

## 研究部レポート

# アクセプトされる失敗した臨床試験の粉飾法： 無駄のない研究推進のためのピアレビュー研究

医療経済研究機構

研究員 奥村 泰之

### 1. はじめに

「ピアレビュー研究に研究助成を！」という主張が2014年3月12日の米国医師会雑誌（JAMA）に掲載されている<sup>1)</sup>。ここで、ピアレビュー研究とは、「同じ領域の他の専門家の研究や研究費申請の質と適切性を専門家が評価するもの」を意味する<sup>2)</sup>。一流の研究者がピアレビュー研究を重要性の高い研究であると認識していることは、2013年9月に開催されたピアレビュー・生物医学出版国際学会に、世界医学五大雑誌の編集委員を筆頭に、世界32か国の編集委員が参集した事実からも推測できる<sup>3)</sup>。しかし、我が国では、ピアレビュー研究の存在すら認知度が低いことが予想される。

そこで、本稿では、①アクセプトされる失敗した臨床試験の粉飾法に焦点をあてることで、ピアレビュー研究の重要性を伝えること、②無駄のない臨床試験を増やすためには、日本発のピアレビュー研究が必要であると提言することを目的とする。

### 2. 失敗した臨床試験の粉飾法の発見

錯乱（spin）とは、「主要評価項目に関して統計的有意性を得られなかった優越性試験において、実験的治療法の有益性を強調する報告戦略、あるいは読者の注意を主要評価項目から反らす報告戦略」を意味する<sup>4)</sup>。つまり、失敗した臨床試験を粉飾する執筆技術である。

錯乱の存在は2010年にBoutronらが発見した<sup>4)</sup>。

Boutronらは、2006年12月から2007年3月に出版された、主要評価項目に関して統計的有意性を得られなかった優越性試験を系統的に展望している。分析対象となった72研究のうち58%は、アブストラクトの結論を錯乱させている事実が明るみに出てしまった。つまり、半数以上の失敗した臨床試験は、研究成果を「盛って」報告していたのである。

