

Title	研究生産性の把握：研究者属性と生産性の対比
Author(s)	寺田, 好秀; 藤田, 裕二; 七丈, 直弘
Citation	年次学術大会講演要旨集, 37: 182-187
Issue Date	2022-10-29
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/18694
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

研究生産性の把握 ～研究者属性と生産性の対比

○寺田好秀（政研大）^{*}，藤田裕二（政研大），七丈直弘（政研大／一橋大）
^{*}yo-terada@grips.ac.jp

1. はじめに

研究者は研究活動を通じて様々なアウトプットを生み出していく。また、科学研究には多くの不確実性が伴い、不連続的変化を誘発する発見を見出す研究者もいれば、漸進的でありながらも多くの研究の積み上げによって社会に大きなインパクトを与える研究者もいる。また、研究アウトプットの学術的インパクトも研究者によって大きく異なっている。これら個々の研究者（研究分野においてはチームサイエンスであって個々の研究者の創意に基づくものだけではないが、ここでは簡便に主体を示す語として「研究者」を用いている）の活動状況の相違は、置かれた環境や特性に起因すると考えられるが、それら属性情報は個人情報であり、広く把握することは困難であった。しかし、科学技術・イノベーションに関する施策を検討する上で、研究活動によって生じたアウトプットの把握は、過去の施策の有効性の評価や将来の施策の立案において重要な役割を持っており、広くその把握を行い分析することによって得られる示唆は有用性が高い。研究アウトプットの中でも論文の執筆とその公刊は多くの研究分野において大きなウェイトを有していることから、本研究では書誌情報データベースを用いて把握された研究資料の公刊状況から研究アウトプットを把握し、内閣府が保有するデータベースによって個々の研究者の属性データ情報をそれに付与することで、研究者属性と研究アウトプットの間を可視化することができるデータ分析ツール（以下、可視化分析ツール）を作成した¹。本研究は内閣府科学技術・イノベーション推進事務局によるエビデンスに基づく政策立案（以下、EBPM）に関連した委託事業として政策研究大学院大学が受託して実施した成果の一部である。

2. 先行研究

科学者の研究生産性についての研究は R. K. Merton, H. Zuckerman, P. Bourdieu らによる科学社会学の先駆的研究を端緒として、多くの研究が行われてきた。研究者の知的創造活動を知識生産関数として表現し、過去の知的ストックが将来の研究生産に影響を与えると考える手法 [1]や、近年では質問票調査によって回答者の属性を把握し、その生産性の変化を分析した研究 [2]や博士論文データベースを用いて博士取得年を把握し、博士号取得後の研究従事年数と研究生産性間の比較を行った研究 [3]などが存在している。どの事例においても、把握可能となった研究者の数は社会における研究者全体の数パーセントかあるいはそれよりも大幅に少なり小サンプルサイズであり、それによる把握可能であった

¹ こうした試みの結果は、『統合イノベーション戦略 2022』（令和 4 年 6 月 3 日閣議決定）や『第 6 期科学技術・イノベーション基本計画』（令和 3 年 3 月 26 日閣議決定）で示された内閣府のエビデンスシステム（e-CSTI）（2022 年 8 月 20 日現在） [11] [12] [13] で報告されているが、本稿はそれらに学術的な意義や位置づけなどを付与してまとめるとともに、分析内容を前進させたものである。例えば、商用書誌情報データベースに J-STAGE から抽出した日本から発表された書誌情報を結合し、研究生産性の指標としてトップ 10%論文を追加し、さらに論文を出版していない研究者も集計の対象としている。そのことから、本稿の可視化の結果は、これまでの報告 [11] [12] [13] には含まれないものである。

り言及可能な事象は極めて限定的である。

3. データ

内閣府が有する府省共通研究開発システム(e-Rad)の研究者属性データと書誌情報データベースを名寄せ²作業により接続し、これにより作成されたデータ(以下、名寄せデータ)により、日本全体の研究者の研究生産性と属性の関係について俯瞰的な分析を行った。

2008年1月から運用が開始されたe-Radは、各府省等が所管する競争的研究費制度に応募する際に使用され、研究者の情報などが登録されている。しかし、e-Radが対象としているのは、競争的資金のすべてとプロジェクト研究資金と呼ばれる研究資金の一部[4]であることに留意が必要である。本研究では、2008年以降にe-Radに情報が登録された研究者が分析の対象となる。

e-Radの各研究者の情報として、生年月日、所属機関名、所属機関への着任日や離任日、任期の有無、常勤または非常勤の別等の研究者属性が収録されている³。今回は2020年10月時点のデータを利用することとし、総件数は1,079,535件、研究者IDの数は758,087件、所属機関数は25,989件となった⁴。なお、以下では、e-Radに収録された個人を「研究者」、各書誌情報データベース上で識別されている個人を「著者」と呼称する。

