

Non-compliance をともなう二値経時観察データに対するランダム化にもとづく解析

京都大学大学院医学研究科 社会健康医学系専攻 医療統計学分野

松山 裕

E-mail: matuyama@pbh.med.kyoto-u.ac.jp

要約

治療を繰り返し実施した場合の二値経時観察データに対するランダム化にもとづく因果効果の推定について考える。特に、二値結果変数が繰り返し測定される場合の共通因果リスク差、およびリスク比に対する推定方法を提案する。この推定方法は、各個人に対する潜在結果変数に対して単純な構造モデルを仮定するものの、治療群間の比較可能性に関してランダム化以外の仮定は何も必要としない。提案する方法は、ITT 解析と同様に治療効果に関する因果帰無仮説の妥当な検定を与える。さらに、この方法は対象者が実際に受けた治療に関する情報 (non-compliance information) を利用でき、ランダムでない non-compliance の場合であっても、すべての個人が割り付け通りの治療に従った場合に観察されていたはずの真の治療効果 (因果治療効果) の推定値を与える。成人急性骨髄性白血病 (AML) 患者に対する寛解後療法 (地固め療法) において、感染症にともなう発熱抑制に対する M-CSF の効果を調べたプラセボ対照のランダム化臨床試験データに提案する方法を応用する。

1. はじめに

ランダム化臨床試験で対象者が割り付け通りの治療に従わないことがよくある。このような治療の不遵守 (non-compliance) が生じているデータに対して標準的に用いられる解析方法は、ITT (intention-to-treat) 解析である。ITT 解析の原則は、ランダム化後の治療状況がどうであろうと、ランダム化の対象となった全ての対象者を割り付けられた治療グループに属するとして解析する方針である。したがって、割り付けは新治療グループであったが実際にはコントロール治療を受けた対象者は、新治療を受けていないにも関わらず、解析の際には割り付け通り新治療グループのメンバーとして取り扱われる。

この ITT 解析は、「治療の割り付け後に起きた治療法変更、副作用による治療中止などの治療の不遵守はその治療が現実の臨床場面に導入された場合にも当然起こりえることであり、ITT 解析はその全てを治療方針として一括して解析している」と解釈でき、実践的な解析方針を前提としている¹⁾。また、治療グループ間の比較可能性がランダム化により保証されている限り、

治療効果に関する帰無仮説の妥当な検定を与えることも ITT 解析が標準的に用いられる理由の一つである。

ITT 解析で問題となるのは、治療効果の推定値が一般に過小評価されることである。ITT 解析による治療効果の推定値は、治療割り付けの平均的な効果を表すので、non-compliance 割合の増加とともに、すべての対象者が割り付け通りの治療に従った場合に観察されていたはずの真の治療効果（因果治療効果）からのズレ（バイアス）が大きくなる。さらに、コンプライアンスに影響を与える要因が、臨床試験と実際の現場ではことなることも問題である。例えば、臨床試験の段階では薬の副作用によりコンプライアンスが悪かったとしても、その後に適当な副作用軽減の方法が見つかり、これにより実際の現場でのコンプライアンスがあがることが考えられる。あるいは、効果が証明されていない治療法を受ける臨床試験の場合と有効性が示された治療法を受ける場合とでは患者のコンプライアンスは大きくことなるかもしれない。つまり、臨床試験で観察されたコンプライアンス割合が実際の現場で観察されるそれとことなる場合には、ITT 解析による治療効果の推定の意義は減じることになる。

ここでは、counterfactual 因果モデル^{2,3)}を用いることによって割り付け通りに治療が守られていた場合に観察されたはずの真の治療効果推定について考える。提案する推定方法は、各対象者に対する潜在結果変数 (potential outcome) に対して単純な構造モデル (structural model) を仮定するものの、治療群間の比較可能性に関してランダム化以外の仮定は何も必要としない。また、ITT 解析と同様に治療効果に関する因果帰無仮説の妥当な検定を与える一方で、対象者が実際に受けた治療に関する情報 (non-compliance information) を利用でき、ランダムでない non-compliance の場合であっても、因果治療効果の推定値を与えるという特徴をもっている。なお、Mark and Robins⁴⁾は生存時間がエンドポイントである場合のランダム化にもとづく因果効果の推定、佐藤⁵⁾は治療が 1 回のみの場合の因果リスク差、リスク比の推定について扱っている。