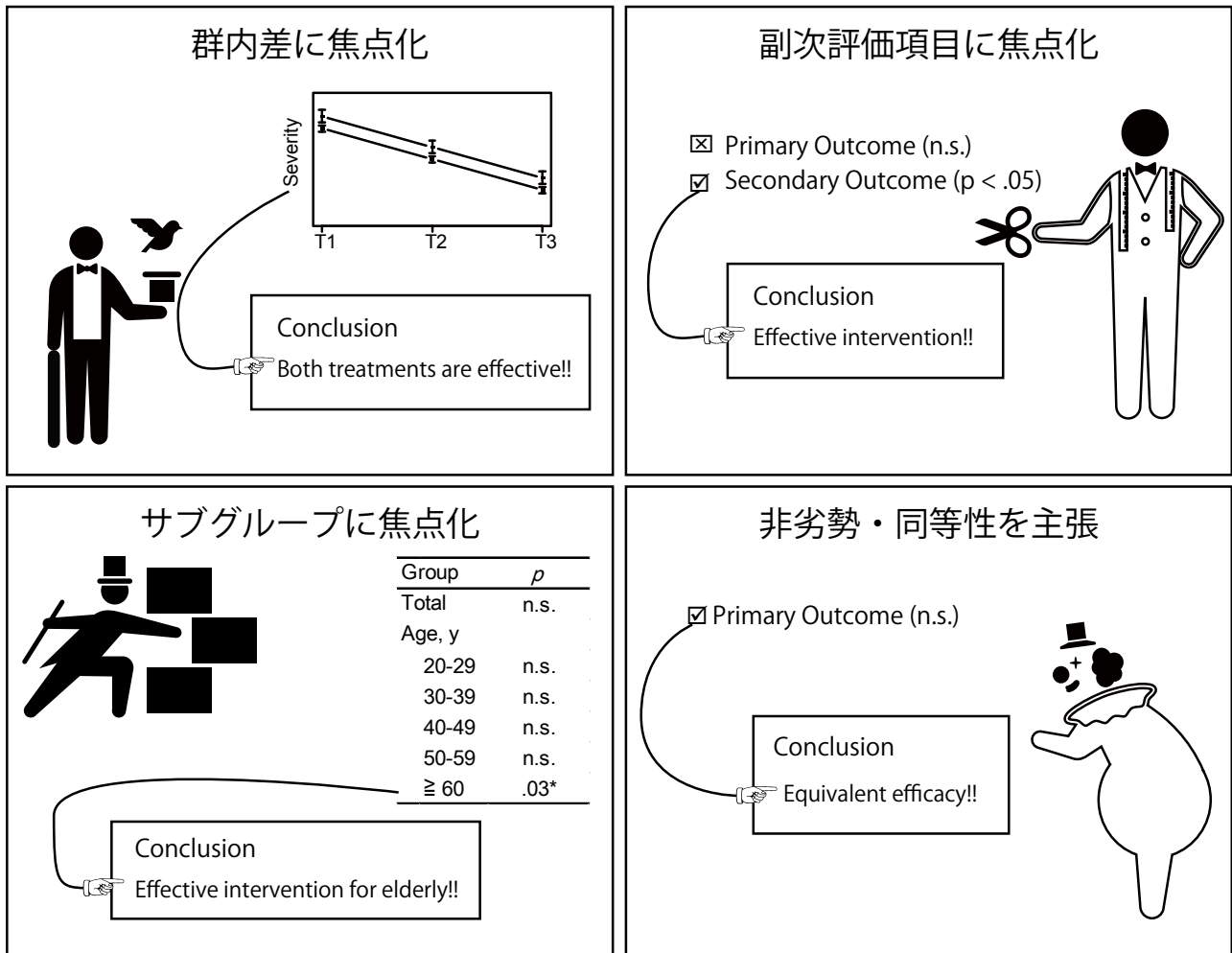
### 3. 実例からみる人気の粉飾法

失敗した臨床試験の粉飾法の中から、巧みな研究者に人気のある4つの執筆技術を具体例と共に紹介する<sup>4)</sup>（図表1）。

#### 1) 群内差に焦点化

Rosaらは、治療抵抗性のうつ病への経頭蓋磁気刺激治療は、電気けいれん療法と比べて、治療開始4週目の治療反応率（介入前後の重症度スコアが50%以上改善）が優れているかを検討している<sup>5)</sup>。無作為化比較試験の結果、治療反応率は、経頭蓋磁気刺激治療が40%、電気けいれん療法が50%であり、統計的に有意な群間差はみられなかった。問題は、アブストラクトの結論で、「治療抵抗性うつ病に対して、両治療法ともに一定の改善に寄与する」と明記していることである。介入前後の各群内の主要評価項目の差をもって、この主張ができるのであれば、並行群間比較試験の存在意義がなくなるであろう。主要評価項目に関して統計的有意性を得られなかった優越性試験であるから、盛らずに、「治療抵抗性

図表1 4つの粉飾法



うつ病治療において、電気けいれん療法と比べた経頭蓋磁気刺激治療の優越性は認められなかった」と結論づけることが望ましかったであろう。

## 2) 副次評価項目に焦点化

Edmondsらは、糖尿病性足部潰瘍への人工皮膚(アプリグラフ®)治療は、通常診療と比べて、治療開始12週以内の創傷完治までの時間が短いかを検討している<sup>6)</sup>。副次評価項目は、治療開始12週時点の、創傷完治率であった。無作為化比較試験の結果、人工皮膚治療の方が通常診療よりも創傷完治までの時間が短い、統計的に有意な群間差はみられなかった。一方、副次評価項目である創傷完治率は、人工皮膚治療が52%、通常診療が26%であり、統計的に有意な群間差がみられた。問題は、アブストラクト

の結論で、「人工皮膚治療は、12週以内の創傷完治率の向上に寄与する」と明記していることである。確かに、副次評価項目の結果は、この結論を支持する。しかし、主要評価項目の結果を無視し、「複数」の副次評価項目のうち、都合の良い結果が得られた一部に焦点化して結論づける科学的妥当性はないのである。仮に、副次評価項目が20個あって、統計的有意性を得られた唯一の副次評価項目に焦点化して結論づけたとしたら、「多数の評価項目が存在することによって偶然得られた差に着目しているに過ぎない」と誰もが理解できるであろう。

