

Regular Article

医学研究評価が直面する構造的な諸問題： 医学研究者を対象としたWEBアンケート調査

島田裕平^{1,2}

箕浦明³

桑原恵介^{4,5}

近藤誠⁶

福島紘子⁷

杉山雄大^{2,8}

Abstract

医学研究評価では、インパクトファクターなど指標への偏重が問題視されてきた。こうした非自律的な研究評価慣行が採用され続ける原因を明らかにするため、医学研究者が自律的な研究評価を行う上で直面する課題を分析する。

先行研究を価値・手段・公正の三つに大別して整理した上で、医学研究者を対象としたアンケート調査の自由回答記述を用いて実証を試みた。98件の回答が対象となった。

既存の評価慣行が教育や臨床など特定の価値を反映しないと指摘される中で、定量的評価は次善の手段として位置づけられていた。また、評価者の主観や認知的限界に起因する不公正な評価にリスクを感じていた。

医学研究者は、評価の前提となる価値体系を変えずに定性的評価へ移行した場合、一層不公正な評価を受ける恐れを感じる中で、定量的評価を諦観的に受容していると考えられる。指標への偏重は、手段自体よりも価値や公正の問題として捉え直すことができる。

以上の構造的問題を前提にすれば、医学研究評価は継続的なガバナンスを通じて不断に再考される必要がある。

キーワード：研究評価、研究公正、インパクトファクター、ガバナンス

¹ 東京大学大学院 法学政治学研究科

² 国立健康危機管理研究機構 国立国際医療研究所 糖尿病情報センター

³ 昭和医科大学 医学部 衛生学公衆衛生学講座

⁴ 横浜市立大学大学院 データサイエンス研究科 ヘルスデータサイエンス専攻

⁵ 横浜市立大学医学部公衆衛生学

⁶ 大阪公立大学大学院 医学研究科 脳神経機能形態学講座

⁷ 筑波大学 医学医療系 小児科

⁸ 筑波大学 医学医療系 ヘルスサービスリサーチ分野

Corresponding Author (E-mail: sugiyama.t@jihs.go.jp)

Medical research evaluation has long been criticized for its overemphasis on metrics such as impact factors. To clarify why non-autonomous research evaluation practices persist, this study analyzes the challenges medical researchers face in achieving autonomous research evaluation.

After categorizing prior research into three main areas—value, means, and fairness—we attempted empirical analysis using open-ended responses from a questionnaire survey targeting medical researchers. A total of 98 responses were analyzed.

While existing evaluation practices are criticized for failing to reflect specific values like education or clinical practice, quantitative evaluation was positioned as a second-best means. Researchers also perceived risks of unfair evaluation stemming from evaluators' subjectivity and cognitive limitations.

Medical researchers appear to accept quantitative evaluation with resignation, fearing that shifting to qualitative evaluation without changing the underlying value system would result in even more unfair assessments. The overemphasis on metrics can be reframed as a problem of values and fairness rather than the means themselves.

Given these structural difficulties, medical research evaluation requires constant reexamination through continuous governance.

Keywords: Research evaluation, research integrity, impact factor, governance

1. はじめに

現代の研究活動は、国家政策や社会的要請と深く関係するようになった(城山, 2018, 100; ネルソン, 2012, 72)。例えば、科学技術・イノベーション基本法を基にした研究費改革や、アウトカム指標に基づく国立大学法人改革が挙げられる。これらの改革は、社会的課題への対応や産業競争力の向上を目的としており、特定の領域に資金を集中する仕組みが導入されている(標葉・林, 2013; 標葉, 2020, 44-69; 国立大学等の独立行政法人化に関する調査検討会議, 2002)。

これに対応して、かつて研究者共同体の内部での研究の質の維持・向上のための営みとして発展してきた研究評価は、近年では国家や社会など外

部からの要請に応えるべく変化を迫られている。1995年に成立した科学技術基本法に基づいて、公的研究費を増額すると同時に、厳正な研究評価の実施が求められるようになった。それを受けて、「国の研究開発全般に共通する評価の実施方法の在り方についての大綱的指針」では、研究資源配分や国民への理解獲得のために、評価基準の明確化が求められた。また、国立大学法人化以降、中期目標・計画に基づく機関評価が導入された。主に定量的指標による研究評価を通じて研究活動を管理する仕組みが制度化された(日本学術会議, 2008)。

こうした科学コミュニティ¹外部から管理された研究評価が研究の多様性を損ない、長期的には

¹ 本稿において科学コミュニティや医学コミュニティとは、研究領域に基づく比較的同質的な研究者の集団を指す。具体的には、学会やジャーナル共同体である。この定義は相対的なもので、医学コミュニティが医学会総会など医学研究者集団全体を指す場合と、個別の診療科や専門領域を指す場合がある。本稿では、後者を特に強調する場合には各医学コミュニティという。

イノベーションを阻害する可能性が指摘されている。大学は、企業における収益のような明快で単一の業績指標を持たないため、評価基準の設定が難しく、指標を単純に向上させることで研究活動の質が上がるとは限らない（大学改革支援・学位授与機，2022；藤田，2001；西山，2019，71-74）。むしろ、指標向上を目指した過度の競争が、研究者間の協力や研究の自由を損なう危険性も指摘される。運営費交付金の削減が進む中で、大学が独自の裁量を発揮しにくい構造になっていることが問題とされる（岡村，2020）。

主に定量的指標を通じて外部から管理された研究評価の限界が指摘される中で、研究者自身が評価基準や評価手法を自ら設定する自律的評価は、新たな役割を期待されている（城山，2018，97-100）。自律的な研究評価を通じて自由な研究環境を維持することで、政策目標を達成しつつ、研究の多様性と創発性を確保できるとされる（有本他，2016，170；山本，2007）。そこに科学コミュニティ自身が自律的に研究評価を行う新たな意義が浮かび上がる。社会的要請に応えるためにも、科学者の専門知識を活用し、研究評価を柔軟かつ多面的に行うことが求められている。

自律的評価の欠缺は、医学研究者の間においても問題である。医学研究領域においては、研究と教育に加え臨床業務も行う研究者がいたり多分野で行う研究もあったりするなど、研究や研究者を評価するにあたっての基準について合意の必要性が高い。また、研究不正の専門分野ごとの事案数は、「医学・歯学・薬学」だけで自然科学系の60.3%、人文社会学を含めた全体の30.7%を占めており、最大である（松澤，2013）。これまでこの問題は主に定量的評価への偏重という形式で議論されてきた。特にインパクトファクター（IF）

の誤用が問題視され、日本医学会による「医学雑誌編集ガイドライン2022」では、その適切な利用が促されている（日本医学雑誌編集者会議，2022）。第31回日本医学会総会でも、学術プログラム「将来に向けた医学研究評価のあり方について考える」が催され、量的指標に偏重した医学研究評価の今後が議論された。日本の医学研究者を対象としたアンケート調査では、67.4%が雑誌のIFを研究評価において重視すると回答し、特に若手や基礎医学（解剖学・生化学など）の研究者が定量的指標を重視していた。一方で、88.8%はIFなどの定量的指標では「重要な研究でも評価されないことがある」と回答したことも報告された（Minoura et al. 2024）。

このように、医学研究者自身が医学研究評価の抱える限界を強く認識するような状況にあるにもかかわらず、なぜIFなどの定量的指標に偏重した既存の研究評価のあり方は変わらないのであろうか。この問いに対する端的な回答は、評価者が定性的な評価を採用しないためというものであるが、それではなぜ定性的な評価は避けられているのかを説明しない。つまり、一見不合理に見える現在の医学研究評価がどのような合理性に支えられているのかを明らかにする必要がある。医学研究領域において研究評価の問題の背景やそうした状況を作り出す構造について全体を整理しかつ実証した研究は私たちの知る限りでは存在しない。

本研究では、医学研究者が自律的な研究評価を行う上で直面する困難性や課題を分析する。研究評価をめぐる問題の背景や構造についての先行研究を価値・手段・公正の三つに大別して整理した上で、アンケート調査の自由回答記述を用いて理論的整理に応じた実証を試みる。その際に、主に評価者の能力的限界あるいは時間など各種の資源

的限界によって、医学研究評価は構造的な困難を抱えていることについて特に議論する。そうすることで、研究評価を行う医学研究者が困難性や課題に直面することで、定量的評価を選択するに至る認識枠組みを明らかにする。この作業を通じて、自律的な医学研究評価ならびに医学コミュニティによる研究評価のガバナンスがいかに難しいのかを理解することを目指す。

2. 問題の整理

研究評価制度を考えるうえでは、優れた指標の開発や評価の方法論の深化を目指すという方向性がある（平澤，2004）。他方で、評価の対象となっている活動それ自体の構造に着目し、対象の活動との関係において評価が機能する条件を検討することに重点を置く分析も存在する（ワイス，2014）。そこで本研究では狭義の評価手段に限らず、評価をめぐるステークホルダーの合意形成を含めた構造を分析した。評価の目的に応じて一義的で明確な評価を目指す理論を前提に、本稿ではむしろ一義的で明確な研究評価がなぜ困難なのか、なぜ合意形成に失敗するのかの構造的な要因に注目する²。ここでは困難性を理論的に議論することにより、実証分析のための視座を設定する。

医学研究者による自律的な研究評価の困難性を考えるうえで、本研究では問題を三つの観点に分類する。すなわち、自律的に研究評価を行うため

には、評価すべき価値基準を明確にし（価値）、適切に測定する手段を選択し（手段）、計画通りに評価を実施する（公正）必要があると考える。そこで以下では、先行研究で指摘された研究評価の抱える困難性を価値、手段、公正の各観点から整理した。それぞれ、何を評価すべきか、どのように評価すべきか、どうしたら適切に評価できるかという問いに関係する。価値と手段については研究評価の制度設計段階の問題であるのに対して、公正については制度の運用面での問題である。

2.1. 価値をめぐる困難性

研究評価において何を評価すべき価値とするかは合意が困難であることが多い³。実効的な評価のためには、評価プログラムを作成する段階で何を評価すべきかを確定させておくことが理想的だが、現実には評価する価値が一義的に決定されていることは稀であり、常に不確実性が存在する。ここでは価値をめぐる生じる困難性を医学コミュニティ内部の合意形成と医学コミュニティを超えた合意形成という二つに分けて整理する。この二つは現実には渾然一体であるが、医学コミュニティ自身による研究評価ガバナンスを考える際には場合分けして検討する意義がある。

