

一般口演

## 一般口演28

## 機械学習・アルゴリズム・解析モデル

2017年11月23日(木) 12:45 ~ 14:15 E会場 (10F 会議室1003)

**[4-E-2-OP28-5] モンテカルロシミュレーションによる追跡率と paired-t 検定結果の関連性の解析**

山下 暁士<sup>1</sup>, 酒井 忠博<sup>2,3</sup>, 船田 千秋<sup>1</sup>, 小林 大介<sup>1</sup>, 大山 慎太郎<sup>1</sup>, 山下 佳子<sup>1</sup>, 佐藤 淳一<sup>1</sup>, 朝田 委津子<sup>1</sup>, 平岩 秀樹<sup>3</sup>, 濱田 恭<sup>3</sup>, 大野 洋平<sup>3</sup>, 宮本 健太郎<sup>3</sup>, 土谷 早穂<sup>3</sup>, 白鳥 義宗<sup>1</sup> (1.名古屋大学医学部附属病院 メディカルITセンター, 2.トヨタ記念病院 整形外科, 3.名古屋大学大学院医学系研究科 整形外科学)

**【目的】**手術の成績を術前後のアウトカムの差で比較する研究がよく見受けられる。この研究手法に対し、追跡率がどの程度検定の結果をゆがませるのかについて、コンピュータシミュレーションを用い検討すること。

**【方法】**Excel VBAを用いて以下のシミュレーションを行った。

1. 症例は100例。術前と術後の調子を表す測定不能な真のスコア(仮に真スコアとする)があると仮定。真スコアは正規分布に従ったランダムな値をとるとした。
2. 真のスコアと相関する測定可能なスコア(仮に測定スコア)があり、それを術前と術後に測定したと仮定した。今回は両スコア間の相関係数を0.6と仮定して計算した。
3. 手術は全く真スコアに影響を与えないとした。つまり、術前後で真スコアの従う分布は同じであると仮定した。
4. 術前後の真スコア変化量の悪い順に追跡できなくなると仮定した。最終的な追跡率を100%、90%、80%、70%、60%とした場合それぞれにおいて、追跡できた症例に対して、術前後の測定スコアを用いて paired-t testを行い、有意水準を5%として統計学的に有意かどうかを検定した。
5. 上記を1万回行い、各最終追跡率において、術前後のスコアの差が統計学的に有意となった確率を算出した。
6. 上記を3回繰り返し、有意になった確率の範囲を求めた。

**【結果】**最終的な追跡率が100%、90%、80%、70%、60%であった場合のそれぞれにおいて、術前後のスコアの差が有意となった確率は4.9-5.2%、18.3-19.1%、50.2-51.0%、74.8-75.5%、88.8-89.3%であった。

**【結論】**スコアの術前後の変化が追跡の可否に影響を与える場合、追跡不能例があるだけで、意味のない手術の術前後のアウトカムの差が有意になる確率が、危険率に比べて何倍も高くなってしまいが確認された。この結果は、術前後のアウトカムを用いた研究において、追跡率がかなり高くなければ結果はゆがんでいる可能性が否定できないということを示唆するものである。

# モンテカルロシミュレーションによる追跡率と paired-t 検定結果の関連性の解析

山下 暁士<sup>\*1</sup>、酒井 忠博<sup>\*2,3</sup>、船田 千秋<sup>\*1</sup>、小林 大介<sup>\*1</sup>、大山 慎太郎<sup>\*1</sup>、山下 佳子<sup>\*1</sup>、佐藤 淳一<sup>\*1</sup>、朝田 委津子<sup>\*1</sup>、平岩 秀樹<sup>\*3</sup>、濱田 恭<sup>\*3</sup>、大野 洋平<sup>\*3</sup>、宮本 健太郎<sup>\*3</sup>、土谷 早穂<sup>\*3</sup>、白鳥 義宗<sup>\*1</sup>

\*1 名古屋大学医学部付属病院 メディカル IT センター、\*2 トヨタ記念病院 整形外科、  
\*3 名古屋大学大学院医学系研究科 整形外科学

## Analysis of the association between follow-up rates and paired-t-test results using the Monte Carlo simulation

Satoshi Yamashita<sup>\*1</sup>, Tadahiro Sakai<sup>\*2,3</sup>, Chiaki Funada<sup>\*1</sup>, Daisuke Kobayashi<sup>\*1</sup>, Shintaro Oyama<sup>\*1</sup>, Keiko Yamashita<sup>\*1</sup>, Jun-ichi Sato<sup>\*1</sup>, Itsuko Asada<sup>\*1</sup>, Hideki Hiraiwa<sup>\*3</sup>, Takashi Hamada<sup>\*3</sup>, Yohei Ono<sup>\*3</sup>, Kentaro Miyamoto<sup>\*3</sup>, Saho Tsuchiya<sup>\*3</sup>, Yoshimune Shiratori<sup>\*1</sup>

