

スター・サイエンティストの検出とリスト構築 (2019.12)



早稲田大学ビジネス・ファイナンス研究センター
科学技術とアントレプレナーシップ研究部会

2019年12月20日

スター・サイエンティストの検出とコホート・データセットの構築

牧 兼充 (早稲田大学ビジネススクール准教授)
菅井 内音 (東京工業大学修士課程)
隅藏 康一 (政策研究大学院大学教授)
原 泰史 (一橋大学経済学研究科特任講師)
長根 (齋藤) 裕美 (千葉大学教授)

早稲田大学ビジネス・ファイナンス研究センター
科学技術とアントレプレナーシップ研究部会
ワーキングペーパーシリーズ No. 001

Working Paper



スター・サイエンティストの検出とコホート・データセットの構築

牧 兼充 (早稲田大学ビジネススクール准教授)

菅井 内音 (東京工業大学修士課程)

隅藏 康一 (政策研究大学院大学教授)

原 泰史 (一橋大学経済学研究科特任講師)

長根 (齋藤) 裕美 (千葉大学教授)

要旨

本稿は、JST-RISTEX「科学技術イノベーション政策のための科学」研究開発プログラム「スター・サイエンティストと日本のイノベーション」が構築した、「スター・サイエンティスト・コホート・データセット」の構築手順と活用方法をまとめたものである。具体的には、スター・サイエンティストの検出方法、暫定リストの概要、正式版リストの概要を解説した上で、データセットの活用法について述べる。

「サイエンスの経済学」分野の研究では、カリフォルニア大学ロサンゼルス校のLynne Zucker教授及びMichael Darby教授により「スター・サイエンティスト(Star Scientist)」の存在が指摘されている。スター・サイエンティストとは、卓越した研究業績を残す少数のサイエンティストのことを指し、通常の研究者に比べて、多くの論文を出版し、多くの被引用を集め、特許を数多く出願する。また通常のサイエンティストと比較して優秀な博士課程の学生やポスドクを育成する傾向がある。スター・サイエンティストは、通常の研究者と比べ、ベンチャー企業を設立する傾向にあり、またスター・サイエンティストの関わるベンチャー企業は他のベンチャー企業に比較して、高い業績を生み出している。更に産業界と関わるスター・サイエンティストは、研究業績も上がるという、サイエンスとビジネスの好循環が発生している。こうしたスター・サイエンティストと企業の連携は、米国のみならず、1980年代の日本においても、バイオテクノロジー分野で観察されている。このように、スター・サイエンティストは研究の学術的インパクトをもたらすのみならず、その経済・社会的インパクトとの相乗効果およびそれらが生じるメカニズムを探る上でも重要である。

このスター・サイエンティストに関する現象は、今現在の日本においても発生しているのであろうか。特に1995年以降の日本のイノベーション・システムの改革は、スター・サイエンティストにどのような影響を与えたのであろうか。日本のスター・サイエンティストの現状を分析するために、筆者らはJST-RISTEX 科学技術イノベーション政策のための科学「スター・サイエンティストと日本のイノベーション」を立ち上げた。このプロジェクトの目的は、スター・サイエンティストとその産業へのインパクトを分析し、それら成果を広く公表・実装することで、日本におけるサイエンスとビジネスの好循環を構築することにある。このプロジェクトの活動の柱の一つが、日本のスター・サイエンティストを分析するためのデータセットの構築を行うことである。このデータセットの活用により、これまで政策形成の課題となっていたサイエンティストに係る課題の意思決定をするための「客観的根拠(エビデンス)」を学術的にかつ政策立案プロセスに対して提供することがはじめて可能となる。

キーワード: スター・サイエンティスト、検出、リスト、日本、データセット

1. はじめに

本項では、JST-RISTEX「科学技術イノベーション政策のための科学」研究開発プログラム「スター・サイエンティストと日本のイノベーション」が作成した、「スター・サイエンティスト・コホート・データセット」の構築手順と活用方法をまとめたものである。具体的には、スター・サイエンティストの検出方法、暫定リストの概要、正式版リストの概要を解説した上で、データセットの活用法について述べる。

日本の科学技術イノベーションを促進するためには、介護および福祉に係る予算枠が増大する背景の下、効果的な研究費の配分が今後更に重要となる。しかしながら、科学技術分野ごとにどのような研究者がどの程度存在するのか、そうした研究者にどの程度の研究費を配分すると良いのか、研究成果のみならずエグジット(産業化可能性)を見据えて研究費をどのように配分すると良いのかに関する、科学的エビデンスは依然限られている。このような課題を解決するためには、「スター・サイエンティスト」という観点から、日本のイノベーション・システムを分析することが有益である。

「サイエンスの経済学」分野の研究では、カリフォルニア大学ロサンゼルス校の Lynne Zucker 教授及び Michael Darby 教授により「スター・サイエンティスト(Star Scientist)」の存在が指摘されている。スター・サイエンティストとは、卓越した研究業績を残す少数のサイエンティストのことを指し、通常の研究者に比べて、多くの論文を出版し、多くの被引用を集め、特許を数多く出願および取得する。また通常の研究者と比べると優秀な博士課程の学生やポスドクを育成する傾向がある。更に、スター・サイエンティストは、通常の研究者と比べ、よりベンチャー企業を設立する傾向にあり、またスター・サイエンティストの関わるベンチャー企業は他のベンチャー企業に比較して、高い業績を生み出している。こうしたスター・サイエンティストと企業の連携は、米国のみならず、1980年代の日本においても、バイオテクノロジー分野で観察されている。このように、スター・サイエンティストは研究の学術的インパクトをもたらすのみならず、その経済・社会的インパクトの相乗効果およびメカニズムを探る上でも重要である。前述の Zucker 教授らは論文データや特許データ、および地域の企業のデータなどを結合し、大規模なデータベースを作り、スター・サイエンティストの特性や産業界へのインパクトをいくつかの観点から定量的に明らかにしてきた。

一つ目はスター・サイエンティストがもたらす地理的効果である(Zucker, Darby and Brewer, 1998)。1976年から1989年におけるバイオテクノロジー分野を対象に、遺伝子配列に関する発見をした327人の研究者をスター・サイエンティストとして定義したうえで、彼らとバイオテクノロジーのベンチャー企業の関係についていくつかの観点から分析している。この研究では、スター・サイエンティストとベンチャー企業の地理的分布を分析し、スター・サイエンティストの所在地にベンチャー企業が集積していることを明らかにした。これによりスター・サイエンティストの分布とベンチャー企業が集積には何かしらの相関があることが示唆された。

