

論文題目 Principal Stratification を用いた代替エンドポイントの評価
—進行性前立腺癌試験データへの適用—
指導教員 松山裕浩教授
所属 東京大学大学院医学系研究科健康科学看護学専攻(平成17年4月進学)
氏名 田中司朗

1 背景と目的

癌や心疾患など致死的疾患の検証的試験では、真のエンドポイントの測定はコスト・時間の観点から難しく、測定の簡便性と臨床的適切性はしばしば矛盾する。代替エンドポイントは、この矛盾を解決するための妥協として広く用いられているが、その利用には功罪がある。患者にとっては、有望な医薬品の早期入手、製薬会社にとっては経済的利益などが良い側面である。一方、代替エンドポイント利用の代償は、医薬品の評価に伴う不確実性である。代替エンドポイントの利用が不適切な医薬品評価に繋がった事例が多々経験され、代替エンドポイントの妥当性を評価する必要が認められるようになってきた。

一方、統計的方法論の観点からは、Prentice 基準や、Freedman の proportion of treatment effect Explained, メタアナリシスアプローチなどの提案があるが、決定版といえるものはない。本論文では、このような現状を鑑み、因果推論に基づく代替性評価へのアプローチを行う。Frangakis と Rubin(2002) は、principal stratification という因果推論の新しい枠組みを提案し、代替エンドポイントの評価への応用を指摘した。しかし、principal stratification の実装のためには、パラメータ推定などの解決すべき課題がいくつかあり、応用例は見当たらない。そこで本論文では、principal stratification に基づく代替エンドポイント評価のための統計手法を提案する。また、提案する手法を進行性前立腺癌試験データに適用し、PSA の代替性評価を行う。

2 Principal Stratification を用いた代替エンドポイントの評価

n 人の患者が、試験治療群と対照群にランダムに割りつけられる臨床試験を考える。 R_i を対象者 $i = 1, \dots, n$ の治療変数とする(試験治療群なら 1, 対照群なら 0)。観察される(打ち切りを伴う)生存時間を T_i^{obs} とする。代替エンドポイントの候補となる変数は、反応有りならば 1, 無ければ 0 となる二値変数 S_i^{obs} とする。また、Stable Unit Treatment Value Assumption, 一致性の仮定, ランダム化の仮定の下で, Neyman-Pearson

の潜在結果変数を定義する。すなわち、 $T_i^{(r)}$ を対象者 i が $R_i = r$ で指示される治療薬を与えられたときの生存時間、 $S_i^{(r)}$ を対象者 i が $R_i = r$ で指示される治療薬を与えられたときの代替エンドポイントの候補となる変数とする ($r \in \{0, 1\}$)。

Frangakis と Rubin は、ベクトル $S_i = (S_i^{(0)}, S_i^{(1)})$ に基づく層別を principal stratification と呼んだ⁴⁰。試験治療と対照治療の間に、代替エンドポイントの候補となる変数への効果に順序性がある状況とする。この下で、 S_i に、 $S_i^{(0)} \leq S_i^{(1)}$ という単調性の仮定を置く。この仮定の下で、全対象者は以下の三つの basic principal strata に分類される。

- 1) Always Responder (AR) = $\{i | S_i = (1, 1)\}$
- 2) Partial Responder (PR) = $\{i | S_i = (0, 1)\}$
- 3) Never Responder (NR) = $\{i | S_i = (0, 0)\}$

本論文では、理想の代替エンドポイントの定義を $AR \cup NR$ における $T_i^{(0)}$ と $T_i^{(1)}$ の比較に差が無いような変数とし、これを principal surrogate と呼ぶ。

また、 $S_i^{(0)} \neq S_i^{(1)}$ となるサブグループにおける $T_i^{(1)}$ の $T_i^{(0)}$ に対するハザード比を Associative Effect (AE)、 $S_i^{(0)} = S_i^{(1)}$ となるサブグループにおける $T_i^{(1)}$ の $T_i^{(0)}$ に対するハザード比を Dissociative Effect (DE) と定義する。principal stratum の定義に用いた潜在結果変数はランダム化と独立と考えられるため、これらの群間比較のパラメータはランダム化に基づく因果効果として解釈出来る。上の三つの basic principal strata を考えた場合、DE は、治療により代替エンドポイントに変化がないサブグループ ($AR \cup NR$) 内での群間比較のパラメータである。この比較が等しくない場合、代替エンドポイントに関係しない作用機序による治療効果があると考えられる。一方、AE は、治療によって代替エンドポイントが改善したサブグループ (PR 群) 内での群間比較に相当し、群間差が大きい方が代替エンドポイントとして望ましい。Proportion Extrapolable (PE) は、真のエンドポイントへの治療効果のうち、外挿可能な部分の割合を表す指標である。PE は、AR 群の割合 P_{AR} と τ 年生存関数 $S(\tau)$ に基づいて、以下のように定義される。