\*1 Medical IT center, Nagoya University Hospital, \*2 TOYOTA Memorial Hospital,  
\*3 Department of Orthopaedics, Nagoya University Graduate School of Medicine

**Objective:** To analyze how much Follow-up rate distorts the result of paired-t test by Monte-Carlo simulation.

**Methods:** The following simulation was performed using Excel VBA.

1. 100 cases. It is assumed that there is an unmeasurable true score (T-score) representing the disease condition before and after surgery. The T-score is assumed to take a random value according to the normal distribution.
2. There was a measurable score (M-score) correlated with T-score, and it was assumed that it was measured preoperatively and postoperatively. Correlation coefficient between both scores was assumed to be 0.6.
3. Surgery didn't affect T-score at all.
4. Assumed that it cannot be followed-up in descending order of the change in T-score before and after surgery. In each case where the final follow-up rate was set to 100%, 90%, 80%, 70%, and 60%, paired-t test was performed using M-score, Test was considered significant when  $p < 0.05$ .
5. The above procedure was performed 10,000 times, and the probability of statistically significant difference between before and after the operation was calculated at each follow-up rate.
6. Repeat the above 3 times.

**Results:** In each of cases where the final follow-up rate was 100%, 90%, 80%, 70%, 60%, the probability that the difference in the M-score before and after the surgery became significant was 4.9-5.2%, 18.3 -19.1%, 50.2 - 51.0%, 74.8 - 75.5%, 88.8 - 89.3%.

**Conclusion:** This result suggests that in a study using differences in outcomes before and after surgery the possibility of distorting the results can not be denied unless the follow-up rate is high enough.

**Keywords:** Follow-up rate, Biases, Computer simulation

### 1. 緒論

今日、多くの外科医が自分や自施設の手術の成績を集計・分析して発表している。過去には case-series による報告が主であったが、現在では症例対照研究やコホート研究、randomized controlled trial (RCT) などが増加傾向にある。研究手法自体は徐々に洗練されたものになっているにもかかわらず、バイアスを引き起こす要因の中にはまだ外科医の意識がほとんど変わらないものもある。その 1 つが追跡率 (follow-up rate) である。

私の専門分野の反復性肩関節脱臼に対する標準治療である関節鏡下バンカート修復術 (Arthroscopic Bankart Repair) の効果に関する論文を PubMed で確認したところ、2017 年 4 月 1 日～6 月 30 日に出版された中で詳細を確認できた 11 論文中、追跡率の記載があったものは 3 本のみであった。また、4 本は inclusion criteria の中に「追跡が〇〇年以上できた症例」という項目が含まれており、手術を行った全例の把握も初めより行われていない状態であった。

追跡率が低いことで起こるバイアスを追跡バイアス (follow-up bias) と呼び、選択バイアス (selection bias) の 1 つとされている<sup>1)</sup>。どの程度、追跡率が下がると考慮しなければならぬくらいにバイアスが大きくなるかについては、今のところはっきりとした見解はない<sup>2)3)4)</sup>。例えば 50% を下回ってもアウトカムに明らかな変化が認められなかったという報告がある一方<sup>2)</sup>で、脱落にアウトカムに関係する場合にはわずかな低下でも影響があると報告もある<sup>1)</sup>。臨床をしていると、受けた手術の成績が不良なために転医してきた患者に遭遇することはよくあることであり、脱落にアウトカムが関係するのは日常的なケースと考えられる。

手術の効果を検討する際に、対照群との比較ではなく手術前後でのアウトカムの比較を行い報告することがよくある。この場合、術後のアウトカムが脱落に関係しているケースでは研究の結果に対する脱落の影響が大きくなるのが予想される。しかし、こういったケースでどの程度の脱落すると結果に影響を与えるかについて検討した報告は渉猟し得る範囲ではほとんどない。

## 2. 目的

本研究の目的は、ある治療(手術を想定しているが、それ以外でも適用は可能と考える)の効果を研究する目的で術前後のアウトカムの比較を paired-t 検定で行った場合に、追跡率が下がることがどの程度検定の結果をゆがませるのかについて、コンピュータシミュレーションを用い検討することである。

## 3. 方法

本研究ではシミュレーションのプラットフォームとして Microsoft Excel 2016、および Microsoft Visual Basic for Application 7.1 を用いた。

### 3.1 シミュレーションモデル

シミュレーションモデルとして、以下のような研究を行ったと仮定した。

**方法:** prospective な case-series study。事前に inclusion criteria を決めて症例を集め、術前にある臨床スコアを測定し、術後一定期間後に外来でフォローできた症例の臨床スコアを再度測定して比較したと仮定した。