二点目はベンチャー企業のパフォーマンスにスター・サイエンティストが与える効果である(Zucker, Darby and Armstrong, 2002)。この研究では、ベンチャー企業のパフォーマンス指標として、(1)特許数とその内容、(2)開発中のプロダクト、(3)上市した(製品化された)プロダクトを取り上げた上で、それらと(A)スター・サイエンティスト、(B)全米のトップ研究大学(スター・サイエンティストの所属有無は問わないものとする)、(C)ベンチャー・キャピタルとのつながりを概観した。定量的な解析の結果、スター・サイエンティストと共著論文が多いベンチャー企業はパフォーマンスが高くなることを示した。一方、トップ研究大学との共同研究やベンチャー・キャピタルからの投資は比較的軽微な効果に留まることも示している。

上記で示唆されたのは、スター・サイエンティストのベンチャー企業への関与が、ベンチャー企業のパフォーマンスの向上に影響すると考えられる点である。それでは、スター・サイエンティストのベン

チャー企業への関与は、スター・サイエンティストの研究業績にどのように影響するのだろうか。Zucker and Darby (2007) ではこの点についても検証し、その結果スター・サイエンティストがベンチャー企業に関わることは、研究者の業績を上げることが、定量的な解析の結果より明らかになった。この論文では、ベンチャー企業と関わりを持つバイオテクノロジー分野のスター・サイエンティストの業績の変化を分析している。なおここでの「関わり」とは、(1)ベンチャー企業との共著論文がある、あるいは(2)ベンチャー企業にて役職を有していることを意味する。米国にて、バイオテクノロジー分野のスター・サイエンティストは 207 名存在し、そのうち 69 名が何らかの形で企業との関係を持つ。また 57 名は企業との共著論文を執筆し、12 名は企業においてサイエンティフィック・アドバイザー(科学顧問)もしくはファウンダー(創業者)としてのポジションを有している。また、彼らの論文あたりの被引用数を算出したところ、企業との関わりが全くないスター・サイエンティストは論文数で平均 1.67、被引用数で平均 13.15 であるのに対して、ベンチャー企業と何らかの関わりをもつスター・サイエンティストは論文数で平均 2.53、被引用数で平均 22.52 であり、明確な差があることが明らかとなった。さらに、ベンチャー企業と何らかの関わりをもつスター・サイエンティストのなかで、企業との共著論文のみ執筆する研究者と、そのみならずベンチャー企業において何らかのポジションを有する研究者を比較すると、公刊論文数の平均はそれぞれ 2.54 と 2.53 と大差ないが、被引用数では前者が 20.74 であるのに対し後者が 31.39 と、圧倒的に多い。

このようにベンチャー企業と関わるスター・サイエンティストは論文数(量)および被引用数(質)ともに大きいことはもちろん、特にベンチャー企業において何らかの役職を保持し、より直接的に関わるサイエンティストのほうが研究業績の質が高いことが示唆される。さらには、時系列データによる分析を行い、スター・サイエンティストがベンチャー企業と関わって以降、スター・サイエンティストの研究業績が向上していることも示されている。

これらのことから、Zucker and Darby (2007)は、スター・サイエンティストと企業が何らかの形で関わると、それぞれ研究業績および企業業績が上がるという“Virtuous Circles in Science and Commerce”(「サイエンスとビジネスの好循環」)の関係を示唆した(図1:サイエンスとビジネスの好循環)。スター・サイエンティストと企業が連携することにより、企業はより高い業績を得る。これにより、企業活動の促進は新たな産業の発展につながる。一方で、企業と関わるスター・サイエンティストは、より多くの論文を生産し、かつより質の高い論文を生み出すようになり、科学的ブレークスルーを生み出す可能性が高まる。

これまでに紹介した結果は 1970 年代から 1980 年代の米国、かつバイオテクノロジー分野を対象としている。では、こうしたサイエンスと商業化における好循環は果たして、普遍的に成立し得る事象といえるだろうか。

Zucker and Darby (2001)では、1970 年代から 1980 年代のバイオテクノロジー分野を対象に、国際比較も行った。その結果によると、国別のスター・サイエンティストの分布について、米国が 50.2%で 1 位であるのに続き、日本が 12.6%を占めている。さらに企業とのつながりがあるスター・サイエンティストの割合をみると、米国が 33.3%であるのに対し、日本は 42.3%である。ただし、スター・サイエンティストの企業とのつながりに着目した場合、1980 年代の産学連携先には日米で大きな違いがある(Zucker and Darby, 2001)。スター・サイエンティストが連携する企業として、米国ではその多くが主にベンチャー企業であったのに対して、日本では大企業であった。

このように、Lynne Zucker 氏や Michael Darby 氏を中心に進められてきたスター・サイエンティスト研究は、定性的に解析されてきたイノベーション・クラスターのモデルの在り方を定量的に指し示したことで重要な学術的貢献を果たしてきた。しかしながら、日本を対象にしたスター・サイエンティスト研究についていえば、依然黎明期にある。両教授による研究対象には日本も含まれているが、データ上の制約および興味関心から 1970 年代から 1980 年代を扱っており、また日本のナショナル・イノベシ

ョン・システムの歴史上の転換点として重要な役割を果たしたと考える 1995 年の科学技術基本法以降の現象を検証・考察していない。

今日の日本には、それぞれの科学分野に、どの程度、スター・サイエンティストが存在するのか。また、スター・サイエンティストの存在や研究成果は、ベンチャー企業をはじめとする新事業創造にどの程度結びついているのであろうか。上記の目的を達成するために、筆者らは JST-RISTEX 科学技術イノベーション政策のための科学「スター・サイエンティストと日本のイノベーション」プロジェクトを立ち上げ、またこのプロジェクトを主体として、日本のスター・サイエンティストを分析するためのデータセットの構築を行った。

本稿では、第 2 節においてプロジェクトの概要を説明し、その後第 3 節ではスター・サイエンティストのリスト構築の先行事例についてまとめる。第 4 節では本プロジェクトで策定した「スター・サイエンティスト」のリストの概要について解説し、第 5 節ではコホート・データセットの概要をまとめる。最後に第 6 節では本データセットを活用した今後の研究課題や展望について述べた上で、全体を総括する。

2. プロジェクトの概要

日本のスター・サイエンティストの現状を分析するために、筆者らはJST-RISTEX「科学技術イノベーション政策のための科学」研究開発プログラム「スター・サイエンティストと日本のイノベーション」を立ち上げた¹⁾。このプロジェクトの最終目的は、スター・サイエンティストとその産業へのインパクトを分析し、それら成果を広く公表・実装することで、日本におけるサイエンスとビジネスの好循環を構築することにある。その目的を達成するために、以下のリサーチ・クエスチョン(RQ)について、研究を行っている。これらの RQ は、関連するデータセットの不足・欠如等の理由により、現時点で学術的な先行研究が不足している領域である。

