.108	0.239	-0.049	0.218
投与量群	-0.896	0.803	-0.079	0.402	-0.265	0.356
年齢	-0.015	0.044	0.020	0.022	0.012	0.020
開始時骨折数	-2.046	1.425	-0.317	0.235	-0.431	0.241*
開始時骨塩量	2.423	2.781	0.434	1.421	1.021	1.249
直前の疼痛有無	-2.269	1.171*	-0.055	0.440	-0.439	0.382
直前の疼痛変化	0.399	0.849	-1.127	0.605*	-0.463	0.469
12 週前の骨塩量変化率	0.152	0.154	-0.198	0.091*	-0.108	0.078

*: $P < 0.10$ の変数

図 1. 被験者ごとの観察確率の分布 (欠測理由を考慮したモデルとしないモデルの比較)

を示す。影響のあった変数は、開始時骨折数のみであった。図 1 に、これら 2 つのモデルから求めた被験者ごとの観察確率の分布を示す。2 つのモデル間で、観察確率の推定値に大きな違いはなかったが、欠測理由を考慮したモデルの方が明らかに小さい値を示した被験者が数名見られた。これらの被験者は、早期に骨塩量の増加が大きく、疼痛が消失した被験者、あるいは早期に骨塩量の減少が大きかった被験者であった。

表 4. IPCW 法による効果の推定

モデル	解析 対象者数	切片 (高用量群)		差 (低用量群 - 高用量群)				
		推定値	標準 誤差	推定値	標準 誤差	P 値	95%信頼区間 下限	95%信頼区間 上限
欠測理由を考慮したモデル	165	3.287	0.542	-0.993	0.709	0.163	-2.393	0.408
欠測理由を考慮しないモデル	165	3.281	0.539	-0.953	0.704	0.178	-2.343	0.437

表 4 に、IPCW 法による治療効果の推定結果を示す。欠測理由を考慮したモデルにおいて、低用量群と高用量群の骨塩量変化率の差の推定値は -0.993、95% C.I. は -2.393 から 0.408 であった。Complete-case 解析の結果と比べ、差の推定値で 0.07 とわずかに縮小した。また、高用量群の骨塩量変化率の推定値についても、0.19 のわずかな減少がみられた。欠測理由を考慮しないモデルにおいても、骨塩量変化率の差の推定値は -0.953、95% C.I. は -2.343 から 0.437 であり、欠測理由を考慮したモデルによる結果と大きな違いはなかった。

5. 考 察

本論文では、中止・脱落による欠測をとまなう臨床試験データの解析に対し、中止・脱落の理由を考慮した IPCW 法を提案した。IPCW 法では通常、観察確率のモデル化において多くの説明変数を含め、予測の精度を上げて重みを推定することで、治療効果推定の効率が増す (Robins, et al., 1995)。我々は、各被験者の観察確率を求めるにあたって、ベースライン共変量に加え時間依存性共変量も説明変数としてモデルに含めた。その結果、観察確率の推定のために提案した polytomous ロジスティック回帰のパラメータ推定値は、「疼痛がなくなった場合は、症状が改善の理由で中止・脱落となる、骨塩量が減少した場合や疼痛重症度が悪化した場合は、症状の悪化やその他の理由で中止・脱落となる」と解釈でき、骨粗鬆症治療に関する常識とあつた結果となった。このことから、中止・脱落の理由によって欠測までの過程が異なっていたと考えられる。一方、これまで提案されてきた IPCW 法である理由を考慮しないモデルでは、骨塩量や疼痛と中止・脱落の関係は明確ではなかった。また、これらのことは、2つのモデルによる観察確率にやや乖離がみられた被験者が、早期に骨塩量の増加が大きく、疼痛が消失した被験者、あるいは早期に骨塩量の減少が大きかった被験者であつたことと対応していると考えられる。