**症例:** 100 例。手術の結果に影響を及ぼすとされる特性は全患者で均質であったと仮定した。

**疾患:** 整形外科慢性疾患のような生死にはかかわらないが、QOL に大きな影響を与えるものを対象としたと仮定した。疾患の QOL に与える影響は時間とともに変化し、また、周囲の状況や現在の状態によっても変わらうものと想定した。

**真のスコア:** 術前と術後の調子を表す測定不能な真のスコア(以後、「真スコア」とする)があると仮定した。真スコアは個々の患者の真の重症度(研究の間は一定)に、上記の疾患に関する仮定を満たすための値として正規分布に従ったランダムな値を加えたものとなると仮定した。全患者において正規分布の平均は 0、分散は一定(以後、 $V_{tr}$ )であると想定した。

**測定可能なスコア:** 実際に計測したのは測定可能なスコア(以後、「測定スコア」とする)であると仮定した。測定スコアは真スコアと相関するものと想定し、本研究では両スコア間の相関係数  $r = 0.6$  となるよう、

測定スコア = 真スコア + (平均 = 0、分散 =  $(1-r^2)^{1/2}r \times V_{tr}$ ) である正規分布に従うランダムな値)

と計算した。

**治療:** このモデルで行った治療は個々の患者の真の重症度に全く影響を与えないものと仮定した。つまり、全く効果のない治療を行ったものとした。

**脱落:** 手術を受けたことで調子が悪くなったと感じた患者から来なくなるものと仮定した(worst case analysis)。しかし、真の重症度は術前後で変化がないことから、純粋に術前後でのランダムな値の差だけが影響を与えるものと想定した。上記の仮定を満たすため、術前後での真スコアの変化量(術後の真スコア-術前の真スコア)が悪い症例から順に外来に来なくなり、追跡できなくなるものと仮定した。その結果、最終的な追跡率が 100%、95%、90%、80%、70%、60%、50% となったと仮定した。

### 3.2 シミュレーションの実行

シミュレーション開始時に乱数を実施日時により初期化し、100 例の術前、術後の真スコアと測定スコアを Excel VBA 標

準の乱数生成関数を用いて算出した。術前後での真スコア変化量を計算し、悪い順に並べ替えを行い、最終追跡率毎に Excel VBA 標準の paired-t test 関数を用いて術前後の測定スコアの差が有意かどうかを有意水準 5% で検定した。

これを 1 万回行い、最終追跡率毎に何回術前後の測定スコアの差が有意と判定されたかをカウントし、最終追跡率毎の有意になった確率を計算した。

上記のシミュレーションをさらに 2 回、合計 3 回行い、最終追跡率毎の有意になった確率の範囲を求めた。

さらに、2 次的な評価項目としてモデルごとに術前の真スコアと測定スコアの相関係数、術後の真スコアと測定スコアの相関係数を計算し、1 万回のシミュレーションで計 2 万の相関係数を導出し、その平均値と標準偏差、中央値、2.5 パーセンタイル値、97.5 パーセンタイル値を求めた。それを 3 回行い、その結果を示した。

表 1 術前後の測定スコアの差が統計学的に有意となった確率(最終追跡率毎)

| 相関係数の代表値      |       |       |       |
|---------------|-------|-------|-------|
| 平均値           | 0.599 | 0.597 | 0.598 |
| 標準偏差          | 0.065 | 0.066 | 0.065 |
| 中央値           | 0.603 | 0.601 | 0.602 |
| 2.5 パーセンタイル値  | 0.436 | 0.433 | 0.439 |
| 97.5 パーセンタイル値 | 0.728 | 0.727 | 0.728 |

## 4. 結果

最終的な追跡率が 100%、95%、90%、80%、70%、60%、50% であった場合のそれぞれにおいて、術前後のスコアの差が有意となった確率は 4.9-5.2%、9.8-10.4%、18.3-19.1%、50.2-51.0%、74.8-75.5%、88.8-89.3%、95.1-95.4% であった(表 1)。

真スコアと測定スコアの相関係数の平均値は 0.597-0.599、中央値は 0.601-0.603 と仮定した相関係数 0.6 に非常に近い値であった。一方、標準偏差は 0.065-0.066、2.5 パーセンタイル値は 0.433-0.439、97.5 パーセンタイル値は 0.727-0.728 とばらつきはある程度あることが分かった(表 2)。