1. 日本におけるスター・サイエンティストの同定手法の開発
2. 日本におけるスター・サイエンティストの現状分析
3. 日本のナショナル・イノベーション・システム改革におけるスター・サイエンティストへの影響
4. スター・サイエンティスト誕生要因の分析と次世代育成手法の検証

本プロジェクトは、スター・サイエンティストに関するデータセットを新たに構築し、スター・サイエンティストの研究開発パフォーマンスを明らかにすること、ベンチャー企業への関与及びそのメカニズムを検証する。これにより、これまで政策形成の課題となっていたサイエンティストに係る課題の意思決定をするための「客観的根拠(エビデンス)」を学術的に政策立案プロセスに対して提供することがはじめて可能となる。

本プロジェクトでは、サイエンティストに係る特許、論文およびファンディングのデータ、およびベンチャー企業の活動に係るデータセットを包括的に収集・突合する。それゆえ、遡及的なデータの活用であり、あくまでオブザベーション・スタディの範疇であり、因果関係の検証の厳密性、すなわち内的妥当性の課題は未だ残る。しかしながら、こうした網羅的なデータセットの構築および整備は、日本のスター・サイエンティストの現状の評価及びメカニズムの探索、更には係る科学技術イノベーション政策の形成において直接的に大きな意義がある。

本プロジェクトは、以下のプロセスにおいて進める。

1. データセットの構築: 論文データベース、特許データベース、ファンド情報データベース、ベンチャー企業データベースを組み合わせることで、スター・サイエンティストを検証するためのデータセットを構築する。
2. インタビューによる RQ の再検証/仮説の導出: 先行研究に基づいたスター・サイエンティストのリストに基づいてインタビューを行い、RQ の再検証を行い、ブラッシュアップを行う。
3. データ分析: 構築したデータセットを活用し、前述の4つの RQ の検証を行う。
4. 研究会の開催: 研究成果の普及及び「サイエンスとビジネスの好循環」を促進するためのコミュニティの形成を目的とした研究会を月に1回程度行う。
5. インタビューによる結果の解釈: 本プロジェクトにおけるデータ分析の結果を再度スター・サイエンティストに提供し、議論の場を持つ。
6. 研究成果のとりまとめ: 本プロジェクトにおける研究成果のとりまとめとして、海外のトップジャーナルへの投稿と、書籍の出版を目指す。本プロジェクトにおける研究成果の還元のために、積極的に政策担当者及び大学発ベンチャー育成担当者を対象としたワークショップなどを

¹ JST-RISTEX 科学技術イノベーション政策のための科学「スター・サイエンティストと日本のイノベーション」: <http://www.jst.go.jp/ristex/stipolicy/project/project27.html>

² プロジェクト・ページ: <http://www.stentre.net/ss/>

スター・サイエンティストの検出とリスト構築 (2019.12)

開催する。

本プロジェクトは、2017年10月に開始し、2020年9月に終了する予定である。

3. スター・サイエンティストのリスト構築の先行事例

「スター・サイエンティスト」という言葉の概念は多義的である。「科学的なブレークスルーを生み出す卓越した研究業績をもつサイエンティスト」といった定義が多くの研究者の間でのコンセンサスであると考えられるが、「ブレークスルー」とは何か、「研究業績」はどのように測るのか、といった統一見解はない。スター・サイエンティスト研究の潮流においては、用語としてスター・サイエンティストをどのように定義するかではなく、定量的にどのように定義するか、というところに主眼が行われている。

表1に、主な先行研究のスター・サイエンティスト抽出方法をまとめた。

現在パブリックに公開されているデータセットの中では、Zucker and Darby (2007)においても用いられている Clarivate Analytics 社の Highly Cited Researchers(以下、CA 版 HCR)が、スター・サイエンティスト研究では活用しやすい。

Clarivate Analytics 社の前身である Thomson Reuters 社は過去に Institute for Scientific Information (ISI)を吸収合併し ISI Highly Cited と呼ばれるデータベースを提供していた。これは同社が提供する Essential Science Indicators (以下、ESI)に従った 21 分野のそれぞれにおいて 250 人の最も高被引用の研究者をリスト化したものであり、最も近年のものは 1981 年から 2008 年までに刊行された論文をカバーしていた。

Clarivate Analytics 社は、2014 年より毎年、同社の論文データベースである Web of Science (以下、WoS) のデータに基づいて、CA 版 HCR のリストを発表している³⁴。2014 年以降の各年の CA 版 HCR のリストは、当該年の 12 年前から 2 年前までの 11 年間を集計期間として、その期間に WoS に収録された論文を対象として、高被引用論文 (Highly Cited Paper ; 以下 CA 版 HCP) を多く刊行している研究者の情報が収録されたものである。CA 版 HCP とは、論文刊行年ごとに、上記の ESI に従った 21 分野それぞれにおいて、被引用数が上位 1%の論文を指す。分野ごとに CA 版 HCP の数で研究者のランキングを作成し、集計期間中に当該分野で論文を刊行した総研究者数の平方根をボーダーとして、CA 版 HCR の選出を行なっている⁵。

なお、この CA 版 HCR のリストを活用して、本プロジェクトでは、いくつかの研究論文をまとめている。例えば代表例として、斎藤・牧(2017)は、日本のスター・サイエンティストの現状についての総合的な分析を行っている。

³ <https://hcr.clarivate.com/> (2019 年 3 月 8 日アクセス)

⁴ 上記のウェブサイト上には 2001 年のものも公表されている。Li (2016)は 2001 年と 2014 年のデータを比較しその間の変化について論じている。

⁵ より詳しくは、同社ウェブサイト上の説明を参照。

<https://hcr.clarivate.com/methodology/purpose-and-methodology/> (2019 年 3 月 8 日アクセス)

4. スター・サイエンティストのリストの構築

本プロジェクトにおけるスター・サイエンティストのリストは以下の5種類がある。

- スター・サイエンティスト・ショートリスト(選定の基準を厳しめにしたもの)
- スター・サイエンティスト・ロングリスト(選定の基準をやや緩めにしたもの)
- クロスフィールド・スター・サイエンティスト・リスト
- スター・サイエンティスト・インテグレートド・リスト
- 機関別高被引用文献数ランキング(時系列)