2.1.1. 医学コミュニティ内部の合意形成の困難性

従来、研究評価において何を評価するかが問題

² 評価制度が不可避免的にステークホルダー間の合意形成の困難を抱えているという理解について田辺（2005）、西出（2020）を参考にした。

³ 研究評価における価値について上手く合意形成がされた事例として米国 CDC における内部クリアランスがある。そこでは所属職員の発表する出版物は事前に内部の研究評価システムである e-Clearance を経ることとされており、これは CDC が組織として一貫したポリシーを保つ目的で行われている（The Centers for Disease Control and Prevention, 2020）。ここでは「国民を健康の脅威から守る」という最終的なミッションを CDC 職員が共有していると言える。

になってこなかった背景として、各医学コミュニティの内部においては価値基準が共有されているという想定(マートン, 1983)があったと思われる。この想定のもとでは、研究評価は同じ専門分野の科学者相互によって行われるべきであり、合意形成が容易である。しかし、現実には同じ研究領域の研究者の間でも何を評価するかは論争的であることが多い。

例えば、研究において重視する要素として、臨床的あるいは政策的な問題解決への貢献といったインパクトの側面を挙げる者もいれば、研究の独創性や解析の網羅性といったプロセスの側面が重要だと考える者もいるだろう。また、研究以外の要素が評価に入り込むこともある。往々にして医学研究者は研究と臨床という異なる価値を複合した評価を受けることがあるが、このような評価においては研究と臨床をそれぞれの程度重視するかといった比重によって評価の結果が変わり得る。

各医学コミュニティの内部において、どの要素をどの程度評価するかについて事前に合意を形成することで一定程度この問題は解決可能であるという点がまず肝要である。評価目的を事前に定めて、それに従って評価を行うことが透明で適正な評価につながる。

そのうえで、こうした合意形成はいくつかの困難性に直面する。例えば研究費配分などのゼロサム的狀況を想定した場合、どの要素をどの程度重視するかの比重を決定することは評価結果を強く予想させるものであるため、利害対立が先鋭化し合意が難しくなる場合もあるだろう。また評価項目間の比重に合意できたとしても、その評価項目を概括的にしてしまえば、ある評価項目の中でどの評価要素を特に重視するかは個別の評価者に委ねられることになってしまう。そもそも、医学コ

ミュニティ内部で序列的な価値基準を作ることによって、自由な研究活動を阻害し研究の多様性を損なう恐れもあることに注意が必要である。

2.1.2. 医学コミュニティを超えた合意形成の困難性

研究評価が評価者の合意形成を必要とするにもかかわらず、現実の評価場面ではそもそも異なる医学コミュニティに属する評価者や非科学者が評価を行う場合もある。例えば、大学における教授採用のための選考において、当該研究領域外の専門家である教授陣による評価が行われることがある。政府の政策的な研究投資には政治家や行政官が評価者となる場合もある。このような場面では、研究評価の目的に即した評価要素について合意形成が必要になる。研究領域が細分化しサイロ化しているという現状も、この問題を一層困難にさせている(日本学術会議, 2007)。

評価者の関心の多様性は評価者間の合意形成を難しくする。科学者は科学自体の発展を志向する一方で、政府機関や政治家など研究への出資者は経済的なインパクトを求めることが多い(ネルソン, 2012)。科学者の中でも基礎研究と応用研究では、科学に内在的な真理追求を目指すのか、社会に与えるインパクトを高めることを目指すのかという方向性の相違があるかもしれない。あるいは、歴史学や人類学に近い分野では、そもそも研究の価値は発見した結果に限られず、資料発掘や保存という活動自体にある場合もある。このような場合、分野を超えた画一的な評価基準を設定することは難しい。

注意が必要なのは、この問題は個別の医学コミュニティごとに適切な評価軸を設定したとしても、解決するわけでは無いという点である。例えば教授選や学会賞選考を考えると、評価者の個別の専

門は異なれども比較的近接領域であるから、各コミュニティの評価基準を相互に調整することで合意形成に相対的には至りやすいと思われる。しかし、他の科学分野の研究者や非研究者を含めた評価の場合には、各コミュニティで研究評価管理を行ったとしても調整がより困難となることが予想され、評価内容が曖昧なまま評価が行われるおそれが高まる。

2.2. 手段をめぐる困難性

たとえいかなる価値を評価するべきかが明確に定まったとしても、実際にその評価を行う手段により価値の測定程度は変動し得る。したがって評価に利用する手段が有する困難性も検討すべき論点である。

2.2.1. 定性的評価の困難性

定性的評価においては、関連分野の専門家によるピアレビュー⁴を採用することがほとんどである。当初ピアレビューは社会的属性に左右されない公平なシステムとして注目された（Zuckerman and Merton, 1971）が、近年ではむしろその機能不全が注目されており（江間, 2013）、万能薬ではないことにも注意する必要がある。

定性的評価においては、定量的評価によって反映することができないような要素についても考慮することが可能になる一方で、比較可能性や一般

化可能性は低下し、状況認識を共有することが難しくなるというトレードオフが存在する（藤垣 他, 2004）。定性的評価を行うことにより評価価値が不透明になりがちであるという特性が、定量的評価への依存を正当化する可能性がある。

2.2.2 定量的評価の困難性

定量的評価に直接的または間接的にしばしば利用されるのは、被引用数である。被引用数はある時点でのある論文に対して規定され、引用の回数を通じて引用された研究の価値を測定する指標である。引用されることは必ずしも好意的に受容されたことを示すわけでは無いが（藤垣, 2003）、引用された科学コミュニティにインパクトを与えたと解釈することができる（藤垣 他, 2004）。しかし、ここでのインパクトとは研究活動における実用上のインパクトであると解すべきことに注意が必要である⁵（Garfield, 1979）。また、引用は研究分野における新規出版論文数や引用慣行により大きく左右されるため、分野を超えて直接比較することが難しい。さらに、単純な被引用数の合計を研究者の評価指標として用いる場合もあるが、各研究に対する研究者の-effortを考慮したものではないため、分野における共著に関する慣行によっても左右される指標である。

医学系分野でしばしば重視されているのが、IFである。IFは1年に1回、雑誌に対して付与され

⁴ 狭義のピアレビューは関連専門分野の研究者によって「研究の手続き的正確さを保証し、結果の妥当性を確認し、そして論文の出版や助成金の配分等研究に対する報奨を与える体系的な方法論」と定義される（江間, 2013；Chubin and Hackett, 1990）が、本稿では人事評価等を議論に含むため、個別の研究に限定せず研究者を含む評価として考える。

⁵ 例えば、新規手法の解説論文の引用数と新規な機序発見の論文の引用数では、学術的な意義が異なるものの、研究上の役に立つと考えられたという「実用的な」点においては比較可能であると解釈されている。

る。よく用いられるのは直近の2年ないし5年の間に雑誌に掲載された論文の平均引用数であり、当該雑誌が研究分野に与えるインパクトを定量化したものだと考えられている。重要なのは、IFが雑誌の評価指標であり、各論文の評価に適さないことである。ある雑誌における論文の引用数の分布が正規分布に近似できるのであればまだしも、実際は負の二項分布に近く、個別論文の質を測る代替指標として雑誌の平均値（つまりIF）を用いることには合理性がないと言われている（小野寺, 2003）。また、藤垣他（2004）は、高IFの雑誌の査読を通過したことをもって論文を評価することを非合理的であるとしている。しかし実際にはIFの流用は広範に行われている。こうした状況に危機感を覚えた研究者を中心に起草された「研究評価に関するサンフランシスコ宣言（DORA）」（2012）では研究評価を行う際にIFを使用しないことが提言されている。

h-indexはHirsch(2005)によって、研究者の生涯の業績を示す目的で作成された。ある研究者について「h回以上引用された論文数がh報以上であることを満たす最大のh」として計算され、ある時点でのある研究者に対して算出される指標である。単純な引用数の和ではなく、論文数も要求することにより、研究活動全体を評価しようとしている。この指標も研究分野に左右されるという被引用数に内在的な問題を抱えている。また、元々研究者の生涯の業績を示す目的で作成されたため、現役研究者の評価では年齢の高い研究者や若くして引用数の高い論文を発表した研究者が有利であ

るという傾向がある（孫, 2017）。

このほかにも定量的評価の方法は様々に開発されており、指標の有する限界を超えるべく改良が続けられている。例えば、Field Weighted Citation Impact(FWCI)⁶は、対象となる論文の被引用数を、その論文と同じ出版年・分野・文献タイプの論文の平均で基準化することで異なる専門分野における研究者の業績を比較することが可能となる（福成, 2014）。たしかに専門分野によって左右されるという被引用数の限界を一定程度緩和しているものの、比較可能な分野を一律で区分することには限界がある⁷。このように、少なくとも現在の段階では、手放しで信頼できる指標は存在しない（調, 2013）。定量的評価指標を開発する立場である計量書誌学の研究者自身が、指標は定性的評価の補完として利用するのにとどまるべきであると諫めた「研究計量に関するライデン声明」を公開したことに象徴的である（Hicks et al., 2015）。評価する価値についての合意形成の不在は、新たな定量的指標を用いることによっては解決せず、むしろ評価に関する本質的な課題から目を背ける可能性があることに注意すべきである。

2.3. 公正をめぐる困難性

評価において公正さも重要な論点である。ここでは、公正さを評価基準と評価実践の乖離として定義する。本稿では恣意的でない乖離も不正として捉えることにより、不公正な評価を帰責するよりも、むしろ多くの場合、評価には一定の不公正さが不可避的に付随する構造にあることに注目

⁶ ある論文P1のFWCIは、(P1の被引用数) / (P1と同じ出版年、同じ分野、同じ文献タイプの全論文の平均被引用数)と定義される（Elsevier, 2019: 46-8）。

⁷ 2025年1月6日時点においては、13分野（Health Sciences, Biological Sciences, Life Sciences など）と14文献タイプ（Article, Review, Erratum など）にとどまっている（Elsevier, 2019）。

