

医学生物学分野における高被引用論文の オブソレッセンス

天野 晃
鹿児島大学

論文（集合）におけるオブソレッセンスの捉え方はさまざまであるが、一般的には（被）引用半減期や Immediacy Index / Impact Factor 比などの指標が用いられることが多い。一方、本研究ではこのような既存の指標のみからでは見えにくい被引用数の変動やこれによる順位の変動のようなダイナミクスに着目する。本研究では FTP 版 PubMed annual baseline（2023）を用い、1946 年から 2020 年までに出版された論文の参照情報を収集、75 年間の引用関係を整理した。この引用関係において、参照を持つ論文はおよそ 700 万、少なくとも 1 回の引用を受けた論文はおよそ 1900 万、延べ参照数はおよそ 2 億であった。

「高被引用論文」の定義は、最も引用を受けた上位 1000 件の論文とした。この高被引用論文群は引用を受けた論文全体の 0.005%であり、被引用数では全体の 1.7%を占めていた。また、被引用数とそのランクの関係を Zipf 式 $[Cr^{-a}; C, a: \text{係数}; r: \text{ランク}]$ にフィットさせたところ、係数 a の値は 0.5 とかなり低く、十分に被引用数の高いもののみを抽出できていると推察される。まず高被引用論文全体の特性を確認した。全体的な傾向として、新しい論文ほど出版直後から引用を受けているように見える（図は省略）。出版年では 10~14 年前のものが最多であった。順位変動については、Kendall の順位相関を年ごとに（前年との比較）算出したところ、初期においては変動が激しいものの概ね 0.7~1 の範囲であり次第に減少傾向にある（図は省略）。次にオブソレッセンスについて観察した。オブソレッセンス到達の定義を、(1) 出版から最大被引用までの期間 + 減衰期間 + 減衰後の期間 ≥ 10 年、(2) 出版から最大被引用までの期間 $\leq$ 減衰期間 + 減衰後の期間、(3) 最大被引用の年を経たあとに年間被引用数が被引用数累積の 5%未満となる年が存在すること（最大被引用の年からこの最初の年までを減衰期間とする）、とした。この結果 142 論文がオブソレッセンス到達と判定された。オブソレッセンス到達論文とそれ以外の論文の特徴差を確認すると、出版から最大被引用までの期間の最頻値が 4~5 年/12~13 年、減衰期間の最頻値が 6~7 年/6~7 年、出版から減衰期間までの最頻値が 16~17 年/18~19 年であった。オブソレッセンス到達論文の出版から最大被引用までの期間および減衰期間については過去の報告（参照は省略）と整合的であった。オブソレッセンスとリバイバル性との関係を検証するために、オブソレッセンス到達論文とそれ以外の論文とで多峰性に差があるか検定を行なった。多峰性テストでは、Gaussian Mixture Model、Kernel Density Estimation、単調増加を検知する独自アルゴリズムによる複合判定を行い、ピーク数を推定した。オブソレッセンス到達論文では 57%、それ以外の論文では 58%に多峰性が認められ、双方のピーク数に対する Mann-Whitney 検定の結果 P 値は 0.97 であり有意差は認められなかった。なお、これは多峰性テストに必要な期間に満たない論文（19 件）を除外した結果である。