本項では、それぞれのリストについて、使用データ、抽出手順、検討課題についてまとめる。

使用データ

本プロジェクトにおけるスター・サイエンティストのリストの作成に当たっては、Clarivate Analytics社により提供されるWoSのCustom XMLデータ(2016年12月時点)と、同社の定義する22研究分野と各文献が発表されるジャーナル名の対照テーブル(2014年時点)を使用した。抽出対象となる文献は各分野と結びついたジャーナルに掲載された文献のみであり、そのほかは抽出対象から外れることに留意されたい。

リスト作成の基準となる高被引用文献の抽出

1981-2014年の各年と各研究分野において、対象となる文献のうち(2016年12月時点での)被引用数が各分野の上位1%に相当する文献を高被引用文献(Highly Cited Papers; 以下、HCP)と定義する。

スター・サイエンティストリストの抽出手順

1. 選定されたHCPのうち、2005-2014年の10年間で発表されたものの著者リストをWoSから抽出した。
2. 各著者についてHCPの数を集計し、研究分野毎にHCP数で著者のランキングを作成した。この際、著者名の名寄せが問題となるが、本プロジェクトでは特別な対応策はとらず、WoSの文献データ中に存在する“display_name”の文字列を著者名として集計を行った。“display_name”は基本的に著者のフルネームを含む(ただし、2007年以前の文献データにおいては“display_name”がフルネームでなく、ラストネーム+ファーストネームのイニシャルという形で表記されている場合が多い)ことと、22の研究分野(先述のESI分類に基づく21分野と、学際領域)毎に集計を行っていることから、“display_name”を単純に用いるのみで、名寄せアルゴリズムを特に定めずとも妥当性のあるランキングを作成可能であると判断した。ただし、中国人、韓国人に多い同姓同名や、日本人の同音異字のような著者については、それらを同一人物として扱ってしまうという課題は残る。HCP数が同数の場合、HCP全体の総被引用数が多い著者をより高い順位につけることとする。
3. 作成したランキングに載っている著者(HCPを1本以上発表している著者)の総数を N としたとき、上位 \sqrt{N} 位までを「スター・サイエンティスト・ショートリスト」として抽出した。さらに、その5倍の順位に相当する著者以上のHCP数を有する者を、「スター・サイエンティスト・ロングリスト」として抽出した。例えば、ある分野のHCP数ランキングにおいて著者総数が10,000人であれば上位100位までが「スター・サイエンティスト・ショートリスト」に掲載される。「スター・サイエンティスト・ロングリスト」には、500位までの著者及び、500位の著者とHCP数が同数である501位以下の著者が掲載される。

4. 著者の所属機関については、WoS に収録された著者-所属機関の対応テーブルを利用した。各著者が HCP を発表した所属機関を集計し、最も多く HCP を発表している所属機関を第 1 所属機関、その次に多く HCP を発表している所属機関を第 2 所属機関と定義する。日本機関所属の HCR は、(1)~(3)の方法で HCR として抽出された著者の中で、第 1 所属機関もしくは第 2 所属機関が日本にある著者のことを指す。なお、2007 年以前の文献データについては WoS の仕様上、複数の著者がいる場合に各著者の所属機関を特定することができない。
5. 日本機関所属の著者については、「スター・サイエンティスト・ロングリスト」と、2014-2017 年の 4 年間の CA 版 HCR を合算したものを「スター・サイエンティスト・インテグレートド・リスト」として構築した。

スター・サイエンティスト・リストの概要

構築したスター・サイエンティストの各リストでは、表 2 に示すように対象著者(“display_name”)ごとに、12 の変数を集計している。各変数の意味や集計方法については表 3 を参照されたい。

上記の手法により、スター・サイエンティスト・ショートリストは 3,808 人、スター・サイエンティスト・ロングリストは 24,431 人が抽出された。その中で、日本機関所属の者についてはショートリストで 83 人、ロングリストで 514 人である。なお、ショートリストのうち 37 人、ロングリストのうち 689 人は所属機関が特定できていない。

2014-2017 年の間に CA 版 HCR に 1 度以上ノミネートされた日本機関所属の著者 121 人のうち、41 人がショートリスト、80 人がロングリストに含まれていた。

本リストの貢献点は以下の通りである。

- WoS を用いて、詳細な情報が記載された個々の書誌情報の単位からスター・サイエンティストのリストを構築したことで、同データベースにある情報に基づき、多数の変数を含めることが可能となった。これにより、スター・サイエンティストのより詳細な分析が可能となる。
- スター・サイエンティスト・ロングリストを構築したことにより、スター・サイエンティストの母集団を増やすことができたので、より柔軟な定量分析が可能となる。

一方で、今後検討すべき課題は、以下の通りである。

- CA 版 HCR との選出基準の違いなどの理由から、高いパフォーマンスを持つ研究者の一部が含まれていない可能性がある。どの基準がより正しいといった判断を行う必要はないと考えられるが、CA 版 HCR と本プロジェクトのリストの結果が異なる理由としては以下が考えられる。
 - CA 版 HCR では、本プロジェクトのリストよりも、被引用数による選出基準が細かく設定されており、その結果近年 HCP を多く出すようになった著者は選出されにくくなっている。本プロジェクトのリストにおいても選出基準に被引用数が含まれているが、HCP 数が多ければ被引用数が少なくてもリストに選出されるようになっていることによる、差異が生じていると考えられる。
 - 本プロジェクトのリストで定義される HCP の対象となるジャーナルは 2014 年時点で選択されたものである。一方で 2015-2017 年に CA 版 HCR として選出された著者の中には、2014 年の時点で対象になっていないが、2015 年以降新たに対象となったジャーナルで多くの HCP を出していた著者も存在すると考えられる。

クロスフィールド・スター・サイエンティストリストの抽出手順

前述の通り、分野別にスター・サイエンティストのリストを構築したが、複数の分野にまたがって顕著な功績を残している研究者は含まれない場合がある。CA 版 HCR においても同様の問題が指摘されて

おり、対応策として、同社は2018年から“Cross-field”という分野のHCRを発表している。“Cross-field”のHCRは、対象期間中の各研究分野におけるHCP数、HCPの総被引用数を著者ごとに集計してから、分野毎にHCR選定基準となるHCP数、HCPの総被引用数でそれらを割り、全分野でその合計が双方とも1を超えた著者を抽出することにより得られる。ただし、ESIに従った21の研究分野のいずれかでHCRに選出されているものは含まれない。