## 3) サブグループに焦点化

Onoらは、自殺率の高い地域への複合的自殺予防

プログラムは、通常の自殺予防対策と比べて、治療開始3.5年後の自殺死亡者と自損行為を複合した発生率が低いかを検討している<sup>7)</sup>。副次評価項目は、複合アウトカムではなく、治療開始3.5年後の自損行為の発生率であった。非無作為化比較試験の結果、複合的自殺予防プログラムの方が通常対策よりも自殺死亡・自損の発生リスクが7%低い、統計的に有意な群間差はみられなかった。一方、サブグループである「男性」と「高齢者」において、複合的自殺予防プログラムの方が通常対策よりも、副次評価項目である自損の発生リスクが65%以上低く、統計的に有意な群間差がみられた。問題は、アブストラクトの結論で、「複合的自殺予防プログラムは男性と高齢者において効果が示された」と明記していることである。つまり、副次評価項目の焦点化に加え、特定のサブグループに焦点化して結論づけているのである。主要評価項目の結果を無視し、「多数」のサブグループのうち、都合の良い結果が得られたごく一部に焦点化して結論づける科学的妥当性はないのである。巨額の研究費が動いた期待の戦略研究の成果が、錯乱の一例として挙げられてしまうことは残念である。

#### 4) 非劣勢・同等性を主張

Damjiらは、開放隅角緑内障への選択的レーザー線維柱帯形成術は、アルゴンレーザー線維柱帯形成術と比べて、治療1年後の眼圧の変化量が低いかを検討している<sup>8)</sup>。無作為化比較試験の結果、眼圧の変化量は、選択的レーザー線維柱帯形成術が5.86 mm Hg減、アルゴンレーザー線維柱帯形成術が6.04 mm Hg減であり、統計的に有意な群間差はみられなかった。問題は、アブストラクトの結論で、「選択的レーザー線維柱帯形成術は、アルゴンレーザー線維柱帯形成術と同等の安全性と効果を有する」と明記していることである。非劣勢や同等性を実証するための研究計画を立案していれば、この主張が成り立つ可能性はある。しかし、優越性試験の「差がない」という結果から、非劣勢・同等性を主張することに科学的妥当性はないのである。

#### 4. 錯乱による被害者

錯乱による研究成果の主張には、科学的妥当性はない。しかし、錯乱はデータ改竄などの不正行為(misconduct)の定義には該当せず、一般的には論文撤回(retraction)や正誤表(erratum)による対応が求められることは少なく、錯乱された主張は修正されずに残り続ける。こうした規制の弱さのためか、Boutronらによる錯乱の発見の後も<sup>4)</sup>、錯乱は極めて一般化可能性の高い現象であるという、研究者コミュニティの倫理観を疑わざるを得ない事実が追認されている<sup>9,12)</sup>。

この錯乱による主な被害者は、研究者ではなく一般国民である。研究法に精通している研究者であれば、錯乱による研究成果を「論文」という媒体から科学的妥当性を批判的に吟味するため、欺かれることはない。しかし、一般国民は、錯乱による研究成果を「報道」という媒体から触れるため、欺かれてしまう。Yavchitzらは、2009年12月から2010年3月までにプレスリリースされた優越性試験を系統的に展望している<sup>13)</sup>。41研究は、「論文」「プレスリリース」「報道」の3つの情報源で研究成果を公表していた。まず、論文のアブストラクトの段階で、17研究は錯乱させていることが明らかになった。次に、プレスリリースの段階で、17研究のうち1件だけが錯乱の修正がされていた。しかし、報道の段階では、17研究すべてが錯乱されたまま報道されていた。つまり、論文のアブストラクトが錯乱されていれば、科学的妥当性のない研究成果が国民に伝わるという事実が確認されたのである。

錯乱防止に一番の責任を有するのは、報道記事を執筆するサイエンスライターではなく、学術雑誌の査読者や編集委員である。査読者や編集委員は、錯乱の事実気づき、論文になる前に修正を求めなければ、国民が科学的妥当性のない研究成果の主張に欺かれてしまうことを認識しなければならない。

#### 5. 日本における臨床研究の問題と対策

国際的には様々なピアレビュー研究が実施され、

錯乱の問題以外に、①研究成果の未公表/公表遅延の問題、②選択的報告の問題、③不完全な報告の問題などが明らかにされ、それらの解決のために研究者と研究費配分機関に対する総合的な対策が行われている<sup>14,15)</sup>。一方、これまでの日本発のピアレビュー研究は限られているものの、以下で示すような臨床研究における科学上の問題点が同定され、その対策が提案されてきている<sup>16-19)</sup>。