各研究者の研究生産性を算出するために、商用書誌データベースを利用するが、主要な書誌情報データベースでは英語以外の文献の収録が少ないという指摘がある。日本語文献を補うため、JSTが運営する論文電子出版プラットフォームであるJ-STAGEが有する書誌情報を併用することとした。具体的には、Scopusの2022年4月時点のデータにJ-STAGEの書誌情報を結合したデータベース(文献情報をマージしただけでなく、ScopusとJ-STAGE間の引用関係も同定されている。以下、Scopus+JSTAGEと称する)を作成することで、分析の対象に日本語での研究活動の拡充を試みた⁵。

Scopus以外にもDimensionsやWeb of Scienceといった商用書誌データベースが存在し、収録されている文献の対象や情報の種類に差異がある。例えば、ScopusにはField-Weighted Citation Impact(以下、FWCI)という「1論文あたりの被引用数を世界平均(出版年・分野・文献タイプ別に算出)で割った数値」とする独自の指標を持ち、さらにField-Weighted Outputs in Top Citation Percentiles(以下、FWCI Percentiles)という「出版年別のFWCIが世界全体の上位X%に含まれる論文」を示す指標などがある[5]。

日本の機関に所属する研究者が著者となっており、かつ、出版年が2008年から2020年までの文献を対象としたScopus+JSTAGE、Dimensions、Web of Scienceの集計値を表1のようにまとめた。Scopus+JSTAGEは、DimensionsやWeb of Scienceといった商用書誌データベースと比較して文献、著者、雑誌、所属機関の数が多い。このことから、内閣府[6]が考える日本語文献が中心の分野についての分析の難さや地方大学の研究活動を過小評価しているという課題に対応し、応用研究や社会実装も日本全体の研究力として分析できるようになった可能性が示唆される。

² 一般的に名寄せは表記ゆれの整理や、同姓同名を生物学的単位に分類することであるが、本稿では内閣府[7]を踏襲し、データを突合することも含めて簡便的に「名寄せ」と称した。

³ 内閣府[7]に詳細がある。

⁴ これまで公表された可視化の結果[11][12][13]より、約2年分のデータが増えている。

⁵ 正確には、Scopusには本文が英語以外(日本語を含む)の論文の書誌情報が含まれている。また、同様にしてJ-STAGEには本文が英語である論文も含まれている。

表 1 主要商用書誌情報データベースの特徴

商用書誌データベース	文献数	著者数	雑誌数	所属機関数
Scopus+JSTAGE	2,955,530	3,064,339	53,208	204,352
Dimensions	775,463	1,731,771	20,590	27,221
Web of Science	728,554	1,070,065	40,328	60,493

出典) Dimensions, Scopus+JSTAGE, Web of Science より筆者作成

e-Rad と各書誌情報データベースのデータを著者単位で結合するため、研究者及び著者の氏名と所属機関の情報で名寄せを行った⁶。

名寄せデータの特徴としては、まず、各研究者が出版した論文（内閣府 [6] [7]を踏襲して以下では、研究アウトプットを表す単位を「論文」と表すが、論文以外の文献も含まれる）の数や出版した論文の被引用数などを研究生産性として利用し、研究者属性や論文属性を活用しながら分析ができる⁷。具体的には、e-Rad には研究者の生年月日の情報があることから、研究者が論文を出版した年齢が算出でき⁸、他にも性別、常勤職か非常勤職か、雇用財源は安定財源か外部資金か、任期の有無、所属する研究機関などの研究者の属性が利用できる。また、各書誌情報データベースからは、出版年、被引用数、著者順、共著者数、国際共著論文かどうか、独自に検討された研究分野や FWCI のような指標を論文の属性として分析に利用できる。また、名寄せデータには、同様の属性を有する多数の研究者が含まれており、それら研究者の出版活動の通時的分析が可能であることから、クロスセクション分析、時系列分析は勿論、パネルデータ分析を行うことができ、EBPM への貢献が期待される。

なお、e-Rad に登録されているものの当該期間に出版された論文が存在しない研究者は、当該年の研究生産性の値を 0 として集計した⁹。これにより、書誌情報データベースに著者情報が含まれない研究者を考慮した分析が可能となり、分析結果の一般性が改善される。

4. 分析手法

前節の検討内容に基づき、研究者の研究生産性と各種属性との関係の可視化を試みた。ただし、本稿では、Scopus+JSTAGE を使った名寄せデータに基づき、いくつかの属性に着目して可視化した。

