

編集後記

2019年6月20日の朝日新聞科学欄に、「統計的有意’は誤解の温床」とか、「‘やめるべき考え方’に研究者800人賛同」というような見出しが踊った。Natureの statistically significant, non-significantに関する記事を紹介したもののだが、極めて誤解を招きやすい表現になっている。実はASA (American Statistical Association) も、2016年に p 値に関する同様のCommentを出し、さらに2019年にまた特集を組んでいる。

検定 p 値は言うまでもなく、それを得た試験の例数、信頼性、再現性をきちんと評価すること、有意とされている効果の大きさを評価すること、同時に類似の他試験との整合性を議論することが必須だが、いつの間にか p 値だけが金科玉条のように独り歩きするケースが生じているのも事実である。心理系のJournalで有意性検定や、 p 値の記載を制限するところも出現し事態は深刻であるが、おそらく、新しい発見を主張するのに過度に p 値を重んじてきたことへの批判と反省が根底にあるのではないかと思われる。

批判すべきは表層的な p 値のみによる二者択一的な有意性議論なのだが、それを統計的検定不要論のように過度に喧伝する向きがあるのは残念である。誤用が問題であるのに、「統計的有意性」そのものが怪しい基準と受け取られるようでは困る。それを突き進めると検定の反転から得られる信頼区間や、統計推測そのものまで否定されかねず、実際、そのような極論も時に見られる。『臨床評価』誌では30年以上も前に、当時まかり通っていたNS (Non-significance) 同等に関連してこの議論を散々しているのので、この機会に、論文「臨床試験における統計的諸問題」(1)(2)(1986年、14巻3,4号)のpdfを下記の同誌サイトにアップした。

<http://cont.o.o7.jp/>

当時は日本の製薬メーカーからFDAへの新薬許認可申請が、統計解析の不備から次々と却下され、とくに多重性問題で製薬界が騒然としていた時期でもある。これらの論争を受けて厚生省は1992年に、同等性、多重性への対処法を盛り込んだ「臨床試験の統計解析に関するガイドライン」を発行している（『臨床評価』20巻1号に収載）。これはICH E9「臨床試験のための統計的原則」に6年先駆けてのことである。今、EBMが叫ばれ、それは1991年に初めてGuyattにより提唱されたとされている。筆者はEBMとは、直観から科学への転換と解釈している。その解釈でよければ、厚生省ガイドライン制定に至った1980年代の官民を巻き込んだ同等性、多重性論争こそが、日本のEBMの先駆けではなかったかと思われる。

なお、拙著「実例で学ぶデータ科学推論の基礎」（岩波書店、2018年）の第8章にも、NS同等や多重性問題で混乱していた当時の日本の状況と、それがどう解決されていったかを述べている。5通りの多重性とそれへの対処が紹介されているが、要は思い入れ、思い込みを廃して如何に科学的アプローチが取れるかということである。本号は国際製薬医学会(IFAPP)の国際会議の特集号であるが、医薬品開発の評価に統計的方法是は不可欠であることを今一度思い起こして欲しい。昨今の統計を取り巻く環境から、統計理論を空論に終わらせず、実問題の解決に適切に応用することの重要性を改めて感じている。

(広津千尋)