本プロジェクトでは、上記のClarivate Analytics社の手法を参考に、研究者の学際性を測るスコアを算出し、クロスフィールド・スター・サイエンティストのリストを作成した。様々なスコアを用いているが、ここでは主に用いるHCP数の学際性スコアの算出方法について説明する。表4のように、3つの分野(数学/微生物学/分子生物学・遺伝学)にてHCPを発表している著者A氏を想定する。各分野においてスター・サイエンティスト・ショートリスト/ロングリストとして選出されるために必要なHCP数は異なる。分野毎にA氏のHCP数をその分野でのスター・サイエンティスト・ショートリスト/ロングリストの最小HCP数で割った値がそれぞれhpc_score_short、hpc_score_longとなる。最下行には分野毎のhpc_score_short、hpc_score_longの合計が示されているが、最終的に学際性スコアとして用いるのはこの値である。

クロスフィールド・スター・サイエンティストリストの概要

構築したクロスフィールド・スター・サイエンティストのリストについては、表5に一部を示した。著者(display_name)ごとに、19の変数を集計している。各変数の定義等は表6を参照されたい。なお、表5では、集計方法や表示形式がスター・サイエンティストリストと同義であるため、country1以降を省略している。このリストでは、HCP数以外に、HCPの総被引用数や責任著者として発表したHCP数についても上記の手法でスコアを算出している。

- スコアの値に関わらず、クロスフィールド・サイエンティストのリストに掲載されている(いずれかの分野でHCPを1本以上発表した)著者(display_name)の総数は658,036人である。このリスト自体も研究者の学際性に関する分析に活用可能である。
- いずれの研究分野においてもスター・サイエンティスト・ショートリストに入っていない著者のうち、hpc_score_shortが1を超える著者をクロスフィールド・スター・サイエンティスト・ショートリストとして抽出する。この手法では5,585人の著者が抽出され、うち日本機関所属のものは120人存在する。
- いずれの研究分野においてもスター・サイエンティスト・ロングリストに入っていない著者のうち、hpc_score_longが1を超えている著者と、上記のクロスフィールド・スター・サイエンティスト・ショートリストを合わせたものを、クロスフィールド・スター・サイエンティスト・ロングリストとして抽出する。このリストには20,348人の著者が抽出され、うち日本機関所属のものは503人存在する。
- クロスフィールド・スター・サイエンティスト・リストについても集計単位となる著者名を示す文字列として“display_name”を用いているが、全分野を合わせて集計しているため、同姓同名の名寄せ問題がより顕在化する。特に中国人の著者に関してこの問題は深刻であり、いずれのリストにおいても上位の多くが中国人となっている。実際に分析をする際はユニークネームを持つ著者に対象を絞るなど、何らかの工夫が必要である。

5. スター・サイエンティスト・コホート・データセットへの拡充

スター・サイエンティストのリストを基盤にしながら、研究を進めるためにより多くのデータセットと結合することにより、スター・サイエンティスト・コホート・データセットを構築した。

機関別 HCP 数ランキング(機関ランキング)の抽出手順

1. WoS の文献-機関の対応に関するテーブルから、1981-2014 年の各 HCP に携わった機関のリストを抽出した。なお、WoS において機関名の表記法は主に 2 通りあり、1 つは“UNIV TOKYO INST APPL MICROBIOL, BUNKYO KU, TOKYO 113 JAPAN”というような機関の正式名称と住所が載ったもので、それに対してもう 1 つは“UNIV TOKYO”のみのような代表的名称(略称)のみを載せるものである。ここでは後者の機関名を全て小文字に変換して利用する。その結果、1 つの論文に対して重複する機関名が存在した場合は重複分を削除した。例えば、“NASA,WASHINGTON,DC 20546”と、“NASA,LYNDON B JOHNSON SPACE CTR,HOUSTON,TX 77058”の 2 通りの機関名が 1 つの論文に記載されている場合があるが、両者とも略称は“NASA”で重複するため、これらの重複分を削除する。
2. 1981-2016 年の各年と各研究分野について、1 で作成したリストから HCP 数の機関ランキングを作成する。HCP 数が同数の場合は、HCP のみでの総被引用数が多い機関が上位にランクするようにする。スター・サイエンティストのリストでは、2005-2014 年の情報から各分野について 1 つのランキングを作成しているが、このランキングにおいては、研究分野毎に 36 年分のランキングを作っている。また、各機関の所属国情報を利用して、各年と各研究分野における HCP 数の国別ランキングも作成した。

国内機関分析用パネルデータの作成手順

上記の手法により作成された機関ランキングのうち日本所属の機関のみを抽出したものと、NISTEP(2018)の提供する大学・公的機関名英語表記ゆれテーブル(以下、NISTEP テーブル)を結合することで、国内機関について、国立大学を中心に研究動向等の変化を分析するためのパネルデータを作成した。以下に、NISTEP テーブルの整形方法と、機関ランキングとの結合方法を示す。

- 1 NISTEP テーブルから、“データ源”列の値が“WoS”及び“辞書、WoS”であるもののみを抽出した。この時点で 7,947 件のデータが残った。その後、“表記バリエーション”列の英語機関名を全て小文字に変換した。
- 2 重複のある列の削除操作を行った。
 - ① “代表機関名”、“表記バリエーション”列の値が共に重複しているものは削除した(26 件)。現状存在しない機関(“現状”列の値が“No”)であり、最終機関が当時から変わっているものについて、“最終機関名”と“表記バリエーション”列の値が共に重複しているものを削除した(317 件)。この操作を行った後、現在存在しない機関については、“代表機関名”の値を継承された最終機関の“代表機関名”の値に置き換えた。置換後、新たに“代表機関名”、“表記バリエーション”列の値が共に重複してしまうものは削除した(2 件)。
 - ② 以上の操作を行ったうえで、“表記バリエーション”列の英語機関名が同じでも代表機関名が異なるものが計 762 件存在していた。このうち、英語機関名の重複が 3 件以上となっている組については、“dept ~”や“res inst ~”など英語名称が一般的で区別不能なものが多く含まれており、一括削除した。英語機関名の重複が 2 件である組(180 件=90 組)についても、英語名称が一般的であるものについては削除するが、そうでない場合は次のように対応した。双方の代表

機関名の表記バリエーションの中で、該当する英語名より“WoS 頻度”の多い英語表記が存在することを“WoS 頻度条件”と定める。一方が WoS 頻度条件を満たし、他方が満たさない場合は満たしていない方のみを残す。双方が条件を満たすか、双方が条件を満たしていない場合は、どちらの代表機関名が適切なかが特定不能であるとして、一括削除する。以下に適用例を示す。