まず、第一に出版時年齢を、第二に性別を分析の視点として採用した。なお、『統合イノベーション戦略 2022』（令和 4 年 6 月 3 日閣議決定）や『第 6 期科学技術・イノベーション基本計画』（令和 3 年 3 月 26 日閣議決定）で若手研究者や女性研究者の活躍促進を政策目標としている点でも意味のある視点である¹⁰。また、科学社会学や科学の経済学でも研究者の研究生産性は個人に内在するケイパビリティに起因すると考えられ、それを律する要因として Fox [8]は年齢を、Stephan [9]は年齢と性別を指摘してい

⁶ 名寄せ作業の詳細は、本学会同日に報告される [10]。

⁷ 内閣府 [7]は研究生産性の向上のために研究者の属性情報の観点などで、要因や経年を分析可能にすることを目指している。

⁸ e-Rad の課題として生年月日に関する情報の入力間違えが導出される。例えば、推計された出版時年齢が負の値や 200 を超えることがある。このことから、名寄せデータで出版時年齢が 26 歳以上または 69 歳以下と推計される場合を分析の対象とし、それ以外は除く。

⁹ これまで公表された可視化の結果 [11] [12] [13]のデータでは、論文を出版した著者や出版された論文のみを対象としている。なお、ここでの「研究生産性の値が 0」とは、分析対象とした書誌情報データベースに該当著者を著者に含むレコードが無いことを示すのであって、書誌データベースに含まれない研究アウトプットが存在しないことを意味するものではない。

¹⁰ この様な背景から、内閣府 [7]は総合科学技術・イノベーション会議として日本の研究力強化のために研究者の性別や年齢を切り口に要因分析するとしている。

る。

次に、研究生産性の指標としては、各研究者が各年に出版した論文数（以下、論文数）、その被引用数の合計（以下、被引用数）、トップ 10%論文の数（トップ 10%論文数）¹¹を採用する。ここでトップ 10%論文は、FWCI Percentiles が 10%以下を示す論文と定義する。また、各指標はサンプルサイズの時間的な変化¹²を考慮して、クロスセクション平均による集計値の平均を算出するものとする。具体的には、研究者ごとの出版した論文の数、その被引用数の合計の平均値、トップ 10%論文の数を出版年ごとに算出し、さらに 2008 年から 2020 年の 13 年分の平均が集計値となる。

以上の議論で導出された重要な研究者の属性を分析の軸として設定する。具体的には、縦軸を研究者の研究生産性、横軸を研究者の出版時年齢とし、更に、それを性別で分類した可視化を行う¹³。

その際、BI（ビジネスインテリジェンス）ツール Tableau により、日本の研究者の研究生産性を可視化する。Tableau で開発した可視化分析ツールにより、政策立案者を含むユーザーは SQL や Python などの知識がなくとも GUI を通じて探索的分析を行うことが可能になる。¹⁴

5. 可視化の結果

分析の対象は、重複を除いた研究者の数が 414,504、その内、男性の数が 313,195 で女性の数が 101,309 となる。また、重複を除いた論文数が 1,810,878 で、その内、トップ 10%論文の数は 142,300 となる。

縦軸を研究生産性とし、横軸を出版時年齢として可視化した結果が図 1 から図 6 である。具体的には、研究生産性の指標を論文数とした可視化の結果が図 1 と図 2、被引用数とした可視化の結果が図 3 と図 4、トップ 10%論文数とした可視化の結果が図 5 と図 6 であり、図 2、図 4、図 6 は性別により分類を行っている¹⁵。

¹¹ 内閣府 [6] [7]がアウトプットや研究生産性の目標例として、論文数、被引用数、トップ 10%論文などを挙げている。これまで公表された可視化の結果 [11] [12] [13]では、研究生産性の指標として論文数と被引用数が利用されているが、本稿では新たにトップ 10%論文を研究者の研究生産性を表す指標として追加している。

¹² 例えば、内閣府 [7]は「大学において任期付きポストにある若手教員が増えている状況」があると考えている。

¹³ 藤田 [14]は分けて視覚化する全景的俯瞰を分析の基礎だと考え、全景俯瞰の基礎は分析の軸の設定だと主張している。また、分析を簡素かつ強力にするためには、軸の数を縦軸と横軸の 2 つに絞って 2 次元で俯瞰することの必要性を述べている。

¹⁴ 内閣府 [7]は EBPM の推進のために「行政官が簡易にデータを可視化・分析すること」を示している。

¹⁵ 図 1 から図 6 は、データベースの作成方法や研究生産性の指標の計算方法が新しいことから、これまで公表された分析結果 [11] [12] [13]には存在しない。