する。

仮に評価価値と手段について合意ができた場合においても、評価が公正に行われるためには、評価者が適切に評価を行うことが必要である。しかし現実には、恣意的な歪曲の問題が存在する。これは評価の透明性には限界があることや、評価における評価者の専門性に基づく裁量が伴うことに付随する。そもそも評価従事者には評価を自身の都合に良いように歪めるインセンティブが存在する（田辺，2005；西出，2020）。評価を機能させる条件の例としては、評価が事後的に検証可能であることがあるが、研究評価は高度に専門的である上、研究機密や不正防止の観点から評価プロセスを完全に透明化することはできない。評価不正のディスインセンティブを高めるためには、罰則を規定したり評価謝礼を高額に設定したりすることが考えられるが、親しい研究者への優遇等の曖昧な事例についてまで適切に監視することは難しい。この場合、評価を機能させる要素としては、評価者の倫理観やコミットメントに依拠するしかなく、研究評価の公正性が立脚する基盤は極めて脆弱である。特に、研究者が競争的な環境に置かれるほど、公正な研究評価の基盤は脆くなる。

以上のような意図的な問題に加えて、非意図的な問題も構造的に存在する。顕著な問題として、評価者のバイアスがある。例えば、査読において医学研究において統計的に有意な研究結果が評価されやすいといった問題（Atkinson, Furlong and Wampold, 1982）や、人事評価において性別や民族によって評価の程度が著しく異なるという問題がある（Fang et al., 2000；Richter et al., 2020）。

仮に評価者にバイアスが一切なかったとしても、評価者の評価能力には限界がある。評価のための情報を完備して判断の結果を事前に予見できるよ

うな理想的な評価者と比べて、現実の評価者は常に情報の不確実性に直面し、専門の研究領域は限られているという限定合理性(March and Simon, 1993)に縛られているからである。Crane (1967)は、雑誌編集者が自身の受けた研究流派や選好する手法や理論に応じて採択する研究に傾向があることを指摘した。そもそも研究活動が学術領域の先端を開拓する営みである以上、評価者は研究者本人よりも当該研究について常に情報量に欠けるという構造的な問題が存在する。

なお、こうした公正の問題は定性的評価にのみ関係すると思われることも多いが、定量的評価でも同様に問題となる。定量的評価により、評価者のバイアス、評価能力、恣意性などの要素を排除しようとしたとしても、いかなる定量的指標も構造的に有利不利が生じる余地が存在する。そのため、ある評価においてどの指標を採用するかという局面で、結局評価者や評価設計者側のバイアスや能力、恣意性が問われる。定量的指標により評価しているから公正だという主張は、脆弱な仮定に立脚している。

2.4. 小括

以上では、評価すべき価値基準を明確にし（価値）、適切に測定する手段を選択し（手段）、計画通りに評価を実施する（公正）必要があることを前提に、その三点を実現する上での構造的な困難性をそれぞれ確認した。

まず価値については医学コミュニティ内部あるいは外部との間での価値基準のばらつきが存在すること、および評価する価値についての合意形成の困難性が予想された。手段については、定性的評価の不透明性が定量的評価の客観性への期待をもたらしたものの、定量的評価にもまた限界が指

摘されていることを確認した。また、価値についての合意形成の不在は手段によっては解決できないことが確認された。公正については、評価者の意図的・非意図的なバイアスや評価者の評価能力の限界が指摘され、研究評価には一定の不正さが不可避免的に付随してしまう構造にあることを確認した。

こうした困難性を乗り越えて自律的評価を実現するためには医学コミュニティ内部における合意形成が必要である。そこで、現実にはどの程度合意形成に成功しているのか、あるいはどのように研究評価の困難性が温存されているのかについて調査する必要がある。

3. 方法

医学研究者が医学研究評価のあり方をどのように認識しているのかについて検証するため、本研究は医学研究者を対象としたWeb式アンケート調査の自由回答の結果を用いる。本研究は Minoura et al. (2024)で行われたアンケート調査の副研究であり、調査全体の詳細についてはそちらを参照されたい。

アンケート調査は、第31回日本医学会総会のU40委員会若手研究者有志が中心となり実施した。質問票の作成にあたり、日本医学会分科会に所属する研究者計22名（シニアおよび初期キャリア研究者、学生）を対象として質問票へのコメントおよびフォーカスグループインタビューを実施した。併せて、質問票の検討のため、研究評価に関する非医学系有識者10名以上から意見を聴取した。完成した質問票を用いて、医学研究者を対象にインターネット調査を実施した。本調査は、日本医学会を通じて各分科会（各学会）に配布を依頼した。調査期間は2022年12月14日から2023

年1月17日までとし、無記名での回答とした。倫理的配慮として、回答内容は個人を識別できないよう加工した上で研究に利用する旨を事前に示し、同意を得た。なお、本研究について国立国際医療研究センター（現・国立健康危機管理研究機構）倫理委員会にて審査・承認を得た（承認番号NCGM-S-004530-01）。

質問票は全体で37の大問で構成されており、枝問を含めると77項目あった。分析対象とした設問は、「問7. 医学研究者、研究の評価について、現在行われている好事例でご存知のものがありませんでしたら教えてください。医学研究者、研究の新たな評価方法（定量的、定性的を含む）について、思いつくところがありましたら教えてください。また、本調査に対するご意見ご感想などがあれば、自由に記述してください。」である。

なお、Minoura et al. (2024)では同設問に対する自由回答の量的な分析を中心に行っているが、本研究では同じデータを用いて自由回答の内容について質的に分析を行った。理論的整理に対応した回答を抽出してカテゴリ化を行った。

4. 結果

アンケート調査全体として3169名から回答が得られたうち、自由回答項目の記載があったのは645名（20.4%）であった。そのうち本研究では、理論的整理に対応した回答98件を抽出してカテゴリ化を行った（表1）。対象となった98件の回答は、価値、手段、公正の各観点それぞれに表2から表4に示した。以下では観点ごとに回答の内容を検討する。

表1：理論的整理と実証的検討の対応関係

整理の観点	理論的整理から示唆される課題	自由回答で観察された研究評価における課題
価値をめぐる 困難性 (表2)	<ul style="list-style-type: none"> ■ 医学コミュニティ内部の合意形成 ■ 医学コミュニティを超えた合意形成 	<ul style="list-style-type: none"> ● 研究活動において重視される要素のばらつき (7件) ● 教育、臨床、社会活動などの研究活動外の活動に対する重要性のばらつき (22件) ● ライフコースやジェンダーに係る要素をどのように位置づけるかについてのばらつき (12件)
手段をめぐる 困難性 (表3)	<ul style="list-style-type: none"> ■ 定量的指標は限界を抱えているという前提への合意が難しい ■ 定性的な因子は客観的に判定可能であるという前提への合意が難しい 	<ul style="list-style-type: none"> ● 定量的指標の妥当性や有効性に関する理解のばらつき (10件) ● 定量的指標の中でも (和文論文数など) 重視されないものもある (6件) ● 定量的指標が評価制度に組み込まれている (4件) ● 定量的指標の解釈の前提としての分野慣行のばらつき (12件) ● 定性的な評価は主観的であるという感覚と、定量的指標の方がまだ良いという諦観 (10件)
公正をめぐる 困難性 (表4)	<ul style="list-style-type: none"> ■ 評価の透明性には限界がある ■ 評価者の評価能力には限界がある ■ 評価者にはバイアスが存在する 	<ul style="list-style-type: none"> ● 評価者個人の専門性や経験への依存 (8件) ● 先端的研究の評価不可能性 (2件) ● 恣意的な評価の可能性 (5件)

4.1. 価値をめぐる課題

表2に示した通り、自由回答を通じて何を評価する価値として含めるべきかについてばらつきが存在することが示唆された。以下では、研究活動、研究外活動、ライフコースやジェンダーの順に確認する。

まず、研究活動とはいえ様々な要素が重視されていることが分かった（表2：研究活動において重視される要素のばらつき）。評価すべきとされた要素として「臨床に役立つ」研究や「政策や施策に影響を与える研究」に言及する回答が見られた。他方で、「学問は、いつ、どのように役に立つのか分かりませんので、それぞれの研究者が、ご自身の興味あるテーマについて、追求できることが大切」という意見も存在した。「楽しさ・わくわく感」といった要素への言及もあった。研究にあたってのリーダーシップや人間性を重視するという回答もあった。

次に、研究外活動の重要性に対する態度のばらつき（表2：教育、臨床、社会活動などの研究活動外の活動に対する重要性のばらつき）である。教育、臨床、社会活動、査読、事務作業などが含まれていた。概して、これらの研究外活動により研究活動に割くエフォートの量が減少してしまう点を問題視する意見が中心であった。「臨床にどれだけ時間をとられているかで研究に使える時間が違う」といったように、研究外活動へ割く労力に格差がある点が問題視されており、その格差は「ますます研究に割く時間がなくなること」で拡大するという指摘があった。こうした状況の構造的な原因として「現在、大学教員は（特に国立では）予算面から人員が削減されてきており、臨床・教育・研究などの従来業務の中で、事務仕事も圧倒的に増加」したという指摘もみられた。

教育、臨床、社会活動については「研究には向かないが教育を得意とする人、実践を得意とする人が、それなりに評価されることが必要」といったように研究とは別の価値を認めるべきとする意見もあった。医局員である大学院生による「在学中にどれだけ非常勤医員として地域医療に貢献したか、救急夜間医療について貢献したか、社会活動に貢献したかなどについて（中略）評価していただきたいです。でないと、研究だけしてCOVID-19への対応も含めて何もしない『研究だけしている人』だけが評価され、（中略）研究時間が取れない人が何も評価されません」との回答は、医療提供体制を支える医局システムの評価体系に再考を迫るものである。

三つ目は、ライフコースやジェンダーの問題である（表2：ライフコースやジェンダーに係る要素をどのように位置づけるかについてのばらつき）。自由回答では子育て、家事、介護などへの言及があった。ここでも研究活動に割くエフォートの量が減少してしまう点を問題視する意見が多く、研究外活動と同列で考えることもできる。一方で、こうした労働の多くが女性により担われるという現実と合わせて考えると、ジェンダーに関連した独立した論点だと捉える必要がある。注目すべきことに本調査では男性からもこのような意見が出ており、例えば「研究に専念したいが、子育て・家事にも時間がとられ、自分の理想通りには進まない。（中略）独身研究者や配偶者がパートや専業主婦（夫）の研究者を羨ましく思う」との回答があった。逆に「女性医師が不当に高く評価されているように感じる」のように、ライフコースやジェンダーに関する既存の手当の仕方が望ましくないという回答も存在した。