- ・ 静岡大学、静岡県立大学の両方が“univ shizuoka”という表記バリエーションを持つが、静岡大学の英語表記名としては“shizuoka univ”の方が適切であり、WoS 頻度も多いため、“univ shizuoka”は静岡県立大学として扱う。
- ・ TOA 株式会社、東亜建設工業株式会社は両方とも“toa corp”というバリエーションを持つが、どちらもそれ以外の表記バリエーションがなく、表記バリエーションから機関名を特定できないため一括削除する。
- ・ 東海大学、独立行政法人国立病院機構は両方とも“tokyo hosp”というバリエーションを持つが、双方ともより WoS 頻度の多い代表的な英語表記名が存在するため、一括削除する。

3 重複削除操作後、機関ランキングに存在する日本所属の機関(英語機関名)と上記で整形した NISTEP テーブルを結合した。英語機関名をキーとした結合操作の結果、NISTEP テーブル側では 5,568 件が結合されなかった(これは、その機関が対象期間中に HCP を 1 本も発表していないことを意味する)。また機関ランキングでは、日本所属の機関 3,629 件中 1,738 件が結合されなかった。1,738 件中 1,674 件は対象期間中に発表された HCP 数が 3 本以下であるので、作業時間を考慮しそれらは先立って削除した。残りの 64 件については、英語表記名から代表機関を特定できた場合のみ、代表機関とそのラベルを新たに加えた。

4 3 での結合後、代表機関名、論文の発表年(1981-2016)毎に変数を集計することでパネルデータを作成した。本稿の執筆時点では以下の変数が含まれている。

- 代表機関の発表した総 HCP 数
- 代表機関の発表した分野毎の HCP 数(22 分野)
- 代表機関の発表した HCP の総被引用数
- 当該年における代表機関の総 HCP 数についての、国内順位
 - ・ HCP 数が同数である場合は HCP の総被引用数の多い方を高い順位に付ける、なお当該年の HCP 数が 0 の場合は、パネルデータに存在する総機関数である 3,555 位とする。
- 当該年における国内および全世界で発表された総 HCP 数
- 1981 年から当該年までの累積 HCP 数
- 被引用荷重 HCP 数

6. まとめ

本稿では、スター・サイエンティストのリスト作成のプロセスをまとめ、また今後の研究の基盤となるデータセットの概要を記した。本稿で紹介したデータセットは、JST-RISTEX 科学技術イノベーション政策のための科学「スター・サイエンティストと日本のイノベーション」プロジェクトと連携することで活用可能である。なお、このデータセットを利用する場合には、必ずこのワーキングペーパーを引用して欲しい。

スター・サイエンティスト研究には今後も調査すべき研究課題が多数存在する。例えば、以下のような課題が考えられる。

- ・ 大学の定年制によるスター・サイエンティストのパフォーマンスの影響
- ・ 大学の知財管理方針の変更によるスター・サイエンティストのパフォーマンスへの影響
- ・ 大学の知財管理方針変更後のスター・サイエンティストの知財取得戦略の変化
- ・ 大学の講座制廃止による若手研究者の研究パフォーマンスへの影響
- ・ スター・サイエンティストの男女差によるパフォーマンスへの影響
- ・ スター・サイエンティストの卵の同定とその後のキャリアに関する調査

スター・サイエンティスト研究は、Zucker 氏、Darby 氏の一連の研究により大きく進展した。しかしながら、地域、時代、研究分野などの外的妥当性の検証はまだ不十分である。またサイエンティストに関するより多様なデータが取得可能となった現在、研究領域としてはこれからも拡大していくことが予測される。

謝辞

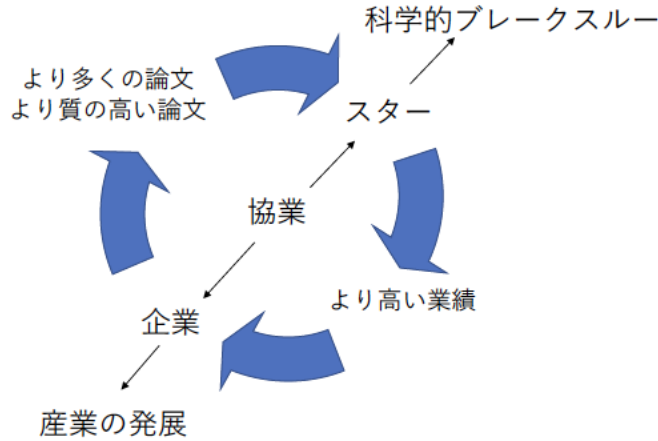
本研究は、JST-RISTEX 科学技術イノベーション政策のための科学「スター・サイエンティストと日本のイノベーション」の助成を受けたものである。その他、JSPS 科研費(15H03377、25705008)の研究成果の一部も含まれている。

本稿執筆にあたっては、プロジェクト・メンバーからの多数のサポートを受けた。データセット構築にあたっては、カリフォルニア大学ロサンゼルス校の Lynne Zucker 氏、Michael Darby 氏、カリフォルニア大学サンディエゴ校の Josh Graff-Zivin 氏の助言を受けた。これら数々のご助力に対して、記して感謝する。