表 2 真スコアと測定スコアの相関係数

| 最終追跡率 | 術前後の測定スコアの差が有意になった確率 |       |       |
|-------|----------------------|-------|-------|
| 100%  | 5.1%                 | 5.2%  | 4.9%  |
| 95%   | 9.8%                 | 10.4% | 10.2% |
| 90%   | 18.3%                | 19.1% | 19.0% |
| 80%   | 50.2%                | 50.8% | 51.0% |
| 70%   | 74.8%                | 75.5% | 75.4% |
| 60%   | 88.8%                | 89.3% | 89.2% |
| 50%   | 95.1%                | 95.4% | 95.4% |

## 5. 考察

術前後の真スコアの差が悪い患者が脱落すると仮定した最もバイアスが強くかかるケースでは、治療自体に意味がないにもかかわらず、脱落が 10-20%程度あるだけで術前後の測定スコアの差が有意になる確率が跳ね上がってしまうことが確認された。これは、アウトカム自体が脱落に影響する場合、少しの脱落率でも結果に影響があったとする過去の報告<sup>1)</sup>と一致する。

しかし、脱落率が高くても(50%を超えてもとの報告もある)アウトカムにほとんど影響しないという報告も複数ある<sup>2)3)4)</sup>。これは、その報告で観察された脱落がアウトカムに寄らないものが多く含まれていたためではないかと考えられる。脱落には大きく分けて3つのパターンがあり、全くの偶然で脱落しただけでありバイアスを引き起こさない Missing Completely At Random (MCAR)、観察可能なデータにより脱落が推定できる(脱落によっておこるバイアスをそれらの説明変数によって調整可能である) Missing At Random (MAR)、脱落が観測可能な変数では予測できない(脱落によるバイアスが推定も調整もできない) Missing Not At Random (MNAR)にわけられる<sup>5)</sup>。本研究での脱落は全て脱落時の真スコアによるものであり、MNAR に属するため、バイアスを生じたものと考えられる<sup>6)</sup>。一方、脱落が影響を及ぼさなかったとする報告では、脱落の理由が転居や忙しさなどおもに MCAR か MAR に属するものであり、バイアスを引き起こさなかったのではないかと考えられる。しかし、術後の成績不良での転医はしばしばみられることであり、実際の手術成績の報告では3つの脱落が状況に応じて混在している状態である可能性が高いと推定される。つまり、本研究のような強いバイアスがかかっている状態とバイアスを認めない状態の中間のところに位置しているものと思われる。これを評価するためには、現実の研究での脱落で3つの類型がどのような割合で起こっているかを検討し、モデルに当てはめる必要があるだろう。

本研究のリミテーションは、1つ目に worst case 分析であり最終追跡率ごとに最大ここのまでのバイアスがかかっていることは分かるが、実際の例ではどの程度のバイアスがかかっているかを予測することはできない点である。しかし、各最終追跡率におけるバイアスの最大値が分かっていることは、ある研究がどの程度の信頼性を持ちうるかを評価するにあたって十分に意義があることであることであると考えられる。

リミテーションの2点目は真スコアを定数である真の重症度と全員が同じ分散の正規分布に従う乱数の単純な和で計算したことである。本来の疾患では真の重症度自体も時間とともに推移していくものであるし、ホワイトノイズの大きさも人により大きい人もいれば小さい人もいる。なにより、単純な線形関数では表されない複雑な反応を本来は示すはずである。

## 6. 結論

モンテカルロシミュレーションを用いて、術前後のアウトカムが脱落と関係しているケースで、脱落率が術前後の測定可能な臨床スコアの差の検定にどのような影響を与えるかについて検討した結果、脱落が 10-20%発生しただけで大きなバイアスが発生していることが確認された。この結果は、術前後のアウトカムの差を用いた研究において、追跡率がかなり高くなければ結果はゆがんでいる可能性が否定できないということを示唆するものである。しかし、本研究は worst case 分析であり、より現実に即した解析を行う必要があるものと考えられた。

## 参考文献

- 1) Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? *Eur J Epidemiol.* 2004;19(8):751-60.
- 2) Howe LD, Tilling K, Galobardes B, Lawlor DA. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. *Epidemiology.* 2013;24(1):1-9.
- 3) Solberg TK, Sørli A, Sjaavik K, Nygaard ØP, Ingebrigtsen T. Would loss to follow-up bias the outcome evaluation of patients operated for degenerative disorders of the lumbar spine? *Acta Orthop.* 2011 Feb;82(1):56-63.
- 4) Powers J, Tavener M, Graves A, Loxton D. Loss to follow-up was used to estimate bias in a longitudinal study: a new approach. *J Clin Epidemiol.* 2015;68(8):870-6.
- 5) Little RJA, Rubin DB. *Statistical analysis with missing data.* New York: John Wiley & Sons,1987.
- 6) 坂本 和靖. サンプル脱落に関する分析—「消費生活に関するパネル調査」を用いた脱落の規定要因と推計バイアスの検証 *日本労働研究雑誌* 2006 ; 48(6) : 55-70.