### 1) 例数設計教育の必要性

奥村らは、2007年1月から2008年12月の間に老年精神医学雑誌に掲載された原著論文を系統的に展望している<sup>16)</sup>。統計的検定を用いた28研究のうち、科学的根拠に基づいた例数設計を行っているものは皆無であることが明らかになった。また、Okumuraらは、1990年から2006年に出版された日本のうつ病研究を系統的に展望している<sup>17)</sup>。臨床サンプルを調査対象とした184研究のうち、80%は科学的に不十分な症例数しか確保していない事実が明るみに出た。これらの結果から、研究者への例数設計教育が喫緊の課題であることを結論づけている。

### 2) 臨床試験登録の義務化徹底の必要性

Yoneokaらは、2010年に出版された日本の無作為化比較試験を系統的に展望している<sup>18)</sup>。1,013研究のうち、臨床試験登録を事前に行っているものは、8%に過ぎないことが明らかになった。また、日本の臨床試験登録システムの一部は、臨床試験登録日や出版日を検索できる状態でないことを指摘している。これらのことから、臨床試験登録システムの改善と、臨床試験登録の義務化の徹底が喫緊の課題であることを結論づけている。

### 3) 研究報告ガイドライン遵守の必要性

Uetaniらは、2004年1月から3月に出版された日本の無作為化比較試験を系統的に展望している<sup>19)</sup>。98研究のうち、研究報告ガイドラインであるCONSORT声明<sup>20)</sup>が規定する報告すべき内容のうち、①例数設計の根拠の記述は23%、②乱数生成法の記述は39%、③割り付けの隠匿化の記述は17%に留まることが明らかになった。これらの結果から、無作

為化比較試験の研究報告の質を高めなければならないと結論づけている。

## 6. 無駄のない研究推進のために

Lancet誌の報告では、生物医学研究の85%以上は「無駄な研究」であり、その原因は、①リサーチクエスションの不適切さ、②研究デザインの不備、③未公表、④研究報告の不備であると指摘されている<sup>21)</sup>。そして、無駄な研究による毎年の損失額は100,000,000,000米ドル以上に相当すると推計されている<sup>21)</sup>。先に紹介した日本発のピアレビュー研究を基に予想すると、我が国における無駄な臨床研究の割合が85%を大きく超える可能性は否定できないであろう。国際的には、科学的根拠に基づいた例数設計や臨床試験の事前登録の必要性は、CONSORT声明やSPIRIT声明などの研究報告ガイドラインで定められているが<sup>20, 22)</sup>、そもそも、我が国では研究報告ガイドラインへの認知度が低いことが予想される<sup>19)</sup>。また、大部分の無作為化比較試験が事前に臨床試験登録を行っていない事実は<sup>18)</sup>、巧みな研究者が錯乱を容易に行える状況であることを意味するのである。

ピアレビュー研究は、科学上の問題点を同定して、その対策を提案することができる。したがって、ピアレビュー研究が盛んになることにより、科学的妥当性のある無駄のない臨床研究が増えることが期待できる。しかし、現状では、日本発のピアレビュー研究は少ないため、他の様々な科学上の問題点が潜んでいる可能性は否定できない。日本における臨床研究の科学上の問題点を同定し、その対策を推進するための研究助成を与える必要があるだろう。臨床研究の実施に伴う、多くの人の善意と時間と資金を無駄にしないためには、質の高い研究を増やす総合的な取り組みが不可欠である<sup>14, 15, 23, 24)</sup>。

