

第11回 消えていった仮説検定 (1)

前回に引き続いて、臨床試験の評価方法や解析方法の変遷を振り返ることにする。今回は「かつては実施されたのに、いつの間にか消えていった仮説検定」として、被験者背景の群間比較に用いられた仮説検定を取り上げる。なお、消えていった仮説検定には、「検定自体は今でも実施されているが、新たな手法が登場したため、現在では実施されない手法」というものも含まれる。こちらについても、別途機会を設けて取り上げたい。

さて、ランダム化比較試験の結果を報告する際には、試験治療群別の被験者の内訳(割り付けから試験完了までの被験者数の推移)を示した後、群別の被験者背景を要約する。具体的には、性別や重症度といったカテゴリカル変数であれば、カテゴリごとの被験者数と割合を示し、年齢や体重といった連続変数であれば、平均値と標準偏差または中央値と範囲(四分位点)を算出する。現在では、こうした記述統計量を表に示すことによって要約が完結する。ところが、かつては、この要約表に群間の分布を比較した仮説検定、たとえば Fisher の直接法や t 検定の P 値が記載されていたのである。

この検定が適切かどうかを考える前に、まず被験者背景を要約する目的を考えてみたい。主な目的は3つある。第1の目的は、計画した対象集団と実際に参加した被験者集団との間に差異がないかどうかを評価することである。たとえば、ヘモグロビン A1c が7%以上の2型糖尿病患者を対象とするように計画したが、実際には7%未満の患者もかなり参加したという事態が生じるかもしれない。こうした事態は試験結果の一般化可能性に影響を及ぼすため、選択基準と除外基準だけでなく、被験者背景を示すのである。第2の目的は、試験に参加した被験者集団がその疾患の患者集団を代表しているかどうかを評価することである。たとえば、骨粗鬆症患者を対象とした臨床試験を実施した結果、被験者の約半数を40歳未満の女性が占めたとしたら、その集団で得られた結果を一般的な骨粗鬆症患者の集団に外挿するのは難しいであろう。第3の目的は、群間の比較可能性を評価することで、被験治療群と対照治療群の被験者背景が同様であれば、アウトカムの差は試験治療の差によるものと推定することが可能になる¹⁾。

被験者背景に対する仮説検定は、上記の第3の目的、すなわち群間の比較可能性を評価するために実施したものである。しかし、このような目的で仮説検定を実施するのは、仮説検定の基本原理に反している。仮説検定では、「複数の試験治療群の背景因子 X の分布は等しい」という帰無仮説下で今回得られた結果またはもっと極端な結果が得られる確率を計算し、その確率(P 値)が事前に定めた閾値(有意水準)を下回った場合には、「珍しいことが偶

然起こった」と考えずに、「最初の帰無仮説が誤っていた」と考え、対立仮説(試験で検証したい仮説)を採択する。すなわち、検証したい仮説が存在する場合に実施するのが仮説検定である。しかし、被験者数が少ない場合には、適切なランダム化を実施しても、ある背景因子では偶然不均衡が生じることがある¹⁾。したがって、 P 値がどれほど小さくても、それは珍しいことが偶然起こったにすぎず、 P 値から群間比較が可能かどうかを結論することはできないのである。

さらには、ある背景因子 X で仮説検定の結果が有意水準を下回った場合、その因子 X を共変量として調整した解析(例: 共分散分析)を実施する企業も存在した。しかし、調整する必要があるのはアウトカムと関連する因子のみである²⁾。たとえば、性別の分布に群間で不均衡が認められたとしても、性別とアウトカムとの間に関連がないのであれば、不均衡を調整する必要はない。逆に、アウトカムと強く関連する背景因子が存在するのであれば、最初からその因子を共変量とするように計画すべきであり、仮説検定の結果に応じて調整するかどうかを決定するのはナンセンスである。

以上の問題点が認識されるようになったのは 2000 年代に入ってからで、きっかけとなったのは、2001 年に発表された Consolidated Standards of Reporting Trials (CONSORT) の解説である³⁾。これは、日本科学技術連盟がメディカルライティング教育コースを開講した時期と一致する。当時から現在に至るまで、私は本コースで「治験総括報告書の作成方法」という講義を担当しており、開講当初はかなりの時間を割いて、この検定の問題点を説明したものである。にもかかわらず、受講者からは「なぜ問題なのか」という質問を頂戴することが多かった。おそらく、ルーチンで実施していた検定を否定されたことに納得がいかなかったのであろう。

しかし、CONSORT が浸透するにつれてこの種の質問を受ける機会は減っていき、ふと気がつくと、仮説検定を実施する企業自体が存在しなくなっていた。医薬品医療機器総合機構のウェブサイトに公開されている医療用医薬品の申請資料を見ても、最近承認された品目で被験者背景に対する仮説検定を実施しているものは、まずない。医学論文も同様で、被験者背景の要約表に P 値を示したランダム化比較試験の論文を目にすることは少ない。このため、最近の治験総括報告書に関する講義では、被験者背景に対する仮説検定の問題点をまったく説明していない。

もちろん、この変遷は喜ぶべきことである。しかし、その一方で、2000 年代初頭の講義を懐かしく思うのも事実である。あの頃、受講者の反応は様々であった。CONSORT という国際的なガイドラインに書かれているのだから、そうなのだろうと自分自身を納得させようとする受講者。説明は理解できるが、今ひとつ釈然としないという表情を浮かべる受講者。私の説明に真っ向から反論してくる受講者。「こいつの言っていることはおかしい」と、はなから説明を無視する受講者...。あの頃の受講者たちは、今どうしているであろうか。

と、ここまで考えたところで、重大な事実気がつく。あれから 20 年がたった。当時の受講

者の平均年齢を考えると、半数は既に定年を迎えて退職したはずで、残る半数もライティングの実務からは離れているはずである。したがって、当時の受講者がランダム化比較試験の結果をどうまとめているのかを考えることには意味がない。そう、ときは流れたのである。確実に。

As times go by.

引用文献

- 1) Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869. doi:10.1136/bmj.c869.
- 2) Senn SJ. Covariate imbalance and random allocation in clinical trials. *Stat Med* 1989;8:467-75.
- 3) Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.