$$PE = \frac{P_{AR} [S(\tau | S_i = (0, 1), R_i = 1) - S(\tau | S_i = (0, 1), R_i = 0)]}{S(\tau | R_i = 1) - S(\tau | R_i = 0)}$$

3 進行性前立腺癌試験データ

本論文の適用事例は、非ステロイド性抗アンドロゲン剤と黄体形成ホルモン放出ホルモンアゴニスト (LHRHA) 併用による効果及び安全性を検討する目的で行われた進行性前立腺癌試験のデータである。試験

デザインは、多施設共同ランダム化第III相試験であった。主な適格基準は、病理組織学的な前立腺癌の確認、ステージD、未治療、測定可能または評価可能な病変の存在、20歳以上85歳以下、performance statusが0から3などであった。

提案する手法の適用では、LHRHA単独療法群（以下LHRHA群）、LHRHAと非ステロイド性抗アンドロゲン剤の併用群（以下CAB群）の二群比較の検討を行うこととする。また、登録された前立腺癌患者167人の内、Full Analysis Setの基準を満たし、かつ治療後にPSAの測定があった156人を解析対象とする。真のエンドポイントは、原病死までの生存時間と再燃までの生存時間の二つを検討する。代替エンドポイントは、治療後のPSAを二値分類したものとし、これを治療後PSA反応と呼ぶことにする。分類の定義は、常に1がPSAが低下する方向で0がその逆とし、以下の因通りについて検討する。

PSA正常化: PSA Nadirが4ng/ml以下であること

PSA正常持続: 治療後のPSAプロファイルに2回連続4ng/ml以下があること

PSA90%減少: 治療後のPSAプロファイルに2回連続ベースラインの10%以下があること

PSA急減: PSA Velocityが、一日当たり <0.05 以下であること

ただし、PSA Nadirは、治療後PSA値の最低値とする。PSA Velocityは、治療後90日以内に対数PSAが低下する勾配を単回帰により推定したものとす。

表1に、進行性前立腺癌試験の患者背景を、表2に、治療後PSAプロファイルの要約を示す。主な試験結果は、CAB療法はLHRHA単剤に比べ、原病死までの時間と再燃までの時間の両方を有意に延長したというものであった（ハザード比はそれぞれ0.56, 0.52）。表3, 4に、提案した手法の適用結果を示す。PSA正常化について原病死への代替性を示唆する結果が得られた（AE=0.28, DE=0.86）。PSAの測定誤差の影響が小さいと思われるPSA正常持続についても近い結果であった（AE=0.34, DE=0.73）。PSA90%減少, PSA急減の代替性は見られなかった。特にPSA急減の結果はAEがDEより大きく不自然であった。再燃に関する結果では、やはりPSA正常化とPSA正常持続について代替性が示唆される結果であった。

4 結論

principal stratificationに基づく代替エンドポイント評価のための統計手法を提案し、進行性前立腺癌試験データへの適用を行った。その結果、治療後PSAが原病死の代替エンドポイントとなり得ることが示唆され、特にPSA正常化に基づく分類（最低値が4ng/ml以下かどうか）が有用であることが明らかになった。

表1. 進行性前立腺癌試験データの患者背景

| | CAB群 (n=104) | LHRHA群 (n=52) |
|-------------|--------------|---------------|
| 年齢 (歳) * | 713±7.5 | 713±7.2 |
| 体重 (kg) † | 58.5±8.3 | 59.8±11.8 |
| LHRHAゾニスト † | | |
| ゴセリン | 27 (26%) | 16 (31%) |
| リュープロレリン | 77 (74%) | 36 (69%) |
| PS † | | |
| 0 | 78 (75%) | 41 (79%) |
| 1-4 | 26 (25%) | 11 (21%) |
| 臨床病期 † | | |
| D1 | 9 (9%) | 5 (10%) |
| D2 | 95 (91%) | 47 (90%) |
| 組織学的分類 † | | |
| 低分化腺癌 | 10 (10%) | 3 (6%) |
| 中分化腺癌 | 57 (55%) | 21 (40%) |
| 高分化腺癌 | 37 (36%) | 28 (54%) |
| EOD † ‡ | | |
| グレード 0 | 18 (17%) | 11 (21%) |
| グレード 1 | 45 (43%) | 21 (40%) |
| グレード 2-4 | 40 (38%) | 20 (38%) |