表2：価値をめぐる課題に言及する自由回答

価値をめぐる課題	自由回答
研究活動において重視される要素のばらつき	<ul style="list-style-type: none"> ● 私は A 診療科医です。私は「臨床に役立つこと」を常に念頭に置いてテーマを選択し、臨床研究論文を書いています。臨床家が研究をすることは、研究者が研究をするのと同じくらい大切なことと自負して論文を書いています。やはり「評価」に関して多くの疑問があります。1つは、雑誌の IF に関してです。我々の業界で格式が高く、ある程度汎用性がある（NEJM や Nature とかではないという意味です）雑誌の IF はおしなべて低く、どんなに頑張っても IF そのものはなかなか上がっていきません。Citation の頻度から考えて、A 診療科というマイナー分野である以上当然なわけですが、それをもって「評価」とされてしまうと頑張りは認められません。 ● 政策や施策に影響を与える研究を行った数 査読数と査読の質 科学的な妥当性を自分で判断できる力をもつ研究者の割合；把握が難しいですが、その国際比較 ● 医学研究（者）の評価は英文誌+インパクトファクターであることが多いように思います。一方で、医学研究、特に社会医学は横断的・応用的な側面があるため、投稿先のジャーナルが人文・社会科学の分野になったり、和文誌になったり、といったケースがあり得るように思います。そのため、評価の基準となるものを見出しにくい状況にあるのではないのでしょうか。 ● 評価を重視すること自体が、学問活動に相容れないと思います。学問は、いつ、どのように役に立つのか分かりませんので、それぞれの研究者が、ご自身の興味あるテーマについて、追求できることが大切だと考えています。 ● 「楽しさ・わくわく感」も役に立つかどうかと同じくらいの重要性があると思います。若い世代（中高生・大学生）が直感的に「すごい」と思える研究の価値を高く評価して欲しい。 ● 必要な人材は組織、タイミングにより異なる、ので、評価法もその都度異なる、というコンセンサスが必要 ● 業績の定量評価も研究者評価のある程度の指標になるとは思いますが、その業績に研究者本人がどの程度コミットしてきたか(まわりの方にどれくらい依存してきたか、リーダーとしてチームを束ねてきたのかチームの歯車に過ぎなかったのか)、人脈や人望はどうなのか、弟子は育てているのか、などなど、他にも重要な要素が多々あるように感じています。 ● 周囲からの信頼を多角的に評価できる指標が求められていると思います ● 所謂 Negative data の発信も価値あるものとして積極的に推奨する雑誌もあります。このような取り組みが bias の問題を解決する一助にもなると思います。
教育、臨床、社会活動などの研究活動外の活動に対す	<ul style="list-style-type: none"> ● 日本に臨床と研究を両立できる physicians scientists の position が確立されることを切に願います。日本の臨床医が行う研究は完全にやりがいの搾取です。研究活動が正当に評価され、それに対する労働対価が支払われることが重要

<p>る重要性のばらつき</p>	<p>です。</p> <ul style="list-style-type: none">● 海外で、PhD を取得しました。現在、市中病院につとめています。日本に帰ってくると、科学研究費もとれない、大学病院への研究する場所の提供のアクセスもない状況です。もっと、研究がオープンにできるようにして欲しい。● 医学部附属病院勤務にしています。学生実習のエフォート、学生講義の評価、臨床における手術件数や新規術式の導入、学会の委員、ガイドライン作成などの数など、無収入、無評価でおわるものに対する適切な評価。● 大学病院では臨床に取り組みながら研究をするが、臨床にどれだけ時間をとられているかで研究に使える時間が違う。一方で、臨床に応用できない研究は、いくらインパクトファクターが高い論文に通ったり、受賞歴があっても意味がない。これらの観点を包括的に評価できる指標、もしくは指標を作れる人はいるのだろうか。● 現在、大学教員は（特に国立では）予算面から人員が削減されてきており、臨床・教育・研究などの従来業務の中で、事務仕事も圧倒的に増加してきており、研究生活を続けたいと思える環境ではなくなってきております。まずは大学などの研究機関に予算面でも人的資産でも余裕を持たせるような体制を国が主体となって作っていただかなければ、日本から世界に発信できるような研究成果はますます少なくなっていくと感じております。● 以前は医育機関が行っていた医学教育を、市中病院が担うようにシフトしてきたおかげでこちらの負担が多く、研究に割く時間が制限される。特に臨床研修病院で教育に対する評価を加味してほしい。● 研究を行うには、時間が必要です。大学（看護 助産学）教員として勤務していますが、実習・授業で研究の時間が持てません。● 研究機関の研究者としての評価と、教育機関での教員としての立場での評価は異なるべきと思います。● 共同研究や博士課程教育などを通して、若い研究者をどれだけ育成したかも、研究者の評価として重要と考えます。● 提案とはならないが、研究は純粋に研究として評価すべきなので、例えば教育に関する実績など、他の能力についての評価は加味すべきではない。● 外科医にとって、トランスレーショナルリサーチは必要と考えます。研究、教育を行っている大学人をもっと正当に評価してもらってもいいと思います。● このような調査をありがとうございます。とても人柄もよく、教育業務にも熱心な教員にばかり教育業務や委員会活動が依頼され、その人はますます研究に割く時間がなくなり、教育業務を他者に押し付けて自身の研究に労力を割いている教員が高い評価を受けること、時間を与えずにあの人は研究は得意じゃないからと教育に忙殺されている人を低く評価する研究者が現実にい
------------------	--

て、悲しい気持ちになります。教育業務や臨床・育成・庶務業務にも正当な評価が必要だと思います。

- 研究には向かないが教育を得意とする人、実践を得意とする人が、それなりに評価されることが必要。
- 研究者としての評価として、特に教育や臨床を含むべきではない。それらはいくまで教育者や臨床家としての評価であり、論文などの研究業績と同程度に比肩しうるものとはなり得ない（例えば臨床を 100%の effort で行なっている医師が論文を書いたものと同様に研究者として評価されるというのはあまりにおかしい。それらは人材や教員としての評価の対象とはなるがやはり研究者としての評価にはなり得ない）。
- 研究者の評価が教育者を選定するための評価になっていると思うので、調査によって現状を明らかにし、評価指標の適切な利用が進むことを望んでいる。
- 研究者は研究内容で評価されるべきだと思います。教育が怠惰の隠れ蓑になっていることを多数経験しているため、「私は教育に熱心です。」と言っている人が信用出来ません。
- 看護系の研究者です。看護系では英語論文数や引用数に対する評価をもっと厳しくする必要があると思います。大学の教授で h-index が 5 もなかったりするのとは問題だと思います。教育が評価されないと皆さん言いますが、研究も評価されていないので業界全体で研究が進んでいないのが実情だと思います。
- 私は英文・和文ともに非常に多くの査読依頼があり、これに非常に時間を割いているが全く評価の対象にならない。投稿を多くするのに全く査読を受諾をしない医師もいるのは不平等を感じる。
- 査読をテーマとした研究もしていただきたいです。論文による科学の進歩には必須ですが、査読の労力は正当に評価されていません。
- 研究者としての評価でなく大学院生としての評価に関する点なのですが、例えば在学中にどれだけ非常勤医員として地域医療に貢献したか、救急夜間医療について貢献したか、社会活動に貢献したかなどについて卒業時の要件でも何でもいいので評価いただきたいです。でないと、研究だけして COVID-19 への対応も含めて何もしない「研究だけしている人」だけが評価され、「研究だけしている人」が研究だけしていればよいように、その他の業務に従事させられているせいで、研究時間が取れない人が何も評価されません。何でもいいので評価してほしいです。
- 医学部生のカリキュラムが大幅に変わってきており、教員の教育へのエフォートが必要とされるのに、教育やその他の業務が転職などの際に評価されにくい。評価されにくいので、教育や管理業務にエフォートをさく教員が減っている負のスパイラルであると思う。評価そのものを見直し、医学そのもの

	<p>が発展し、いい人材が育つような仕組みを作って欲しい。</p> <ul style="list-style-type: none"> ● 公衆衛生学は社会活動（実践）も研究と同等と評価すべきと考える。
<p>ライフコースやジェンダーに係る要素をどのように位置づけるかについてのばらつき</p>	<ul style="list-style-type: none"> ● ダイバーシティに関することを学問として認める気運を作ってほしい。臨床と基礎に大きく分かれているが、医学に関する学問はもっと多くあるということの認知を進めてほしい。 ● 本調査には、最近の行き過ぎた若手優遇策、女性優遇策に対する問題意識を問う項目がない。過度の優遇作により、人材評価が後回しになり、文科省からのノルマ達成のため（大学レベルで）能力いかに関わらず若手・女性人材を登用する風潮がある。海外における Equality 問題解決の表面的な部分だけを「ものまね」しているだけになっているため、日本の研究能力は下降していくと思われる。 ● ロスジェネ世代のため昇進もなく、人生設計が破綻している。このような社会を改善してほしい。 ● リーダーを選ぶ際の良い指標が構築されることを期待しています。女性の割合が低い原因の一つには定性評価が影響していると思います。 ● 研究分野においても男女間の差が生じていると感じています。女性研究者にとっても研究しやすい環境改善を期待しております。 ● 研究に専念したいが、子育て・家事にも時間がとられ、自分の理想通りには進まない。それを睡眠不足で補う方法もあるが、健全でないように思う。独身研究者や配偶者がパートや専業主婦（夫）の研究者を羨ましく思う。 ● 今後、女性研究者の妊娠出産に関してのデータも出していただけたらと思います。妊娠出産育児で大きくビハインドを感じてしまっています。 ● 最近、増えてきた、女性限定採用の公募は、男性やマイノリティーに対する差別であるので、やめたほうがいいと思う。 ● 女性医師が不当に高く評価されているように感じる。17時に帰るのと、遅くまで残った仕事をするのと、同等の評価では納得できない。また各種学会にも女性専用のポジショニングがあるのは、逆に男卑女尊の流れになっていると思うこともある。 ● 健康状態と研究生活の状況については、家族の出産・育児タスクの発生によって生活様式が変化した影響を受けているため、参考値として頂ければと思います ● 子育てをしながら研究を行っている研究者をプラスで評価する仕組みがあればよいと思います。現職場では子育て中は昇進はないと上司に言われており、子育て中は研究にさける時間が減るため日本ではネガティブに受け取られていますが、子育てによって得られる危機管理能力、柔軟性、責任感、教育に関する知見は、研究を行う上でも大きなプラスになると信じています。評価者が中年男性に偏っているのも大きな問題だと考えます。 ● 臨床業務が長く、子育て世代になってから研究活動を開始したため、時間や