2. M-CSF 試験

M-CSF 試験は、178 人の成人急性骨髄性白血病 (AML) 患者に対する寛解後療法（地固め療法）において、感染症にともなう発熱に対するマクロファージコロニー刺激因子（以下 M-CSF）の効果を調べたプラセボ対照のランダム化臨床試験である⁶⁾。対象者は M-CSF グループ（87 人）とプラセボグループ（91 人）の 2 群にランダムに割り付けられ、3 コースからなる地固め療法を受けた。各コースにおける地固め療法終了後に対象者は M-CSF またはプラセボをそれぞれのコースで投与された。ここでは、各コース中に感染症にともなう 38 度以上の発熱を起こすかどうか（二値データ）をエンドポイントと考える。ここでの関心事は M-CSF が繰り返し生じる感染症にともなう発熱のリスクを減少させるかどうかということである。

各コースでの発熱の頻度を表 1 に示す。M-CSF グループでは、1 コース終了後に 5 人、2 コース終了後に 7 人、プラセボグループでは、それぞれ 2、8 人が、おもに地固め療法の拒否、急性白血病による死亡のため打ち切りを受けている。

3 コース目において、M-CSF グループの 2 人、プラセボグループの 2 人は実際には割り付けられた治療法を受けていない。ここでは、プラセボ効果が無視できるとして、プラセボグループの 2 人は compliant ケースとみなし、M-CSF グループの 2 人の対象者を 3 コース目で

non-compliant ケースと考える。

表1. M-CSF試験の結果の要約

時点	治療群	event		No event	合計
1コース	M-CSF	56	(64.4%)	31	87
	Placebo	69	(75.8%)	22	91
RD = -0.115 ^a RR = 0.849 ^a P = 0.07 ^b					
2コース	M-CSF	49	(59.8%)	33	82
	Placebo	62	(69.7%)	27	89
RD = -0.099 RR = 0.858 P = 0.12					
3コース	M-CSF	21	(28.0%)	54 ^c	75
	Placebo	35	(43.2%)	46 ^d	81
RD = -0.152 RR = 0.648 P = 0.04					

^a RD：リスク差，RR：リスク比 ^b Fisher's exact test

^c 実際にはM-CSFを受けていない患者が2人

^d 実際にはプラセボを受けていない患者が2人

3. 因果治療効果の推定

対象者 i ($i=1, \dots, N$) に対して時点 t ($t=1, \dots, T$) において観察される二値結果変数（発熱の有無）を Y_{it} とし、各対象者は時点 $t-1$ で試験治療（M-CSF）かコントロール治療（プラセボ）のどちらかの治療法を受けるとする。ただし、時点 0 はランダム化の時点である。ここで、何人かの対象者は最初に割り付けられた治療法に従わず、他方の治療を受ける状況（治療法のクロスオーバー）を想定する。M-CSF 試験では、3 コース目における M-CSF グループの 2 人がそのような対象者である。3.3.節まで欠測データは存在しないと仮定する。

われわれが知りたい平均的な因果治療効果（average causal effect of continuous treatment）は、全ての対象者が試験治療を最初から受け続けていた場合の平均的な結果と全ての対象者がコントロール治療を最初から受け続けていた場合の平均的な結果の違いである⁷⁾。

ここで、対象者 i ($i=1, \dots, N$) に対する時点 t における 2 つの潜在二値結果変数（1：発熱あり、0：発熱なし）を次のように定義する。

Y_{it}^M ：対象者 i が時点 $t-1$ において試験治療を受けた場合の時点 t での結果

Y_{it}^P ：対象者 i が時点 $t-1$ においてコントロール治療を受けた場合の時点 t での結果

ただし、上記の潜在変数は次の 2 つの前提を満たすとする。

1. 対象者 i に対する潜在結果変数は別の対象者 j ($j \neq i$) が受けた治療に影響を受けない
2. 対象者 i に対する潜在結果変数は割り付けられた治療の関数ではなく、対象者 i が時点 $t-1$ において実際に試験治療（コントロール治療）を受けた場合には、潜在結果のうちの一方が実現したと考え、 $Y_{it} = Y_{it}^M$ (Y_{it}^P) とする