治療効果の推定について、IPCW 法による骨塩量変化率の差の推定値は、わずかではあるが complete-case 解析より小さく、標準誤差も小さくなった。Complete-case 解析や LOCF 解析では、欠測メカニズムが MAR の場合には治療効果の推定にバイアスが入る。一方、IPCW 法は、解析対象者は complete-case 解析と同じ最終時点まで観察された被験者であるものの、それらの被験者をそれぞれの観察履歴に応じて重み付けている。したがって、被験者特有の何らかの特性に応じて選択的に中止・脱落が生じている場合には、そのような特性(結果変数、および中止・脱落に影響を与える要因)を考慮することにより、標準誤差が小さくなり、バイアスが減少すると考えられる。また、IPCW 法による骨塩量変化率の推定値そのものが complete-case 解析よりわずかに小さくなったことについても、同様に IPCW 法が、観察期間が短く、骨塩量の増加が不十分な被験者による影響を修正した結果と考えられる。

このように、IPCW 法は MAR の状況で中止・脱落による欠測がもたらすバイアスを減らす方向に働いていると考えられるものの、本論文で提案した欠測理由を考慮した IPCW 法の結果は、欠測理由を考慮しなかった場合の結果や、complete-case 解析の結果と大きな違いはなかった。これは、比較する 2 群で用いている薬剤は同じであるので、2 群間で中止・脱落までの過程が類似していたためと考えられる。このことは、2 群間で中止・脱落理由別の例数や中止・脱落時期が似ていることから裏付けられる。

以上のことから、今回のデータでは骨塩量変化率の差の推定値は大きく変わらなかったものの、欠測理由を考慮した IPCW 法では骨塩量変化率や疼痛について骨粗鬆症治療の常識とあつたモデル化が可能であつたので、欠測理由を考慮した解析が適切だと考えられた。このように、欠測メカニズムとして MAR が考えられる状況において、中止・脱落の理由を考慮した IPCW 法を用いて、中止・脱落による欠測が治療効果に及ぼす影響を検討することが可能であつた。

本論文では、治療効果の指標として最終時点での連続データを扱ったが、2 値データや生存時間を扱うことも可能であり、また最終時点のみではなく経時データを扱うことも可能で (Rotnitzky, et

al., 1998; Robins and Finkelstein, 2000), 多くの臨床試験で中止・脱落が起こった場合に IPCW 法は適用可能である。実際に適用するにあたっては, ランダムな中止・脱落という仮定はデータからは検証不能なので, 欠測理由のモデル化において, 欠測を説明できるような多くの変数を検討することが重要である。また, ランダムな中止・脱落という仮定に対しては, non-ignorable な欠測の状況を想定した感度解析 (Rotnitzky, et al., 1998) が有用であると思われる。

参考文献

- Diggle, P. J. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics*, **43**, 49-93.
- Fitzmaurice, G. M., Laird, N. M. and Zahner, G. E. P. (1996). Multivariate logistic models for incomplete binary responses. *Journal of the American Statistical Association* **91**, 99-108.
- 藤田拓男, 折茂 肇, 井上哲郎ら. (1993). 退行期骨粗鬆症に対するアルファカルシドールを対照としたエチドロン酸二ナトリウム (EHDP) の二重盲検比較試験. 臨床評価 **21**, 261-302.
- Greenland, S. and Finkle, W. D. (1995). A critical look at methods for handling missing covariates in epidemiologic regression analyses. *American Journal of Epidemiology* **142**, 1255-1264.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663-685.
- 厚生省医薬安全局. (1998). 臨床試験のための統計的原則. 医薬審第 1047 号.
- 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. (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. (1987). *Statistical analysis with missing data*. New York: Wiley & Sons Inc.
- Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.
- 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.
- Robins, J. M. and Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* **56**, 779-788.
- Rothman, K. J. and Greenland, S. *Modern Epidemiology*, Chapter 20. (1998). Philadelphia: Lippincott Williams & Wilkins.

- Rotnitzky, A. and Robins J. M. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika* **82**, 805-820.
- Rotnitzky, 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.
- SAS Institute Inc. (1999). SAS/STAT® User's Guide, Version 8. Cary, NC, USA, SAS Institute Inc.
- Wu, M. C. and Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* **45**, 939-955. Corr.: 1990, **46**, 889.
- Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175-188.