	給与等の面でこのまま続けていくか悩みが尽きないです。省察する良い機会となりました。ありがとうございました。
--	---

- ※ 回答には誤字・脱字や文章の誤りが含まれていたが、筆者らは改変をせず、そのまま記載した。ただし、個人や対象を特定できると思われる箇所については、可能な限り趣旨を変えない形で修正を施した。

4.2. 手段をめぐる課題

自由回答では IF への批判的な意見のみならず、一定の有用性を主張する回答が散見された。まず IF の限界を指摘する意見としては、オンライン雑誌のように戦略的に操作可能である点が懸念されていた。さらに、「臨床系に比べ、社会医学系のインパクトファクターが小さいので、数値で単純比較することに不公平感がある」のように、研究分野間での差についての言及が見られた（表3：定量的指標の妥当性や有効性に関する理解のばらつき）。

他方で IF が有効な指標であるという意見として、「内容の評価が大事であると思うが、論文の重要性と緻密さと、インパクトファクターはおおよそ比例していると思われる」⁸ や定量的評価全体についての意見であるが「透明性のある評価基準と評価結果を出すべき」といった点が挙げられた。また、IF などの定量的指標を、研究活動が極端に少ない研究者の「足切りに使う」ことを提案する回答もみられた（表3：定性的な評価は主観的であるという感覚と、定量的指標の方がまだ良いという諦観）。

こうした IF の諦観的受容の背景には代替的な評価の難しさがあるかもしれない。「インパクトファクターや citation index の問題点は、自分が研究をはじめた 20 年以上前から指摘されていたが、現在も当時と同じ問題点が延々と議論されていて進歩がない。研究者を公平・公正に評価することは極めて困難であり、2 週(ママ)くらい回って、結局のところインパクトファクターのような客観的・定量的指標による評価が最も費用対効果がよいという感覚に達しています」という回答があっ

た。また定性的評価について「評価者の主観が入り込みやすい」などの言及があった（表3：定性的な評価は主観的であるという感覚と、定量的指標の方がまだ良いという諦観）。

IF の重要性を規定する別の原因として、評価における指標として用いられていることが指摘された（表3：定量的指標が評価制度に組み込まれている）。「人事評価のため毎年提出する書類には、論文の IF（中略）などで加算があるということになっています」、「インパクトファクターが評価に関わるので、インパクトファクターを見ながら論文を投稿します」や「若い時には、インパクトファクターを考慮しないと就職に不利になるため、また、言い訳だと言い割れる(ママ)可能性もあるため、インパクトファクターのついている学術誌以外には投稿しませんでした」という回答があった。

他方で、定量的指標の限界と密接に結びついた形で言及されていたのが、人的・資金的リソースの多寡やオーサーシップについての扱いが、研究チームごとに異なる点であった（表3：定量的指標の解釈の前提としての分野慣行のばらつき）。

「現在、評価の高い雑誌に論文を掲載するために求められるデータの量は、一人の研究者で出来る範囲を超えている」ため「同じデータを出すので、資金のある研究室では外注する。一方で、資金に乏しい研究室の研究者は長時間かけてすべての実験を自分で実施している」というようにリソースの多寡が影響するという議論や、オーサーシップについて「有名大学出身者あるい(ママ)大きな研究室出身者は論文の共著者になる機会も多く、それだけインパクトファクターも多くなる」

⁸ この点については上記 2.2 を参照。

ケースや、「論文には研究室員の名前を全て載せる」のようなギフトオーサーシップに至っている事例について言及があった。

なお、被引用数と h-index については全体的に

否定的な意見が少なく、特に上記 2.2 において指摘された指標としての限界について言及する回答は無かった。

表 3：手段をめぐる課題に言及する自由回答

手段をめぐる課題	自由回答
定量的指標の妥当性や有効性に関する理解のばらつき	<ul style="list-style-type: none"> ● このアンケートは研究の質が定量化できることを前提に作られているので、違和感を感じる。本来 1 個 1 個の論文の質、目的が違うことを考えると、定量化することはできないのではないかと。同じ impact factor でも reviewer の質で通りやすさが違うので、1 本の論文を定量化して評価できないのだから、鼻から無理だと思う。そういう意味で、first author, second author, last author, correspondence author に絞って評価が行われべきと思われる。 ● 私のみならず海外の共同研究者もインパクトファクターが評価に関わるので、インパクトファクターを見ながら論文を投稿しますが、オンラインジャーナルの乱立などでインパクトファクターの浮き沈みが激しく、投稿先に頭を悩ませています。本調査を基に何らかの提言がなされると嬉しいです。 ● 年間平均インパクトファクターなどを使って給与などに反映するべきだと思う。 ● h-index はよりよい指標だと思います。 ● 社会における研究者の役割が多様化しているのではないのでしょうか。そうなりますと、評価軸も多様化する必要があると思われまます。IF の高い論文を載せるということに過剰に局所最適化していないのでしょうか。それは何に対するインパクトなのでしょう。それは出版ビジネスに研究費をこれ以上注ぐ目的はなんのでしょうか。 ● 研究評価は発表時点で受けることも重要であるが、10 年後 20 年後に人類にとって生産的であったと評価されることもあるのでむづかしいのではないかと。この点で科学的興味を想起される研究対象は重要と思われる。 ● 私自身、雑誌のインパクトファクターは研究のレベルの指標にしますが、近年はインパクトファクターと論文の質に乖離が大きい論文が多いように感じます。また、研究（ラボ）が評価されているのか、研究者が評価されているのか曖昧な部分なども感じるが多々あります。色々曖昧に感じる部分が多い部分ですので、このようなあり方を調査いただき嬉しく思います。ぜひ、調査結果を元に、多くの方が納得いく評価方法のご提案をお願い申し上げます。 ● 研究者を「選択」「選別」するための指標として（雇用時、あるいは契約延長時）、論文数のみでは「短期間で結果が出る研究」を多く行っている研究

	<p>者に有利となるが、本当に求められている研究は（アカゲザルを20年育てたときのカロリー制限が老化に及ぼす影響、という有名な研究のように）論文数が多い研究者以外からも出ていると思われる。このような新奇性、発展性のある研究に対して高い評価点を付与するシステムが望ましい。新奇性・発展性がある研究は、各分野の雑誌よりは、分野をまたいだ雑誌（Nature, Science, ..）に多いのかもしれない。（※各分野の雑誌は、現在「雨後のタケノコ」のように乱立しており、論文は増えてしまっており、かつお互いが引用し合っているので、IF もそれに引っ張られて上昇してしまう。）このような新奇性（誰も思いつかなかった）・発展性（別分野に応用できる；例、循環器→感染症、医学→地学）という採点基準が欲しい。</p> <ul style="list-style-type: none"> ● 発表論文数は必須条件です。英文原著論文が発表できた上で新たな評価基準を考えてもいいと思います。科研費の申請に業績欄をなくしておきながら、日本の科学論文数が少なくなったとされることに違和感を感じます。まず論文発表、それなくして別の評価法は無意味だとすら感じます。 ● 論文本数や IF、被引用数、h-index や g-index が硬直した数字争いになるのは好ましくないが、広島大学のように足切りに使うのはそう悪くない話のように思えます。少なくとも常勤の教員で論文を全く書かないような研究者をフィルタリングする役割としては良いと思う。しかし絶対的な業績評価として使うのは下策であると思う。
<p>定量的指標の中でも（和文論文数など）重視されないものもある</p>	<ul style="list-style-type: none"> ● 現状ではとにかくインパクトファクターのついているジャーナル（英文誌）に出すことで、自分自身の評価に繋がるように工夫しています。しかしテーマによっては、和文誌のほうが高度な議論を行うことができる場合があります。ここで生じるのは、「業績上の評価を得るために、研究の品質を意図的に下げる」というトレードオフです。つまり、研究の評価が、研究の品質と必ずしも一致しない状況が進行しているように思われ、そこに一抹の不安を覚えています。 ● 和文論文がもっと評価されるべきである ● 英論文だけを良しとせず和文であっても適切に評価される必要がある。 ● 質の高い論文が英文誌に投稿される傾向がみられるため、日本で行われた研究を日本の社会で役に立てるために使いにくいのではないかと感じています。和文誌に掲載された論文を評価する手法か価値基準を確立することが必要ではと思います（引用数や言語の違う同じ論文投稿を二重投稿とみとめない価値観など）。和文誌の編集担当をすることが多いのですが、和文誌への投稿論文の質と量が低下してきていることも気になっています。 ● 日本国内の研究を、和文で発表した際の評価機関をつくる必要があると考えます。分野・領域によっては、国内の雑誌に和文で発表することの方が、影響が大きいと思います。 ● 専門医などの申請には研究業績とみなしても、医学研究者としての業績対象

	<p>からと和文を外す。</p>
<p>定量的指標が評価制度に組み込まれている</p>	<ul style="list-style-type: none"> ● A 診療科という狭い分野において、人事評価のため毎年提出する書類には、論文の IF や Q1-3 に当たるかどうかなどで加算があるということになってはいますが、一度もその加算対象となる論文を公表できたことはなく、当該分野の最も IF の高い雑誌もせいぜい IF3-5 であるため、10 以上、15 以上、30 以上、、、というような加算項目を見るたびにため息しか出ません。これらの評価指標は特定の分野のみ配慮した指標で、やはり、臨床がメインのコアな分野の医師に対する救いがなければ、同じ土俵では働いていけません。早く辞めたいのに、近年の集約化のためやめる先がなく、仕方なく大学に踏みとどまっている、というのが現状です。 ● 日本（特に臨床）は、研究評価に関する意識改革が大きく遅れていると思う。たとえば、履歴書にインパクトファクターや H index を記入させるような形式を未だに要求されることがあるが、これはその履歴を審査する側が、研究の価値を評価する能力に欠けていることを自白するに等しい行為である。 ● 若い時には、インパクトファクターを考慮しないと就職に不利になるため、また、言い訳だと言われ割れる可能性もあるため、インパクトファクターのついている学術誌以外には投稿しませんでした。しかしながら、研究者の少ない分野では、インパクトファクターがつかないことも多いため、あくまで、参考程度の指標として用いるべきだと思います。最近では、意図的にインパクトファクターを知らべることなく（インパクトファクターを無視して）、論文を投稿しています。 ● 偏った評価指標で順位をつけて、給与を 5～7% 減らす施策をとっている大学がある（当方の大学）。非常にナンセンスであり、人のやる気を損なう、経営学的にはやってはならないシステムと思う。評価するのは良いが、その使い方は慎重にすべきである。
<p>定量的指標の解釈の前提としての分野慣行のばらつき</p>	<ul style="list-style-type: none"> ● 臨床系に比べ、社会医学系のインパクトファクターが小さいので、数値で単純比較することに不公平感がある。何らかの形で、研究分野間で公平に比較できる評価方法の創出が必要と感じる。 ● 共著論文で荒稼ぎしているのを排除できる評価方法が望ましい。ただの環境ガチャを当てただけにしか見えない。 ● 日本の共著制度にも疑問があります。日本は論文に関わってなくてもその科に所属していれば共著になります。ですから、筆頭とセカンドオーサーとラストオーサーくらいしか事実上の実績として認めてはもらえません（臨床の世界だけ・・・?）。これだと、論文を書いた人ばかりが苦勞をしますし、また何もしなくても共著になるわけなので、人の研究に協力しようという気持ちがなくなります。互いは身内であると同時にライバル関係であるこ