Y_{it}^M と Y_{it}^P の比較は、同一対象者に対することとなる治療状況を問題としているので、比較可能性があり、その対比により真の治療効果を求めることができる。例えば、時点 t における平均因

果リスク差は次のように表すことができる。

$$E(Y_{it}^M) - E(Y_{it}^P) \quad (1)$$

(1) 式を時点に関して平均を取れば、全ての対象者が試験治療を最初から受け続けていた場合の平均的な結果と全ての対象者がコントロール治療を最初から受け続けていた場合の平均的な結果の違いが得られる。

ここで、 Y_{it}^M と Y_{it}^P は同一対象者がことなる治療を受けた場合の結果（counterfactual モデル）なので、何らかの仮定を設けない限り上述の平均的な因果治療効果は識別可能ではない。

3.1. 共通因果リスク差の推定

次のような単純な加法構造モデル（additive structural models）を想定する。

$$E(Y_{it}^M) - E(Y_{it}^P) = \delta_0 \quad (2)$$

このモデルのもとでは、 δ_0 は観察期間を通して共通の因果リスク差と解釈することができる。

δ_0 を推定するために、 $U_i(\delta) = \sum_t (Y_{it} - \delta Z_{it-1})$ という関数を考えてみる。ただし、 Z_{it-1} は対象者 i が時点 $t-1$ において実際に受けた治療法を表す変数で、試験治療を受けていれば $Z_{it-1} = 1$ 、コントロール治療を受けていれば $Z_{it-1} = 0$ をとるとする。前述の潜在変数に対する 2 つの前提とモデル (2) のもとでは、 $E(Y_{it} - \delta_0 Z_{it-1}) = E(Y_{it}^P)$ なので、 $U_i(\delta_0)$ の期待値は対象者 i がコントロール治療を受け続けた場合に観察されるイベント発生回数の期待値である。

共通因果リスク差の真値 δ_0 を与えたもとでは、 $U_i(\delta_0)$ は対象者全員がコントロール治療を受け続けた場合の結果なので、性別や年齢のような背景情報と同様に $U_i(\delta_0)$ は治療法の割り付けとは統計的に独立である。すなわち、 $\delta = \delta_0$ の場合、

$$E[U_i(\delta) | R_i = 1] = E[U_i(\delta) | R_i = 0] \quad (3)$$

が成立する。ただし、 R_i は割り付け治療グループを表す変数で、試験治療の場合 $R_i = 1$ 、コントロール治療の場合 $R_i = 0$ をとるとする。

(3) 式を解くことによって、 δ_0 の推定量 $\hat{\delta}$ は次式のようにもとまる。

$$\hat{\delta} = \frac{\frac{\sum_j jX_{Mj}}{n} - \frac{\sum_j jX_{Pj}}{m}}{\frac{\sum_j jn_j}{n} - \frac{\sum_j jm_j}{m}} \quad (4)$$

ただし、(4) 式の記法に関しては、以下の表 2、表 3 を参照。

表2. 観察データの要約： $2 \times (J+1)$ 表

割り付け治療群	イベントの回数						合計
	0	1		j		J	
M-CSF	X_{M0}	X_{M1}	$\cdots$	X_{Mj}	$\cdots$	X_{MJ}	n
プラセボ	X_{P0}	X_{P1}	$\cdots$	X_{Pj}	$\cdots$	X_{PJ}	m
合計	T_0	T_1		T_j		T_J	N

表3. Non-complianceの要約

割り付け治療群	M-CSFを受けた回数						合計
	0	1		j		J	
M-CSF	n_0	n_1	$\cdots$	n_j	$\cdots$	n_J	n
プラセボ	m_0	m_1	$\cdots$	m_j	$\cdots$	m_J	m
合計							N

注意：コンプライアンスが100%の場合、

$$n_J = n, n_{J-1} = \cdots = n_0 = 0, m_J = \cdots = m_1 = 0, m_0 = m$$

治療が 1 回の場合($T = J = 1$)には、推定量 (4) は、操作変数推定量 (instrumental variables estimator)⁸⁾ に一致する。

(3) 式は、次のように表現できるので、

$$\frac{nm}{N} \left\{ \sum_j j \left(\frac{X_{Mj} - \delta n_j}{n} \right) - \sum_j j \left(\frac{X_{Pj} - \delta m_j}{m} \right) \right\} = \sum_j j \left\{ \left(X_{Mj} - \frac{m}{N} \delta n_j \right) - (T_j - \delta m_j) \frac{n}{N} \right\} \quad (5)$$