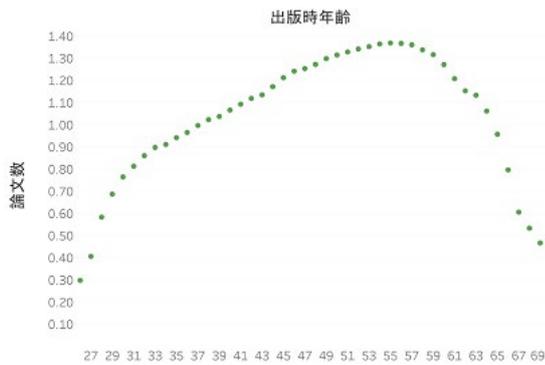


図 1 論文数と出版時年齢の関係

出典) 名寄せデータにより筆者作成

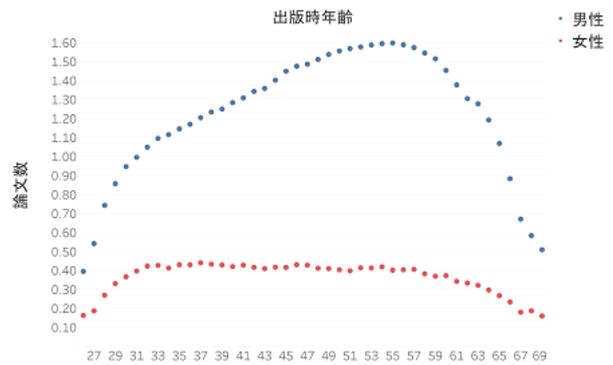


図 2 性別ごとの論文数と出版時年齢の関係

出典) 名寄せデータにより筆者作成

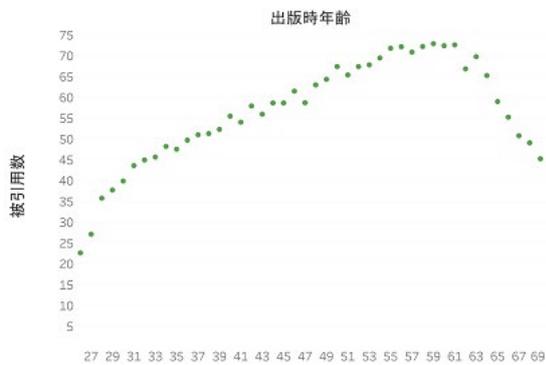


図 3 被引用数と出版時年齢の関係

出典) 名寄せデータにより筆者作成

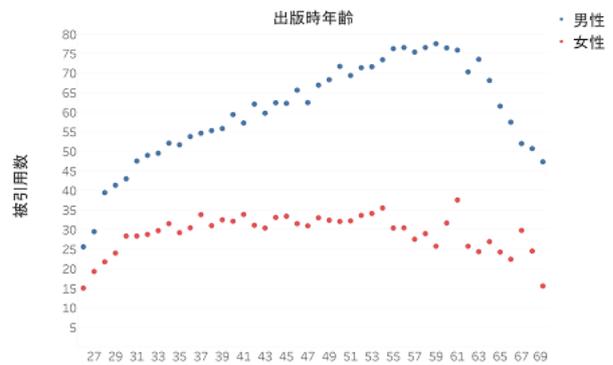


図 4 性別ごとの被引用数と出版時年齢の関係

出典) 名寄せデータにより筆者作成



図 5 トップ10%論文数と出版時年齢の関係

出典) 名寄せデータにより筆者作成

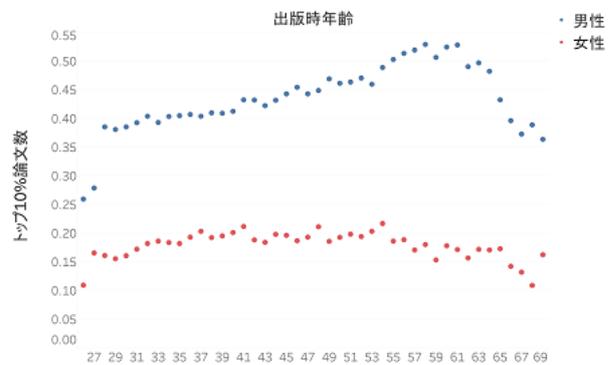


図 6 性別ごとのトップ10%論文数と出版時年齢の関係

出典) 名寄せデータにより筆者作成

6. 結語

本稿では、内閣府が有する e-Rad と書誌情報データベースを名寄せして結合することで、日本社会全体の研究者の姿を表すデータベースを構築し、研究者の研究生産性と属性の関係を俯瞰的に可視化する分析ツールを開発した。研究者や論文の属性情報に着目し、日本の研究生産性の向上を目標とした科学技術・イノベーション政策に資するマクロ分析が可能となる。