とも事実なので、「実績にならない人の研究」を手伝う気持ちなんてなくなります。（本当は共著に入るので厳密には実績ですが、前述したように日本の臨床ではセカンドオーサーとファイナルオーサーくらいしか「実績」とされないの。）そうなるとそれぞれがそれぞれに研究をするだけで、それらが有機的に結びついたりはしません。海外では教授ではなく直属のスーパーバイザーがラストオーサーのこともありますし、共著は本当にかかわった人だけしか載りません。いつも日本が研究の世界においていられないか心配しています。

- 若い研究者を評価するのは難しい。所属する大学、研究室により研究環境の違いが大きすぎる。スタートラインの違いが研究者としての能力・才能の評価を歪めている可能性があるのではないか。現在、評価の高い雑誌に論文を掲載するために求められるデータの量は、一人の研究者で出来る範囲を超えていることが多く、若手研究や基盤研究 C 程度の研究費では全く足りない。このような現状を改善しない限り、研究者の正当な評価はできないのではないかと日頃思っている。
- 規模の大きい大学では、業績の水増しが横行している。例えば、論文には研究室員の名前を全て載せるなど、不正行為に含まれるギフトオーサーである。言いようによってはボーダーに行くこともでき、業績水増しの逃げ道となっているが、このあたりを適切に評価できる指標があればうれしい。同じデータを出すのでも、資金のある研究室では外注する。一方で、資金に乏しい研究室の研究者は長時間かけてすべての実験を自分で実施している。資金のある研究室の研究者は、実験手法の説明すらできないこともあり、この両者が同じ評価というのは不適切なので、新たな指標で解決してほしい。
- 有名大学出身者あるいは大きな研究室出身者は論文の共著者になる機会も多く、それだけインパクトファクターも多くなる。CI や HI-index もしかり。上記のような研究者が教授などになってもしっかりと後輩の研究者を育成する能力がない事例に度々遭遇してきた。
- 調査の目的から逸れるのかもしれませんが、研究者の評価にあたっては、任期と予算の問題が極めて大きいと思います。有為な人材であっても、任期が来れば雇い止めとなること、無期転換は予算上無理なことが多いためです。このことは研究者の福祉だけでなく、研究界の発展を著しく妨げる大きな要因となっています。任期が来ても、本人と研究機関の合意で延長できる法改正が必要だと思います。
- 医局員の多い教室の研究者の業績が高くなりやすい。この点を踏まえた評価が難しい。
- 評価基準が明確に示されていない状態で業務に取り組んでいる研究者（臨床医）が多いことが問題かと思います。臨床医の研究活動については、First author や corresponding author を中心に評価するのではなく、研究チーム単位

	<p>での評価を重要視することが必要かと思います</p> <ul style="list-style-type: none"> ● 研究者業績のアウトカムを、社会への還元（雇用を生み出す、製品開発と流通、起業など）と科学への貢献（論文数、被引用数、講義数、担当学生数）に分けて、併せて100%となる評価基準の作成と、定期的な繰り返しの評価の実施はいかがでしょうか。 ● 研究の環境も評価する（実験室、オフィス、事務補佐員の有無などで研究関連業務の負担が異なるので業績にも影響するため）。自己申告評価の重視（職場の各委員会に入っていないなくてもその資料をつくるなど実質的に委員と同等のことをしているケースもある、ゴーストライトも普通にある）。 ● 1.自分が筆頭著者として載っている論文のみを h-index のように数値化、2.自分が責任著者として載っている論文のみを h-index のように数値化したものをそれぞれ出すこと。つまり共著者として載っている論文は h-index から外す。大きな研究室の出身者や、研究室の性格により皆載せている場合などは、h-index が非常に高くなってしまう。これを防ぐ方法が必要である。
<p>定性的な評価は主観的であるという感覚と、定量的指標の方がまだ良いという諦観</p>	<ul style="list-style-type: none"> ● 定性的評価は総論賛成だが、評価者の主観が入り込みやすいので、双方向チェック機能が働くような形を望む。 ● インパクトファクターや citation index の問題点は、自分が研究をはじめた20年以上前から指摘されていたが、現在も当時と同じ問題点が延々と議論されていて進歩がない。研究者を公平・公正に評価することは極めて困難であり、2週くらい回って、結局のところインパクトファクターのような客観的・定量的指標による評価が最も費用対効果がよいという感覚に達していません。 ● 現在の定量的評価に限界や問題はあるが、ある程度しょうがないと考える。ただし学閥や個人的繋がりによる評価は問題があり、透明性のある評価基準と評価結果を出すべき ● 研究の評価は論文の内容でおこなうより客観的なものはない。内容の評価が大事であると思うが、論文の重要性と緻密さと、インパクトファクターはおおよそ比例していると思われる。インパクトファクターによる研究の評価は、実はそれほど問題ではないのではない。また、引用数を軸とした評価系とあわせることで、基礎的でインパクトファクターが高い雑誌に掲載されにくいが高評価の研究をすくい上げて、内容で評価するほかない。一つの指標はある側面しか評価できないのは当然であるので、複数の指標を組み合わせるようなガイドラインがあると良いと思う。また、よく第一著者と責任著者により重み付けがなされるが、責任著者による恣意的な重み付けがあり得るため重要視されるべきでないと思う。 ● 定性的評価とされている要素も、極力定量化の評価にシフトさせることが重要と考える。研究者であれば、研究資金の獲得が最重要項目の一つに挙げられるべきで、かつ自身の給与の一部も研究資金から支出可能な体制の整備が

	<p>必要と考える。米国 NIH のグラント獲得形式、透明性の高い評価基準に見習う部分は多くあろうと考える。</p> <ul style="list-style-type: none">● どの論文に掲載されたかで評価するのもある程度やむを得ないと思う● 研究に対する評価が、量的評価に偏っているという指摘は妥当であると思う。所属する施設では研究に対する”質的評価”の方法が決まっていない。● 定性評価は必要と思うが、方法論が見当たらないので適正なツールを教えてください。● 客観的指標については今のもので十分と思います。● 研究評価に関して、問題があるとはいえインパクトファクターはいい判断材料になっていると思う。逆に和文の論文？を業績に含めるという内向的な評価査定は全くの論外。いわゆるハゲタカジャーナルと何が違うのか？
--	--

※ 回答には誤字・脱字や文章の誤りが含まれていたが、筆者らは改変をせず、そのまま記載した。
ただし、個人や対象を特定できると思われる箇所については、可能な限り趣旨を変えない形で修正を施した。

4.3. 公正をめぐる課題

公正をめぐる課題については、研究評価が評価者の認知的制約に限界づけられていることが理解されていた（表4：評価者個人の専門性や経験への依存）。特に先端的な研究については評価者の理解可能な範囲を超えることがあるとの指摘がなされた（表4：先端的な研究の評価不可能性）。年長の研究者が研究の先進性を理解できないことを指摘する点は、Minoura et al. (2024)において年齢

が高いほど定性的評価のアンコンシャスバイアスが問題視されない傾向が見られたのと整合的であった。さらに、恣意的な評価の不正については、仲間内での引き立てや逆に出身研究室への忖度への言及があった（表4：恣意的な評価の可能性）。「お友達」や「好き嫌い」への言及もあり、そうした内容が真実かどうかは検証が必要であるものの、そのような回答が生じる程度には、既存の評価制度の公正性に対する信頼が低いことが分かる

表4：公正をめぐる課題に言及する自由回答

公正をめぐる課題	自由回答
<p>評価者個人の専門性や経験への依存</p>	<ul style="list-style-type: none"> ● 過去に実績があっても、現在の研究についていけない年長の研究者では、研究の先進性が理解できず、十分な評価ができていない可能性がある。研究評価自体を若い研究者が新しい視点で評価を行う。 ● 評価者の評価が不可欠 評価結果の FEED BACK が必要 ● 大学等の評価に影響を与えるものであることを祈願しています。特に reviewer の仕事の評価が必要と感じます。 ● 評価は何らかの定規を使って行われるものです。定量的なもの以外にも、個人の経験に照らしてという評価もあります。評価者が知っている内容については高評価になり、理解できないと低評価になります。研究費の分配は、研究で成功した人が差配することが多いと思いますが、評価は成功体験に影響されます。教科書に書かれていることに合致していないと、評価されません。ネイチャーの宣伝に「あなたの知っていることは何も書いてありません」というのがありましたが、公的な立場の人が評価すると「知っていることしか評価できない」になり、「有用」とは教科書に書いてあることだけになってしまいます。 ● 科研費は「話題のテーマ」や「仲間内での引き立て」による採択が生じやすく、本来行うべき領域や内容に適切に配分されない傾向があります。 ● 日本には、無意識のバイアス・差別が多くあると思う。研究費審査、採用審査などの様々な局面で、性別、現在のポジション、年齢等を隠したうえで審査すべきである。 ● 評価後に評価の基準を明らかにした評価書類の作成を義務付ける。閉鎖的で均質な選考委員（50代以上の男性が主、それ以外はいても意見が言いにくい環境）にならない方策が必要かと考えています。インパクトファクター偏重はご自身がそれで教授になった成功例と考えておられるからと拝察していません。