各グループで観察されるコンプライアンス割合を与えたもとでは、(5) 式の分散は Mantel 傾向検定の分散の公式から求めることができる。したがって、 δ_0 に対する $1-\alpha$ 信頼区間は、次式を満たす δ の組である。

$$|\chi_M(\delta)| \leq z_{1-\alpha/2}$$

ただし、 $\chi_M(\delta)$ は $U_i(\delta)$ に対する Mantel 傾向検定統計量であり、 $z_{1-\alpha/2}$ は標準正規分布の $1-\alpha/2$ パーセント点である。

因果帰無仮説 $\delta_0 = 0$ のもとでは、上述の検定統計量は表 2 に対する ITT Mantel 傾向検定統計量に一致する（次式参照）。したがって、帰無仮説のもとでは提案する方法はノンパラメトリックである。

$$\chi_M(\delta=0) = \frac{\sum_j j \left(X_{Mj} - T_j \frac{n}{N} \right)}{\sqrt{\frac{nm}{N^2(N-1)} \{ N \sum_j j^2 T_j - (\sum_j j T_j)^2 \}}}$$

3.2. 共通因果リスク比の推定

前節の因果リスク差と同様の考え方が因果リスク比の推定にも応用できる。

次のような単純な乗法構造モデル (multiplicative structural models) を想定する。

$$\log[E(Y_{it}^M)] - \log[E(Y_{it}^P)] = \phi_0 \quad (6)$$

ϕ_0 を推定するために、 $V_i(\phi) = \sum_t Y_{it} \exp(-\phi Z_{it-1})$ という関数を考えると、モデル(6)のもとでは、 $\phi = \phi_0$ のとき、

$$E[V_i(\phi) | R_i = 1] = E[V_i(\phi) | R_i = 0]$$

が成立する。この式から、共通因果リスク比の推定量は次式のようにもとまる。

$$\exp(\hat{\phi}) = \frac{\frac{\sum_k kX_{Tk|T}}{n} - \frac{\sum_k kX_{Ck|T}}{m}}{\frac{\sum_k kX_{Ck|C}}{m} - \frac{\sum_k kX_{Tk|C}}{n}}$$

ただし、 $X_{Ak|B}$ は、治療法 A に割り付けられたが、実際には治療法 B を受けて k 回イベントを起こした対象者の数を表す。

3.3. 欠測値

ITT 解析の原則に従えば、ランダム化の対象となった全ての対象者を最後まで追跡し結果変数を測定することが要求されるので、ランダム化研究では結果変数に関する欠測は本来大きな問題ではないかもしれない。しかしながら、M-CSF 試験のように途中で打ち切りを受ける対象者が実際には存在する。例えば、M-CSF 試験では試験途中での化学療法の拒否、急性白血病による死亡のため、その後の感染症の発症に関するデータが存在しない対象者が存在する。

このような脱落は治療効果の推定値にバイアスを与える一つの原因となる。特に、エンドポイントである感染症発症に関する（未測定の）リスク要因によって脱落が生じている場合が問題である。ここでは、そのようなリスク要因が存在しないと考え、例えば、共通因果リスク差の推定に関しては、3 節の $U_i(\delta)$ を各対象者の観察回数 T_i で調整した $U_i(\delta)/T_i$ に置き換えて同様の推定を行うことにする。

4. 結果

表 1 のデータに対する Generalized Estimating Equations (GEE) による ITT 解析の結果を表 4 に示す。結果変数間の相関構造には compound symmetry 構造を仮定した結果である。治療法の回帰係数がゼロかどうかのロバスト p 値は、リスク差モデル、リスク比モデルいずれも 1%程度で

あり、また共通リスク差の点推定値は-0.122（ロバスト 95%信頼区間は-0.211 から-0.032） 共通リスク比の点推定値は 0.834（ロバスト 95%信頼区間は 0.717 から 0.969）である。