本研究は、科学社会学や科学の経済学などの学術分野の蓄積となるとともに、『統合イノベーション戦略 2022』(令和 4 年 6 月 3 日閣議決定) や『第 6 期科学技術・イノベーション基本計画』(令和 3 年

3月26日閣議決定)で示されるエビデンスに基づく戦略策定のための有用な手段となることが期待される。

分数カウントを用いた集計や筆頭著者論文に絞った集計、年齢や性別以外の研究者属性や論文属性を切り口とした分析は、今後の課題としたい。

謝辞

本研究は内閣府令和4年度科学技術基礎調査等委託事業「エビデンスデータベースの利活用高度化に関する調査」の成果の一部である。データの前処理や分析補助を行った加瀬豊氏(政研大)に感謝する。

参考文献

- [1] S. G. Levin and P. E. Stephan, "Research Productivity Over the Life Cycle: Evidence for Academic Scientists," *The American Economic Review*, vol. 81, no. 1, pp. 114-132, 1991.
- [2] C. Lawson, A. Geunabed and U. Finardi, "The funding-productivity-gender nexus in science, a multistage analysis," *Research Policy*, vol. 50, no. 3, pp. 1-20, 2021.
- [3] H. Sauermann and P. Stephan, "Conflicting Logics? A Multidimensional View of Industrial and Academic Science," *Organization Science*, vol. 24, no. 3, pp. 889-909, 2013.
- [4] 府省共通研究開発管理システムポータルサイト, "府省共通研究開発管理システム (e-Rad) について," Available: https://www.e-rad.go.jp/dl_file/particulars_e-rad.pdf. [アクセス日: 2022年8月20日].
- [5] エルゼビア・ジャパン株式会社, "研究力分析ツール SciVal の使い方～ 主な評価指標 ～," 2020年12月. Available: <https://rao.web.nitech.ac.jp/wp-content/uploads/2021/03/0a5fb96373aab2f500971208dc6cb541.pdf>. [アクセス日: 2022年8月20日].
- [6] 内閣府, "EBPM等の推進に係る取組状況について," 2019年3月14日. Available: <https://www8.cao.go.jp/cstp/gaiyo/yusikisha/20190314-1/siryo1.pdf>. [アクセス日: 2022年8月20日].
- [7] 内閣府政策統括官(科学技術・イノベーション担当)付, "研究力の分析に資するデータ標準化の推進に関するガイドライン," 2019年4月5日. Available: https://www8.cao.go.jp/cstp/evidence/guideline_honbun.pdf. [アクセス日: 2022年8月20日].
- [8] M. F. Fox, "Publication Productivity among Scientists: A Critical Review," *Social Studies of Science*, vol. 13, no. 2, pp. 285-305, 1983.
- [9] P. E. Stephan, "The Economics of Science," in *Handbook of the Economics of Innovation*, vol. 1, Edited by Bronwyn Hall and Nathan Rosenberg, Chapter 5, North-Holland, 2010. (後藤康雄訳・解説(2016)『科学の経済学-科学者の「生産性」を決めるものは何か』日本評論社.)
- [10] 藤田裕二・宇佐美徳隆・藤井俊彰・永井博昭, "e-CSTIにおける研究データ連結の確からしさ," 研究・イノベーション学会第37回年次学術大会, forthcoming, 2022.
- [11] 内閣府, "研究アウトプットと研究者属性の関係性分析," Available: https://e-csti.go.jp/downloads/2-kenkyu/rdcap/20210113_e-Rad_rdcap_tab.pdf. [アクセス日: 2022年8月20日].
- [12] 内閣府政策統括官(科学技術・イノベーション)付参事官(エビデンス担当), "女性活躍に係るガラスの天井の見える化の試み-eCSTIを通じた分析-", 2020年12月. Available: https://e-csti.go.jp/downloads/about/women_empowerment_201228.pdf. [アクセス日: 2022年8月30日].
- [13] 内閣府政策統括官(科学技術・イノベーション担当)付, "e-CSTIを通じたEBPM等の推進に係る取組状況について," 2021年1月. Available: https://e-csti.go.jp/downloads/about/EBPM_Initiatives_20210105v2.pdf. [アクセス日: 2022年8月30日].
- [14] 藤田康範, 経済戦略のためのモデル分析入門, 慶應義塾大学出版会, 2011.