PS: Performance Status, EOD: Extent of Disease

* 平均±SD, † 頻度(%), ‡ 一人が欠測

表2. 治療後PSAプロファイルの要約

| | CAB群 (n=104) | LHRHA群 (n=52) |
|------------------|---------------|---------------|
| 治療後PSA値の時点ごとの分布* | | |
| 第0週 (0.0%, 74%) | 179.07±361.80 | 191.22±544.40 |
| 第4週 (9%, 84%) | 11.80±24.10 | 63.80±239.75 |
| 第8週 (9%, 82%) | 2.18±4.80 | 16.78±70.30 |
| 第12週 (8%, 82%) | 1.12±3.10 | 11.24±39.50 |
| 第16週 (7%, 79%) | 0.88±2.05 | 10.12±37.90 |
| 第20週 (6%, 76%) | 0.78±2.00 | 8.18±24.10 |
| 第24週 (6%, 73%) | 0.69±1.25 | 5.66±16.50 |
| 第28週 (5%, 72%) | 0.60±1.00 | 4.35±11.55 |
| 治療後PSA反応率 † | | |
| PSA正常化 | 84 (81%) | 18 (35%) |
| PSA正常持続 | 78 (75%) | 18 (35%) |
| PSA90%減少 | 92 (88%) | 35 (67%) |
| PSA急減 | 77 (74%) | 9 (17%) |

* 平均は四分位範囲, 平均は対象値の最小二乗平均を
変換し指定, ()内は, 測定がなされた人数/全体と
予定の前週3日以内だった人数/測定がなされた人数
† 頻度(%)

表3. PSA代替性評価の結果 (原病死)

| | 点推定値 | 95%信頼区間 |
|---------------|------|----------------|
| PSA正常化 | | |
| AE (46%)* | 0.28 | (0.13 to 0.61) |
| DE (19%+35%)* | 0.86 | (0.46 to 1.60) |
| PE | 0.89 | (0.53 to 1.48) |
| PSA正常持続 | | |
| AE (40%)* | 0.34 | (0.16 to 0.72) |
| DE (25%+35%)* | 0.73 | (0.39 to 1.36) |
| PE | 0.72 | (0.37 to 1.41) |
| PSA90%減少 | | |
| AE (21%)* | 0.51 | (0.22 to 1.16) |
| DE (12%+67%)* | 0.50 | (0.27 to 0.93) |
| PE | 0.24 | (0.04 to 1.31) |
| PSA急減 | | |
| AE (57%)* | 1.44 | (0.70 to 2.96) |
| DE (26%+17%)* | 0.19 | (0.10 to 0.36) |
| PE | <0 | |

AE: Associative Effect, DE: Dissociative Effect

PE: Proportion of Extrapolable

*ハザード比 (95%信頼区間), CABが高いと<1

()内はPRの割合又はNR, ARの割合

表4. PSA代替性評価の結果 (再燃)

| | ハザード比 | 95%信頼区間 |
|---------------|-------|----------------|
| PSA正常化 | | |
| AE (46%)* | 0.36 | (0.23 to 0.55) |
| DE (19%+35%)* | 0.65 | (0.37 to 1.15) |
| PE | 0.61 | (0.15 to 2.40) |
| PSA正常持続 | | |
| AE (40%)* | 0.39 | (0.25 to 0.61) |
| DE (25%+35%)* | 0.62 | (0.36 to 1.08) |
| PE | 0.55 | (0.13 to 2.28) |
| PSA90%減少 | | |
| AE (21%)* | 0.22 | (0.13 to 0.37) |
| DE (12%+67%)* | 0.54 | (0.35 to 0.85) |
| PE | 0.20 | (0.02 to 1.90) |
| PSA急減 | | |
| AE (57%)* | 0.95 | (0.63 to 1.43) |
| DE (26%+17%)* | 0.12 | (0.07 to 0.22) |
| PE | <0 | |

AE: Associative Effect, DE: Dissociative Effect

PE: Proportion of Extrapolable

*ハザード比 (95%信頼区間), CABが高いと<1

()内はPRの割合又はNR, ARの割合