参考文献

- Clarivate Analytics. "Purpose and Methodology" , 2018.
<https://hcr.clarivate.com/methodology/purpose-and-methodology/> (2019/6/13 閲覧)
- Hara. Y. (2017) "Creating Star-Scientist Data Platform", GRIPS-UCLA Dialogue for Evidence-based STI Policy (Semi-Closed Meeting), GRIPS, Tokyo, September 12, 2017.
- J. T. Li, What we learn from the shifts in highly cited data from 2001 to 2014? *Scientometrics*, 108, 57–82 (2016).
- Nagane H. Fukudome Y, Maki K. "An Analysis of Star Scientists in Japan," ICE/IEEE conference , Stuttgart, Germany, June 17-20, 2018
- Saito, H., Fukudome, Y. and Maki, K. "Analyzing the Star Scientists in Japan," Star Scientists and Knowledge Transfer between Academia and Industry: Towards Promotion of Innovation., GRIPS, Tokyo, Sep 11, 2017.
- Saito, H. and Maki, K. "Star Scientists and Entrepreneurship: A Consideration on University Ranking in Asia" Global Entrepreneurial University Metrics, Triple Helix, Workshop III, Stanford University, Palo Alto, 10-11 June 2017.
- Sasaki, T, Nagane H, Fukudome Y, Maki K Asia Innovation Conference by Stanford university, "Innovation policies and star scientists in Japan"(presented by co-author), Tsinghua University in Beijing, China, September 10th, 2018.
- Sasaki, T and Maki, K, "Innovation Policies and Star Scientists in Japan", CEAFJP WORKSHOP "Who Changes the Status Quo? The Role of Star Scientists in Science Intensive Industry", December 2018
- Sumikura K., Sugai N., and Maki K., "The involvement of San Diego-based star scientists in firm activities," ICE/IEEE ITMC International Conference on Engineering, Technology and Innovation, Stuttgart, Germany, June 19th, 2018.
- Sumikura, K., Saito, H., Sugai, N. and Maki, K. "Function of academic knowledge in innovation," Star Scientists and Knowledge Transfer between Academia and Industry: Towards Promotion of Innovation., GRIPS, Tokyo, Sep 11, 2017.
- Zucker, L. G. and Darby, M. R. (2001) 'Capturing Technological Opportunity Via Japan's Star Scientists: Evidence from Japanese Firms' Biotech Patents and Products', *The Journal of Technology Transfer*, 26(1/2), pp. 37-58.
- Zucker, L. G., Darby, M. R. and Armstrong, J. S. (2002) 'Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology', *Management Science*, 48(1), pp. 138-153.
- Zucker, L. G., Darby, M. R. and Brewer, M. B. (1998) 'Intellectual human capital and the birth of US biotechnology enterprises', *American Economic Review*, 88(1), pp. 290-306.
- 齋藤裕美・牧兼充、「スター・サイエンティストが拓く日本のイノベーション」、一橋ビジネスレビュー、pp. 42～56.、Summer 2017.
- 菅井内音・隅藏康一・福留祐太・長根(齋藤)裕美・牧兼充、「スター・サイエンティストに着目した日米の特許分析」、日本知財学会第16回年次学術研究発表会(大阪工業大学、2018年12月1日)
- 隅藏康一・菅井内音・牧兼充(2019)「日本における高被引用研究者の現状～東大・京大とUCSDに着目して」『研究技術計画』、34巻2号、139-149。

- 隅藏康一・菅井内音・牧兼充「サンディエゴ地域におけるスター・サイエンティストと企業との関わり」第32回研究・イノベーション学会年次学術大会、京都大学、2017年10月29日
- 隅藏康一・菅井内音・牧兼充、「特許データから見た地域内外の人材移動：San Diegoのケース」第15回日本知財学会年次学術研究発表会、国士舘大学、2017年12月3日
- 隅藏康一・菅井内音・福留祐太・牧兼充、「スター・サイエンティストの日米比較：東京大学とUCSDに着目して」、日本機械学会2018年次大会（関西大学、2018年9月10日）
- 長根(齋藤)裕美・福留祐太・牧兼充、第32回研究・イノベーション学会年次学術大会「日本のイノベーション政策とスター・サイエンティスト」（共著）、（東京大学、2018年10月26日）
- 長根(齋藤)裕美・林元気・牧兼充、第16回日本知財学会年次学術研究発表会「スター・サイエンティストに着目した日米の特許分析」（共著者による報告）、（大阪工業大学、2018年12月1日）
- 長根(齋藤)裕美・牧兼充、「日本のイノベーションとスター・サイエンティストの役割：現状と課題」、SciREX Working Paper [SciREX-WP-2018-#01]、2018年6月、
https://scirex.grips.ac.jp/resources/archive/180601_1119.html
- 原泰史 (2017) 「バイオエコノミーの「現在地」と「未来地図」-特許、論文、ファンドおよびソーシャルデータを用いた解析と展望-」、バイオインダストリーとイノベーション、Vol 75、No.5
- 原泰史 (2018) 「ビジネスとサイエンスの循環が生み出したノーベル賞」、産学官連携ジャーナルバイオインダストリーとイノベーション、Vol 75、No.5
- 原泰史「ドアをノックするのは誰だ？：スターサイエンティストの検出と特性分析」Institut Mines-Télécom・立命館大学イノベーションマネジメント研究センター共催、2019年2月22日
- 原泰史 (2017) 「日本のスター・サイエンティスト分析に係るデータプラットフォーム整理」、研究・イノベーション学会、京都大学、2017.10.29
- 牧兼充・長根(齋藤)裕美、「1.1.4 スター・サイエンティスト サイエンスとビジネスの好循環が新産業を創出する」、科学技術イノベーション政策研究センター編「科学技術イノベーション政策の科学: コアコンテンツ」、2019年4月、<https://scirex-core.grips.ac.jp/1/1.1.4/main.pdf>
- 文部科学省科学技術・学術政策研究所(NISTEP). “大学・公的機関名英語表記ゆれテーブル (Ver. 2018.1) ”、2018. <http://www.nistep.go.jp/research/scisip/data-and-information-infrastructure>