## 謝辞

本稿の草稿の段階で、精読し有益な御助言を下された、赤羽隆文様 (医療経済研究機構) に深謝いたします。



## 引用文献

- 1) Malicki, M., von Elm, E., Marusic, A.: Study design, publication outcome, and funding of research presented at international congresses on peer review and biomedical publication. *JAMA* 2014; 311: 1065-1067.
- 2) Peer Review, Research [<http://www.ncbi.nlm.nih.gov/mesh/68017712> (accessed 2014.4.29) ].
- 3) International Congress on Peer Review and Biomedical Publication [<http://www.peerreviewcongress.org/> (accessed 2014.4.29) ].
- 4) Boutron, I., Dutton, S., Ravaud, P. et al: Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010; 303: 2058-2064.
- 5) Rosa, M. A., Gattaz, W. F., Pascual-Leone, A. et al: Comparison of repetitive transcranial magnetic stimulation and electroconvulsive therapy in unipolar non-psychotic refractory depression: a randomized, single-blind study. *Int J Neuropsychopharmacol* 2006; 9: 667-676.
- 6) Edmonds, M.: Apligraf in the treatment of neuropathic diabetic foot ulcers. *Int J Low Extrem Wounds* 2009; 8: 11-18.
- 7) Ono, Y., Sakai, A., Otsuka, K. et al: Effectiveness of a multimodal community intervention program to prevent suicide and suicide attempts: a quasi-experimental study. *PLoS One* 2013; 8: e74902.
- 8) Damji, K. F., Bovell, A. M., Hodge, W. G. et al: Selective laser trabeculoplasty versus argon laser trabeculoplasty: results from a 1-year randomised clinical trial. *Br J Ophthalmol* 2006; 90: 1490-1494.
- 9) Le Fourn, E., Giraudeau, B., Chosidow, O. et al: Study design and quality of reporting of randomized controlled trials of chronic idiopathic or autoimmune urticaria: review. *PLoS One* 2013; 8: e70717.
- 10) Latronico, N., Metelli, M., Turin, M. et al: Quality of reporting of randomized controlled trials published in *Intensive Care Medicine* from 2001 to 2010. *Intensive Care Med* 2013; 39: 1386-1395.
- 11) Lockyer, S., Hodgson, R., Dumville, J. C. et al: "Spin" in wound care research: the reporting and interpretation of randomized controlled trials with statistically non-significant primary outcome results or unspecified primary outcomes. *Trials* 2013; 14: 371.
- 12) Patel, S. V., Chadi, S. A., Choi, J. et al: The use of "spin" in laparoscopic lower GI surgical trials with nonsignificant results: an assessment of reporting and interpretation of the primary outcomes. *Dis Colon Rectum* 2013; 56: 1388-1394.
- 13) Yavchitz, A., Boutron, I., Bafeta, A. et al: Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study. *PLoS Med* 2012; 9: e1001308.
- 14) 奥村泰之, 吉田和樹, 清水沙友里. 臨床疫学研究における報告の質向上に向けて: 研究者の倫理. *精神科* 2014; 24: 551-557.
- 15) 奥村泰之. 臨床研究の再現性向上に向けた国際動向: 無駄のない研究のために. *週刊社会保障* 2014; 2767: 40-41.
- 16) 奥村泰之, 伊藤弘人. 老年精神医学研究に求められる検定力分析、標本効果量および信頼区間. *老年精神医学雑誌* 2010; 21: 93-100.
- 17) Okumura, Y., Sakamoto, S.: Statistical power and effect sizes of depression research in Japan. *Psychiatry Clin Neurosci* 2011; 65: 356-364.
- 18) Yoneoka, D., Hisashige, A., Ota, E. et al: Are Japanese randomized controlled trials up to the task? a systematic review. *PLoS One* 2014; 9: e90127.

- 19) Uetani, K., Nakayama, T., Ikai, H. et al: Quality of reports on randomized controlled trials conducted in Japan: evaluation of adherence to the CONSORT statement. *Intern Med* 2009; 48: 307-313.
- 20) Moher, D., Hopewell, S., Schulz, K. F. et al: CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2011; 340: c869.
- 21) Chalmers, I., Glasziou, P.: Avoidable waste in the production and reporting of research evidence. *Lancet* 2009; 374: 86-89.
- 22) Chan, A. W., Tetzlaff, J. M., Gotzsche, P. C. et al: SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* 2013; 346: e7586.
- 23) 奥村泰之, 原井宏明, 谷晋二, 佐藤寛. 行動療法研究における研究報告に関するガイドラインの特集に向けて. 行動療法研究 印刷中.
- 24) 臨床疫学研究における報告の質向上のための統計学の研究会 [<http://blue.zero.jp/yokumura/workshop.html> (accessed 2014.4.29) ].