

「P 値」至上主義の終焉にむかって

私が疫学・生物統計学の研究・教育のかたわら、臨床研究支援を始めてから20年あまりが経ちました。今から20年前といえばウィンドウパソコンが発売され皆がインターネットを普及しだした頃で、携帯電話はなく外出先で連絡を取るのには電話ボックスといった時代でした。さて研究支援、特に統計相談に目を転じるとどうでしょうか？最近では高度で難解な質問は増えたとはいえ、いまも「P値の出し方を教えてほしい」という質問は現役として活躍？しています。おそらくこの質問の根底には「臨床研究ではP値を出さなければならない」という呪縛・幻想があるからだと思います。一方海外に目を転じてみるとどうでしょうか？海外の一流学術雑誌では「P値のみ」書かれた論文が掲載されることは皆無であり、ハザード比やリスク比などに95%信頼区間をつけることがもはや常識とされています。医学分野ではありませんが、編集委員会がP値を禁止する学術雑誌まで現れる事態になっています (Psychology journal bans P values. Nature 2015; 519: 9)。

この一連の「P値のみ」研究への批判は、科学実験の成果に対する「再現性の危機 (replication crisis)」の問題と密接な関連があります。この「再現性の危機」とは「統計的有意性のあった研究成果について、再試験すると結果が再現できない (有意差が出ない)」という深刻な問題のことで、主に心理学分野で発生した問題でもあります。この問題は統計的には検出力 (パワー) の問題であり、研究目的にあった適正なサンプルサイズ的设计、研究計画の問題までさかのぼることができるのですが、それとともに研究者の「P値崇拜」の弊害とも捉えることもできます。

この深刻な事態に対応するため、米国統計学会は統計的有意性とP値に関するASA声明「The ASA's Statement on p-Values: Context, Process, and Purpose」を発表しま

した (なお日本の生物統計学の学会である日本計量生物学会のホームページ (<http://www.biometrics.gr.jp/news/all/ASA.pdf>) にはこの翻訳が掲載されています)。声明の内容は6項目で構成されていますが、重要な2つを抜粋します (【 】内は筆者注)。

1. 科学的な結論や、ビジネス、政策における決定は、P直がある値 (訳注: 有意水準) を超えたかどうかにのみ基づくべきではない。【P値のみ判断の否定】
2. P値は、それだけで統計モデルや仮説に関するエビデンスのよい指標とはならない。【P値以外の指標の重要性】さて医療・公衆衛生学分野に戻ると、この「P値のみ」を使用する論文への批判は今に始まったことではなく、古くは1978年のRothman KJ. A show of confidence. N Eng J Med. 1978; 299: 1362-3. までさかのぼります。その後、疫学では「検定を使うべきか推定を用いるべきか？」の論争が繰り返され、最終的には区間推定 (95%信頼区間) の有用性が広く理解されることとなりました。この区間推定の有用性については私も拙文 (東邦医学会雑誌. 2014; 61: 238-243) を書きましたので、興味を持たれた方はご一読下さい。

20年経って電話ボックスに入り電話をしていた時代は終わりを告げ、携帯電話に移行しました。臨床研究の分野でも、P値と有意性のみで研究の善し悪しを判断した時代が終わり、研究デザインの良さや治療効果の大きさを議論する時代に移行しています。医療統計学分野ではこれからもこの潮流に沿った研究・教育などの活動を展開し、先生方との共同研究を通じ本学に普及させたいと思っておりますので、ご支援のほどよろしくお願いいたします。

(医学部社会医学講座医療統計学教授: 村上義孝)

DOI: 10.14994/tohoigaku.2017.64-03-208