出典：Zucker & Darby(2007) "Virtuous Circles in Science and Commerce"より転載。翻訳は筆者による。

図 1: サイエンスとビジネスの好循環

「スター・サイエンティスト・コホート・データセット」

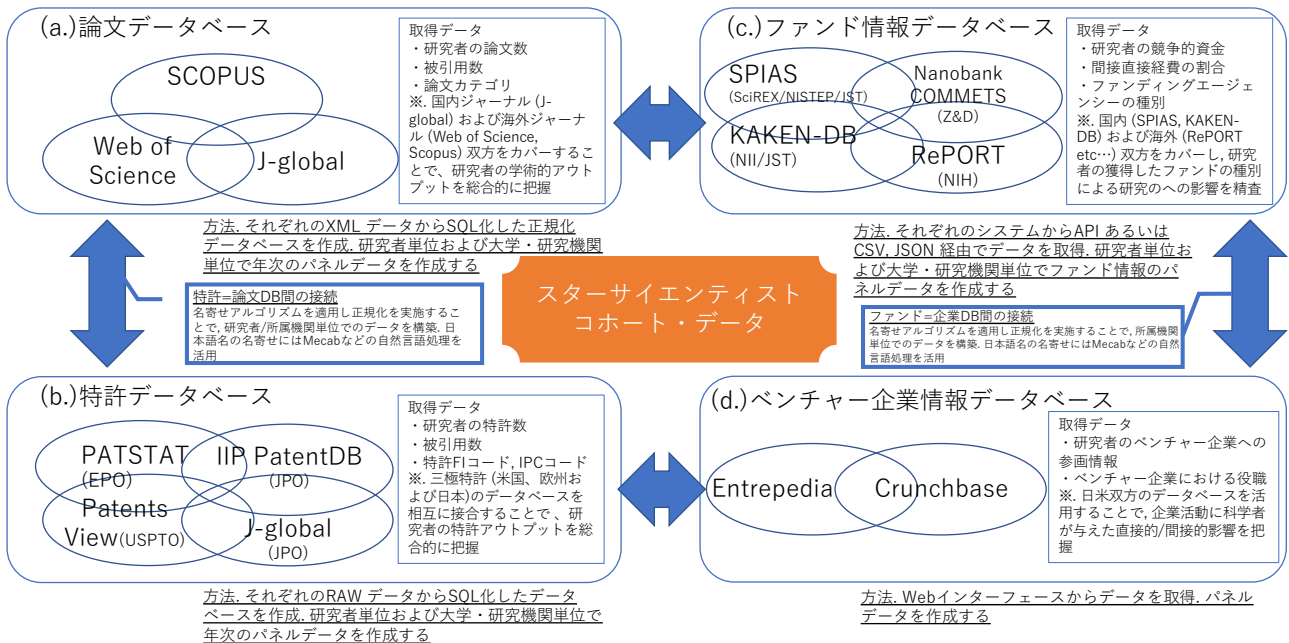


図 2: スター・サイエンティスト・コホート・データセット

表 1: 先行研究におけるスター・サイエンティストの抽出方法

著者	論文名	出版年	スター・サイエンティストの抽出方法
Zucker et al.	Intellectual capital and the birth of US biotechnology enterprises	1998	GenBank に登録されている研究者の中でトップ 0.75%に入る者
Zucker et al.	Virtuous circles in science and commerce	2007	Clarivate Analytics が公表する Highly Cited Researchers の該当者
Azouley	Superstar Extinction	2010	PubMed におけるトップ 1%の引用数を多数出版する研究者
Oettl	Reconceptualizing Stars: Scientist Helpfulness and Peer Performance	2012	他の研究者の助けになっているか (helpfulness) を指標としている。具体的には、論文の謝辞に名前が登場しているかで測定。

表 2: スター・サイエンティスト・リストの一部

display_name	hcp_count		wos_total	
yamanaka, shinya	8		3337	
	country1	organization1		c1
	Japan	Kyoto Univ		8
	country2	organization2		c2
	USA	Gladstone Inst Cardiovasc Dis		5
	organization_count	reprint_count	star	hcp_esi
	4	3	2	biology

表 3: スター・サイエンティスト・リストの変数についての定義

変数名	変数説明
display_name	WoS データベース上での著者の名前。以下の変数の集計単位となる。名前は「(ラストネーム), (ファーストネーム) (ミドルネーム)」の順に掲載されている。ただしこのミドルネームはイニシャルのみであることが多く、場合によってはファーストネームもイニシャルのみであることがある(2007年以前のデータについてはファーストネーム以降全てイニシャルである)。
hcp_count	対象となる研究分野において、2005-2014年の間に著者が発表した HCP 数の合計。
wos_total	対象研究分野において著者が発表した HCP の、2016 年末時点での総被引用数の合計。
country, organization, c	著者の所属機関に関する情報 (6 つの変数) である。c はその所属機関で著者が発表した HCP の数を示し、この値が最も多い機関と 2 番目に多い機関の機関名 (organization) と国名を (country) 示している、機関名は WoS 内で正規化された値を用いている。
organization_count	著者が対象研究分野・対象期間内に HCP を 1 本以上出した機関の数を示す。特に、3 つ以上の機関で HCP を発表した場合、発表した HCP 数が 3 番目以降の機関名を明記しないので、この列に機関数のみが記録されることとなる。
reprint_count	著者が対象研究分野・対象期間において責任著者として HCP を発表した数。スター選出の基準にこの数は含まれていないが重要な変数である。
star	その著者が short list star である場合は 2、long list star である場合は 1 とする (star でない著者は 0)。
hcp_esi	研究分野を示す。研究分野ごとに作成しているランキングにおいてはこの変数は存在しないが、日本機関所属のスターをまとめたリストなど複数分野のスターをまとめている場合この変数を追加する。

表 4: 研究者の学際性スコア算出例

研究分野	A 氏の HCP 数	star(long list)の最小 HCP 数	star(short list)の最小 HCP 数	hcp_score_long	hcp_score_short
Mathematics	2	3	6	0.667	0.333
Microbiology	3	2	4	1.5	0.75
Molecular Biology & Genetics	2	5	12	0.4	0.167
			合計	2.567	1.25

表 5: クロスフィールド・サイエンティスト・リストの一例

display_name	hcp_score_short	citn_score_short	reprint_score_short
knight, rob	15.743	52.575	3.697
	hcp_score_long	citn_score_long	reprint_score_long
	30.633	188.368	7.167
	esi_count	star_short	star_long
	10	4	4
	hcp_count	wos_total	reprint_count
	108	47799	29

表 6: クロスフィールド・サイエンティストリストの変数定義

変数名	変数説明
display_name	著者の名前(各変数の集計単位)、表 3 と同様。
hcp_score_short	各分野におけるショートリストのスターのうち、HCP 数が最小のものを基準とし、各著者について分野毎に HCP 数を基準数で割ったものの総和
citn_score_short	各分野におけるショートリストのスターのうち、HCP の総被引用数が最小のものを基準とし、各著者について分野毎に HCP の総被引用数を基準数で割ったものの総和(選出条件上、基準数より総被引用数が多くてもショートリストに入らない場合があるため有効性は現時点では低い)。
hcp_score_long	各分野におけるロングリストのスターのうち、HCP 数が最小のものを基準とし、各著者について分野毎に HCP 数を基準数で割ったものの総和
citn_score_long	各分野におけるロングリストのスターのうち、HCP の総被引用数が最小のものを基準とし、各著者について分野毎に HCP の総被引用数を基準数で割ったものの総和(ロングリストのスター選出条件に被引用数を含めていないため、ショート同様現時点での有効性は低い)
reprint_score_short	hcp_score_short の"各著者の HCP 数"を、その著者が責任著者となった HCP 数に置き換えて計算したもの
reprint_score_long	hcp_score_long の"各著者の HCP 数"を、その著者が責任著者となった HCP 数に置き換えて計算したもの(現時点では利用していないが有意性はあるはず)
esi_count	各著者が集計期間中に HCP を出した分野の数
star_short	各著者がショートリストのスターに選出された分野の数
star_long	各著者がロングリストのスターに選出された分野の数 (ショートリストにも選出されている場合は除く)
hcp_count	著者の発表した HCP 数(全分野合計、基準数で割っていない)
wos_total	著者の HCP の総被引用数(全分野合計、基準数で割っていない)
reprint_count	表 3 と同様。
country, organization, c	表 3 と同様、著者の所属機関に関する情報 (6 つの変数) であり、c が 1 番目に多い機関と 2 番目に多い機関の機関名と国名を示している。
organization_count	表 3 と同様。