

# p 値に関する最近の議論

池田敏広 (参天製薬株式会社)

## 1. はじめに

p 値が発表されてから多くの批判が繰り返されているが、最近の議論では適用場面において改善提案が多くみられる。American Psychological Association (APA) (2001) では、重要な知見あるいは p 値を報告するときにはエフェクト・サイズを記すことを指示した。Nuzzo (2014) では、心理学の 2 回の試験結果が異なる結果であったことに触れ、p 値が多くの科学者が考えているような信頼できるものではないと指摘した。さらに有意性は「効果があるか」を問うものであるが、これにより有意性を追跡する者 (p-hacker) が現れる問題点を指摘している。また「効果があるか」を指摘するのではなく「どの程度の効果があるか」を問うことに言及している。その後、社会心理学分野の論文誌である Basic and Applied Social Psychology (BASP) における Tranfimow and Marks (2015) の編集指針では、論文誌内での帰無仮説有意性検定方式と p 値の提示を禁止した。これらを受け American Statistical Association (ASA) は、Wasserstein (2016) において p 値における声明を発表した。最近では、実際の適用場面において p 値の使用方法の見直しが進められている。日本国内での議論では、河合ら (2003) が p 値のみで重要な研究を要約することの問題点を指摘している。大久保・岡田 (2012) では Cohen (1994) より派生したこれらの p 値の議論についてまとめているが、実際の適用での議論には至っていないのが現状である。

## 2. p 値の問題点

### 2.1. 文献の渉猟

最近の論文に見る議論としては①p 値の誤用②p 値による 2 値の結果により、重要な意思決定をすべきでない③「効果があるか」を提示するべきではなく「どの程度の効果があるか」を示すべきである④良い結果のみを提示すべきでない、とする 4 点に要約できる。

p 値は多くの科学者が想定しているほど信頼できるものではなく、さまざまな解析、他の臨床試験の結果を鑑みて意思決定を下すべきである。この際に「効果があるか」を確認するよりも「どの程度の効果があるか」を多角的な観点から検討すべきであり、都合のよい結果のみを示すのではなく、すべての解析結果を提示すべきである。

## 3. 解析方法

2.1 の②③に着目し、最近議論されている 2 つの解析方法について紹介する。1 つは効果の大きさを示すエフェクト・サイズであり、もう 1 つは p 値を算出したもとで帰無仮説が真である確率を示す Bayes 因数を用いた方法である。

## 4. 結びにかえて

Fisher により p 値が紹介されてから 100 年近くがたち、今日までさまざまな批判が繰り返されてきたが、実地の場面での改革はほとんどみられていなかった。p 値による決定は便利であるが、本来、臨床試験以外のデータも含めてさまざまな角度からデータを解析し、出力された結果によって、意思決定を行うべきである。この判断材料として、「どの程度の効果があるか」という視点で解析を進めていくべきであり、解析した結果についてはすべてを提示すべきである。最近の動向として、実地におけるガイドラインや編集指針に p 値の問題点を考慮した対応がとられ始めており、今後、改革がどのように進んでいくか見ていく必要がある。一方で柴田 (2016) が指摘しているように、河合ら (2003) にみる多くの議論はあるものの、日本で実地の場面で適用される動きは、あまり見ることがない。日本でも今後に精力的な検討が是非とも必要である。

(参考文献)

- [1] Wasserstein, R. L. and Lazar, N. A. (2016). The ASA's Statement on p-value: Context, Process, and Purpose. *The American Statistician*, 70(2), 129-133. [丸尾和司・後藤昌司. (2016). p 値に関する ASA の提言: 背景, 過程, 目標. 特定非営利活動法人医学統計研究会. 論文講究録. 2016-G2-1].
- [2] Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, 49(12), 997-1003
- [3] Nuzzo, R. (2014). Scientific method: statistical errors. *Nature*, 506, 150-152.
- [4] Tranfimow, D. and Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, 37, 1-2.
- [5] 大久保街亜・岡田謙介. (2012). 伝えるための心理統計: 効果量・信頼区間・検定力. 勁草書房.
- [6] 河合統介・栗林和彦・濱崎俊光・後藤昌司. (2003). 統計科学の足もと: p 値と信頼区間. *天理医学紀要*, 6(1), 71-79, 2003.
- [7] 柴田義貞. (2016). 放射線リスク-確定的影響と確率的影響-. *行動計量学*, 43(1), 35-43.