	<ul style="list-style-type: none"> ● 勝者たちによる偏った評価、は是正されるべきと思いますが...難しい問題と思います。
先端的研究の評価 不可能性	<ul style="list-style-type: none"> ● 審査委員の65%は専門外審査と知り愕然とした。良い研究程その分野に明るいものしか評価できない。 ● 医学系研究者では理解できない全く独創的な研究をすると、reviewerが理解できないため、論文がrejectされる。独創的な研究はreviewerがrejectしてもeditor権限で掲載するというようなシステムが必要である。
恣意的な評価の可能性	<ul style="list-style-type: none"> ● 有意義な試みと思うが、高齢の研究者に定着した評価の概念は永久に払拭できないと思う ● 結局教授の好き嫌いでしょう ● 財布を握る人の好きな研究（お友達）につけるという制度が気に食わないが、それが現実でしょう。 ● 日本では基礎/臨床系の学会とも総じて「お友達クラブ」化しており、「クラブ」から外れた研究者は国内で全く評価されません。まずはこの状況の見直しが必要と思われます。 ● 大学の教授の選考を、自分たち大学で行わせると、身内の人間関係のしがらみで、明らかに劣る研究者が教授二なり、その人間たちが徒党を組んで、公正に能力のある人間を選ぶことを拒むので、日本の大学ランキングは上がりません。そのためには、1) 内部昇格を禁止する制度、2) 大学内部の互選で管理職を選ぶ制度の排除=研究をする側と研究者を評価して大学の運営をする側は全く別の組織にして公正な研究評価をし、また、大学運営側が自分の言う事を聞く人間を昇進させたり、聞かない人間の昇進をストップさせるようなことが出来ない公正な形で選ばれるようにすることなど、大学内の権力構造によって教授人事が選ばれるような制度を排除しないことには、どんなに良い評価法が出来ても、その評価法を利用して自分たちの都合で動かすので意味がありません。大学の研究側と運営側を完全に分離しない限りには新たな評価法は意味がないと思いますし、逆に、人事のシステムが良ければ現行の評価法でも十分良い人材が生き残り、不当な人材をなくすことが出来ると思います。

※ 回答には誤字・脱字や文章の誤りが含まれていたが、筆者らは改変をせず、そのまま記載した。
ただし、個人や対象を特定できると思われる箇所については、可能な限り趣旨を変えない形で修正を施した。

5. 考察

自律的に研究評価を行うためには、評価すべき価値基準を明確にし（価値）、適切に測定する手段を選択し（手段）、計画通りに評価を実施す

る（公正）必要があるが、その実現のためには医学コミュニティ内部における合意形成が必要である。にもかかわらず、アンケート調査の自由回答の分析からは医学研究評価に対する医学研究者た

ちの問題認識のばらつきが観察された。研究者が重視する価値については、研究活動や研究外活動からライフコースやジェンダーまでにわたる、幅広いばらつきが存在した。手段については、IFなどの定量的指標への賛否が分かれたが、純粋に評価手法としての適否よりも人的・資金的リソースやオーサーシップ慣行などの研究を取り囲む環境と結びついて議論されていた。公正については、評価者の限界やバイアスへの言及があり、その結果生じる恣意的な評価の可能性が懸念されていた。こうした結果から、医学コミュニティ内部における合意形成は課題を抱えており、自律的な医学研究評価が難しくなっていることが示唆された。

特に、手段に関する自由回答では、定量的評価の限界を理解しつつも、定性的評価に比べて相対的に信頼できる評価手段であるという諦観的な受容が見られた。他方で価値に関する自由回答では、学問的追究や独創性などの指標で測定するのが難しい要因や、研究外活動やライフコースといった指標外の要因を加味すべきであると考えられていた。公正に関する自由回答では、評価者の認知的な限界や恣意性が介入する可能性のために、公正な研究評価を行うことは難しいと考えられていた。このことから、医学研究者は現在の研究評価のあり方が特定の価値に偏重していると認識しており、評価の前提となる価値体系を変えることなく定性的評価に移行したとしても一層不公正な評価を受ける恐れがあると感じていた可能性がある。このジレンマの中で、諦めを伴いつつ定量的評価を受け入れるという指標の諦観的受容が発生していたと考えられる。このように考えると、医学研究者が定量的評価へ傾倒する原因は評価の手段自体の問題というよりも、むしろ価値や公正に関わる問題として捉え直すことができる。これは、評価が

機能する前提としてステークホルダー間の合意形成が必要だという政策評価の議論と整合的である(田辺, 2005; 西出, 2020)。

医学研究者による自律的な研究評価の意義が高まっているにもかかわらず、本研究では自律的な医学研究評価は困難な状況にあることが観察された。むしろ研究評価を定量的指標に依存することにはある種当然の帰結であることを示した。しかしそれは現状のままであることを良しとするものには無い。こうした構造的環境にあるからこそ、医学コミュニティ自身による研究評価のガバナンスが求められている。一つには研究評価における価値体系を適切に反映するようなプロセスを制度化することが求められるだろう。医学コミュニティとして、どのような研究や研究者が求められているのかを議論する場を設定することが必要である。もう一つには公正な研究評価を行うための十分なリソースと透明性を備えた評価制度を構築することが求められている。審査プロセスの透明性の確保や審査基準の公開は当然であるが、評価疲れと言われるような人的リソースの限界にも注意する必要がある。

現状の研究評価に対して再考を制度化する組織の例として、専門家集団によって構成される学会が想定される。例えば、研究評価に関する委員会などの設置が考えられるだろう。しかし学会による管理にも当然限界がある。まず一般論として医学コミュニティ内部でのガバナンスを行う上では、研究評価が既存の序列を再生産する契機を含むことに十分注意が必要である。単なる現状追認はガバナンスとしての意義に欠ける。加えて、個別の学会で扱う課題とアンブレラ組織で扱う課題は異なる。学会において研究の価値を一元的に序列化することは学会内政治を反映する危険性や多様性

を失うおそれがある一方、総花的な羅列をしてしまうと価値基準としての意義が薄くなってしまふことに注意が必要である。また、価値のみならず手段や公正を含めてモニタリングするなど、広範で継続的な点検作業が望まれる。

本稿の限界として、今回扱ったデータが任意のアンケート調査における自由回答項目を通じて得られた点が挙げられる。アンケート調査全体として3169名から回答が得られたうち、自由回答項目の記載は20.4%に当たる645名にとどまった。したがって回答は必ずしも網羅的ではないうえ、解釈に当たっては社会的望ましきバイアス等の限界を抱えている。その意味で本研究の分析は、少なくとも上記で示したようなコンフリクトが生じていることを間接的に推認させるものにとどまる。そのため、ここで言及されたことが直ちに解消すべき課題であることを意味せず、逆にここで言及されなかった課題は重要でないことを意味しないことに注意が必要である。本研究が探索的に示した関係について、今後より厳密な因果の検証が求められる。また、直接関係しないために本稿には掲載しなかったが、自由回答の中には本調査自体に対する批判的なコメントも多かった（本調査に対するコメントのうち91件が肯定的だったが77件が否定的だった）（Minoura et al., 2024）。

また、本研究は医学研究評価再考の必要性を示したにとどまっているため、個別の点について具体的にどのような評価が望ましいのかについては各医学コミュニティの判断に委ねられる。再考の結果、既存の評価が正しいと判断されることもあるだろう。しかし、形骸的な検討や問題提起が妨げられるような状況では再考したとは言えない。研究評価について実質的な検討を行う環境整備が求められる。

本稿の分析では医学研究評価ガバナンスの必要性が示されたが、一方で現実の医学コミュニティは分断を抱えていることも理解しなければならない。回答の中には医学研究評価のあるべき姿を真摯に論じるものも多かった一方で、「結局教授の好き嫌いでしょ」のような研究評価への頹廃的な信頼喪失も垣間見られた。また繰り返しになるが、「問題があるとはいえインパクトファクターはいい判断材料になっていると思う」、「客観的指標については今のもので十分と思います」といった既存の研究評価慣行の健全性を肯定しようとする態度も存在する。そして、本アンケート調査の自己選択バイアスにも関わるが、そもそも研究評価について思いを巡らせる必要のない有利な立場にいることに非自覚的な研究者もいるだろう。こうした分断を超えて、医学コミュニティ内部での対話がまず求められている。

6. 結論 今後の研究評価再考に向けた政策的示唆

研究評価を医学コミュニティ内部で自律的に管理することは重要であるにもかかわらず、研究評価は価値・手段・公正において構造的な困難性を抱えており、調査からは現実に軋轢を生んでいることが示唆された。研究評価に構造的な困難が存在するということを自覚すると、既存の研究評価の慣行の健全性を盲信する態度こそが研究評価ガバナンス不全の原因となる。したがって、本稿で確認した医学研究評価の構造的困難を前提にするならば、医学コミュニティは研究評価を不断に再考する必要がある。

価値や公正に関する課題がもたらす不確実性は、医学コミュニティ内部のガバナンスによってある程度まで縮減することができる。そうすることで医学研究評価に対して信頼が与えられ、定性的な

評価が選択される素地が生まれる。問題点を指摘するだけでは定量的指標の諦観的受容という現象に対する解決策にはならず、医学研究評価を取り囲む環境を医学コミュニティが協力して整える体制が望まれる。

謝辞

第31回日本医学会総会の役員の方、学術委員会U40委員会委員の方々に感謝を申し上げます。加えてアンケート調査の周知にご協力頂きました日本医学会分科会の先生方に感謝を申し上げます。日本学士院若手アカデミーおよび日本学術会議の会員の方々には、有意義なご助言を賜りました。本研究におけるインタビュー調査およびアンケート調査にご回答いただいた皆様に、感謝を申し上げます。また、今井健二郎先生、井花庸子先生、湧井-木村晶子先生には、本研究を長きにわたりサポートして頂きました。東尚弘先生、福田吉治先生、城山英明先生には、専門的見地から多くの助言をいただきました。最後に、高瀬堅吉先生、小泉周先生、林和弘先生には、研究期間を通じてたいへんお世話になりました。