表4. GEEによるITT解析

リンク関数	パラメータ	点推定値	ロバスト 95% CI	ロバスト P値
Identity	1コース	- ^a	-	-
	2コース	-0.055	-0.149, 0.038	0.247
	3コース	-0.342	-0.439, -0.245	0.000
	治療効果	-0.122	-0.211, -0.032	0.008
Log	1コース	- ^a	-	-
	2コース	-0.082	-0.220, 0.057	0.249
	3コース	-0.655	-0.876, -0.435	0.000
	治療効果	-0.182 ^b	-0.333, -0.032 ^b	0.018

^a 1コース目がレファレンスカテゴリー

^b 共通リスク比の点推定値 = 0.834 (95% CI : 0.717 - 0.969)

3 節で紹介した方法で求めた共通因果効果の推定値を表 5 に示す。

表5. 因果治療効果の推定

因果パラメータ	点推定値	95% CI
リスク差	-0.121	-0.215, -0.027
リスク比	0.814	0.688, 0.956

5. 考察

治療を繰り返し実施した場合の二値経時観察データに対するランダム化にもとづく因果リスク差、およびリスク比の推定について考えた。何らかの仮定を設けない限り、全ての対象者が割り付け通りの治療を受けたときの真の治療効果を推定することはできない。本論文で提案した方法は、治療群間の比較可能性に関してランダム化以外の仮定は何も必要としない。特に、「治療法の不遵守がランダムに起きている」と仮定していないことに注意を要する。したがって、提案した方法は、測定された既知のリスク要因、あるいは未測定のリスク要因によってコンプライアンスが規定されていたとしても（ランダムでない non-compliance が生じていたとしても）因果効果の推定値を与えてくれる。しかしながら、推定したい因果治療効果を識別可能にするために、(2), (6)式のような単純な構造モデルを仮定した。このモデルは観測不能な量にもとづいて定義されているので、データからその正しさを検証することはできない。この問題に関しては感度解析が必要であると思われる⁹⁾。

提案した推定方法は、未測定のリスク要因によって脱落が生じていないこと（ランダムな欠測）を仮定している。脱落に影響を与えるリスク要因が既知の場合には、通常の層別解析の技法、あるいは IPCW (Inverse Probability of Censoring Weighted) 法¹⁰⁾によりその要因で調整した解析が必要である。本論文で提案した方法をそのような場合に拡張することは容易である。しかしながら、脱落に影響を与える未測定のリスク要因が存在するかどうかはデータから決めることはできないので、ランダムでない脱落が生じていることを想定した場合の感度解析を行うことが重要である。

最後に、M-CSF 試験では non-compliant ケースが 3 コース目の M-CSF グループに 2 人であったために、ITT 治療効果と因果治療効果の推定値の違いは見られなかったが、non-compliance(特に、ランダムでない non-compliance)割合の増加とともにその違いは大きくなると思われる^{4,11)}。

参考文献

- 1) Lagakos SW, Lim LLY, Robins JM. Adjusting for early treatment termination in comparative clinical trials. *Statistics in Medicine* 1990; 9:1417-1424.
- 2) Holland PW. Statistics and causal inference (with Discussions). *Journal of the American Statistical Association* 1986; **81**:945-970.
- 3) Greenland S and Robins JM. Identifiability, exchangeability, and epidemiologic confounding. *International Journal of epidemiology* 1986; **15**:413-419.
- 4) Mark SD and Robins JM. A method for the analysis of randomized trials with compliance information: An application to the multiple risk factor intervention trial. *Controlled Clinical Trials* 1993; **14**:79-97.
- 5) 佐藤俊哉 ランダム化にもとづいた intent-to-treat 解析. *応用統計学* 1994; **23**:21-34.
- 6) Ohno R *et al.* Human urinary macrophage colony-stimulating factors reduces the incidence and duration of febrile neutropenia and shortens the period required to finish three courses of intensive consolidation therapy in acute myeloid leukemia: a double-blind controlled study. *Journal of Clinical Oncology* 1997; **15**:2954-2965.
- 7) Robins JM, Greenland S, Hu FC. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association* 1999; **94**:687-700.
- 8) Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables (with Discussions). *Journal of the American Statistical Association* 1996; **91**:444-472.
- 9) Balke A and Pearl J. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* 1997; **92**:1171-1176.
- 10) Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* 1998; **93**:1321-1339.
- 11) Korhonen PA, Laird NM, Palmgren J. Correcting for non-compliance in randomized trials: An application to the ATBC study. *Statistics in Medicine* 1999; **18**: 2879-2897.