本研究は公益財団法人 日立財団 2021 年度（第53回）倉田奨励金 人文・社会科学研究分野の支援を受けて行われました。

参考文献リスト

〈英語文献〉

- Atkinson, D. R., Furlong, M. J., & Wampold, B. E. (1982). Statistical significance, reviewer evaluations, and the scientific process: Is there a (statistically) significant relationship? *Journal of Counseling Psychology*, 29(2), 189–194. <https://doi.org/10.1037/0022-0167.29.2.189>
- Chubin D. E., & Hackett E. J. (1990). *Peerless Science. Peer Review and U.S. Science Policy*. New York: SUNY Press.
- Crane, D. (1963). The Gatekeepers of Science: Some Factors Affecting the Selection of Articles for Scientific Journals. *The American Sociologist*, 2(4), 195–201.
- Elsevier. (2019). *Research Metrics Guidebook*. https://elsevier.widen.net/s/chpzk57rqk/acad_rl_elsevierresearch-metricsbook_web
- Fang D, Moy E, Colburn L, et al. (2000). Racial and ethnic disparities in faculty promotion in academic medicine. *JAMA*, 284(9), 1085-92. doi:10.1001/jama.284.9.1085
- Garfield, E. (1979). Is citation analysis a legitimate evaluation tool?. *Scientometrics*, 1, 359–375. doi: 10.1007/BF02019306
- Hicks, D., Wouters, P., Waltman, L. et al. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520, 429–431. doi: 10.1038/520429a
- Hirsch JE. (2005). An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA*, 102, 16569-16572.
- Al-Jabi SW. (2023). Global research trends and mapping knowledge structure of depression in dialysis patients. *World J Psychiatry*, 13(8), 593-606. doi: 10.5498/wjp.v13.i8.593
- March J., Simon H. (1993). *Organizations* (second edition). New York: John Wiley & Sons. 高橋伸夫訳 (2014) 『オーガニゼーションズ: 現代組織論の原典』ダイヤモンド社
- Minoura A, Shimada Y, Kuwahara K, et al. (2024). Medical researchers' perceptions regarding research evaluation: a web-based survey in Japan. *BMJ Open*, 14, e079269. doi: 10.1136/bmjopen-2023-079269

- Miura, T., Asatani, K., & Sakata, I. (2021). Classifying Sleeping Beauties and Princes Using Citation Rarity. In: Benito, R.M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L.M., Sales-Pardo, M. (eds) *Complex Networks & Their Applications IX. COMPLEX NETWORKS 2020*. Studies in Computational Intelligence, vol 943. Springer, Cham. https://doi.org/10.1007/978-3-030-65347-7_26
- Polanyi, M. (1962). *The Republic of Science: Its Political and Economic Theory, Knowing and Being: Essays by Michael Polanyi*, ed. by M. Green. Chicago: University of Chicago Press. 佐野安仁他監訳 (1985) 「科学の共和国—その政治的・経済的理論—」, 『知と存在』晃洋書房
- Richter KP, Clark L, Wick JA, et al. (2020). Women Physicians and Promotion in Academic Medicine. *N Engl J Med*, 383(22), 2148-2157. doi: 10.1056/NEJMsa1916935. PMID: 33252871
- Zuckerman, H., & Merton, R.K. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva*, 9, 66–100. doi: 10.1007/BF01553188
- 〈日本語文献〉
- 芦部信喜 (高橋和之補訂) (2019) 『憲法 第七版』岩波書店
- 天野郁夫 (2006) 「国立大学の法人化—現状と課題—」 『名古屋高等教育研究』 6 : 147-69.
- 天野郁夫 (2008) 『国立大学・法人化の行方—自立と格差のはざままで』 東信堂
- 有本建男, 佐藤靖, 松尾敬子, 吉川弘之 (2016) 『科学的助言 21 世紀の科学技術と政策形成』 東京大学出版会
- 江間有沙 (2013) 「科学知の品質管理としてのピアレビューの課題と展望：レビュー」 『科学技術社会論研究』 10 : 29-40.
- 岡村浩一郎 (2020) 「「大学の研究力低下」をめぐる議論—論点と政策—」 国立国会図書館調査及び立法考査局 『ポスト 2020 の科学技術イノベーション政策』 93-108.
- 小野寺夏生 (2003) 「引用文献数で研究を評価できるか？」 『学士館会報』 803 : 50-57.
- キャロル・H・ワイス (2014) 『入門評価学 政策・プログラム研究の方法』 日本評論社
- 小林信一 (2012) 「公的研究開発の現状と課題—本調査研究のねらい—」 『調査報告書 国による研究開発の推進』 5-19.
- 小林信一 (2015) 「大学改革と研究費—運営費交付金と競争的研究費の一体的改革をめぐる—」 『レファレンス』 775 : 1-30.
- 標葉隆馬, 林隆之 (2013) 「研究開発評価の現在—評価の制度化・多元化・階層構造化—」 『科学技術社会論研究』 10 : 52-68.
- 標葉隆馬 (2020) 『責任ある科学技術ガバナンス概論』 ナカニシヤ出版
- 調麻佐志 (2013) 「科学計量学と評価」 『科学技術社会論研究』 10:16-28
- 白川展之 (2021) 「多様な研究評価の方法論と科学技術政策の評価」 『日本評価研究』 21 (1) : 5-16
- 城山英明 (2018) 『科学技術と政治』 ミネルヴァ書房
- 孫媛 (2017) 「研究評価のための指標：その現状と展望」 『情報の科学と技術』 67:179-184.
- 竹内健太 (2019) 「国立大学法人運営費交付金の行方—「評価に基づく配分」をめぐる—」 『立法と調査』 413 : 67-76.
- 田辺國昭 (2005) 「中央省庁における政策評価の現状と課題：期待と失望のスパイラルを超えて」 『政策形成支援のための政策評価：NIRA 型

- 政策評価モデルの提言（NIRA 研究報告書）』、41-80.
- 西出順郎（2020）『政策はなぜ検証できないのか 政策評価制度の研究』勁草書房
- 西山慶司（2019）『行政サービスの外部化と「独立行政法人」制度』晃洋書房
- マックス・ウェーバー，尾高邦雄訳(1936)『職業としての学問』岩波文庫
- 原現吉（1982）『科学研究費—その成り立ちと変遷』科学新聞社
- 平澤冷（2004）「我が国の公共部門における研究開発評価の課題（公的資金による研究開発の評価）」、『研究技術計画』、17（3）：128-141.
- 福成洋（2014）「研究戦略のための計量書誌学の実践的活用と応用」『情報管理』57（6）：376-386.
- 藤垣裕子（2003）『専門知と公共性』東京大学出版会
- 藤垣裕子，平川秀幸，富沢宏之，調麻佐志，林隆之，牧野淳一郎（2004）『研究評価・科学論のための科学計量学入門』丸善株式会社
- 藤垣裕子（2018）『科学者の社会的責任』岩波書店
- 藤田宙靖（2001）『行政組織法〔新版〕』良書普及会
- 松澤孝明（2013）「わが国における研究公正 公開情報に基づくマクロ分析（1）」『情報管理』56(3)：156-165.
- 山本隆司（2007）「学問と法」城山英明＝西川洋一編 法の再構築[Ⅲ]科学技術の発展と法(東京大学出版会)143-167頁
- リチャード・R・ネルソン，後藤晃訳（2012）『月とゲッター 科学技術と公共政策』慶応義塾大学出版会
- ロバート・K・マートン，成定薫訳（1983）『科学社会学の歩み エピソードで綴る回想録』サイエンス社
- 〈報告書など〉
- 大学改革支援・学位授与機構（2022）「大学への民間的発想・手法導入の有効性の検証と自律的活用のための指針の開発」『大学改革支援・学位授与機構研究成果および平成27～29年度科学研究費補助金研究成果報告書』
- 国立大学法人評価委員会（2004）「国立大学法人の中期目標・中期計画（素案）についての意見」
- 「国の研究開発全般に共通する評価の実施方法の在り方についての大綱的指針」（平成9年7月28日内閣総理大臣決定）(https://www.mext.go.jp/b_menu/shingi/kagaku/hyoka/index.htm) 2023年6月26日アクセス
- 総合科学技術・イノベーション会議 有識者議員（2015）「第5期科学技術基本計画における指標及び目標値について」(<https://www8.cao.go.jp/cstp/kihonkeikaku/5sanko.pdf>) 2023年6月26日アクセス
- 「第5期科学技術基本計画」（平成28年1月22日閣議決定）(<https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>) 2023年6月26日アクセス
- 「国の研究開発評価に関する大綱的指針」（平成28年12月21日内閣総理大臣決定）(<https://www8.cao.go.jp/cstp/kenkyu/taikou201612.pdf>) 2023年6月26日アクセス
- 国立大学法人評価委員会（2004）「国立大学法人の中期目標・中期計画（素案）についての意見」(https://www.mext.go.jp/b_menu/shingi/kokuritu/sonota/1223085.htm) 2023年6月1日アクセス
- 文部科学省（2023）「成果を中心とする実績状況に基づく配分について」(https://www.mext.go.jp/content/20230329-mxt_hojinka-100014170.pdf) 2023年6月1日アクセス
- 国立大学等の独立行政法人化に関する調査検討会議

- (2002)「新しい「国立大学法人」像について」(<https://www8.cao.go.jp/cstp/siryo/haihu16/siryo2-2.pdf>) 2023年6月1日アクセス
- 大学基準協会 (2023)「大学評価基準ハンドブック (2023 (令和 5) 年改訂)」(<https://www.juaa.or.jp/accreditation/institution/handbook/>) 2023年6月1日アクセス
- DORA (2012) “San Francisco Declaration on Research Assessment.” (<https://sfdora.org/read/>) Accessed June 1, 2023.
- 日本医学雑誌編集者会議 (2022)「医学雑誌編集ガイドライン 2022」https://jams.med.or.jp/guideline/jamje_2022.pdf) 2023年6月26日アクセス
- 日本学術会議 (2007)「対外報告 提言：知の統一社会のための科学に向けてー」(<https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-20-t34-2.pdf>) 2025年3月18日アクセス
- 日本学術会議 (2008)「対外報告 我が国における研究評価の現状とその在り方について」(<https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-20-t51-3.pdf>) 2023年6月26日アクセス
- 日本学術会議 (2021)「提言 学術の振興に寄与する研究評価を目指してー望ましい研究評価に向けた課題と展望ー」(<https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-25-t312-1.pdf>) 2023年6月26日アクセス
- The Centers for Disease Control and Prevention (2020) “CLEARANCE OF SCIENTIFIC INFORMATION PRODUCTS DISSEMINATED OUTSIDE OF CDC FOR PUBLIC USE” (<https://www.cdc.gov/os/policies/docs/CDC-GA-2005-06.pdf>) 2025年2月18日アクセス

Received 12 September 2025

Accepted 2 March 2026