

Title	研究プロジェクトの支出構造 : 科学者サーベイを用いた基礎的な分析
Author(s)	伊神, 正貫; 長岡, 貞男; Walsh, John P.
Citation	年次学術大会講演要旨集, 26: 655-658
Issue Date	2011-10-15
Type	Conference Paper
Text version	publisher
URL	<a href="http://hdl.handle.net/10119/10204">http://hdl.handle.net/10119/10204</a>
Rights	本著作物は研究・技術計画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 Science Policy and Research Management.
Description	一般講演要旨

## 2 G 2 5

### 研究プロジェクトの支出構造：科学者サーベイを用いた基礎的な分析

○伊神正貫（文科省・科学技術政策研），長岡貞男（一橋大），  
John P Walsh（ジョージア工科大）

#### 1. はじめに

研究活動を考える上で、研究開発費は重要なインプットであり、これなしでは多くの研究は実施不可能である。しかし、研究プロジェクトにおいてどのような活動がなされ、それぞれの活動に対して、どの程度の研究開発費が用いられているかについての分析は殆どなされていないのが現状となっている。

そこで、本報告では日本の科学者を対象とした知識生産過程についての大規模調査[1]から得られたデータを用いて、大学における研究プロジェクトの支出構造について分析を行う。具体的には、研究プロジェクトに費やした人月、研究チームの構成(ポストドクターの参加など)、研究施設・設備等へのアクセス、研究マネジメント、研究成果の商業化といった研究プロジェクトにかかわるさまざまな活動の状況と研究プロジェクトで費やした研究資金の関係について分析する。特に被引用数が上位1%の論文を生み出した研究プロジェクトとそれ以外の研究プロジェクトの違いに注目し、両者に活動形態の違いが存在するかについて考察する。

#### 2. 分析に用いたデータ

分析には、一橋大学イノベーション研究センターと科学技術政策研究所が共同で実施した「科学における知識生産プロセスに関する調査」(以降、科学者サーベイと呼ぶ)から得られたデータを用いた[1]。

科学者サーベイでは、研究チームの構成(論文著者の地位や専門分野など)、研究マネジメントの状況、研究プロジェクトで使用した研究資金額、研究資金の資金源、研究プロジェクトから生み出された論文等のアウトプットなど、

研究プロジェクトについての包括的なデータが収集されており、研究プロジェクトにおける活動と研究プロジェクトの支出との関係の分析が可能である。

また、科学者サーベイでは、被引用数が上位1%である高被引用度論文をもたらした研究プロジェクトとそれ以外の論文(通常論文)をもたらした研究プロジェクトについてデータを収集しているので、両者のプロジェクトの違いについても分析が可能となっている。

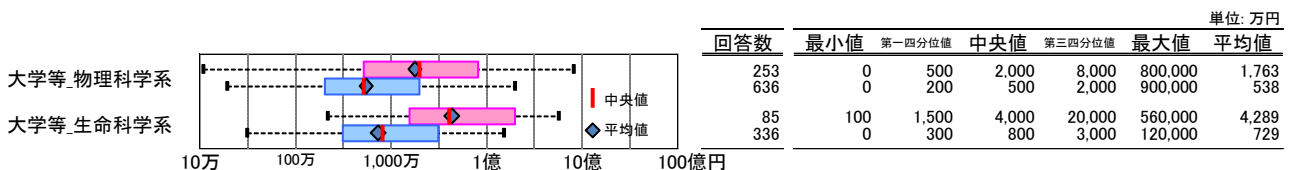
本報告では、大学等における物理科学系、生命科学系の研究プロジェクトに注目する。

#### 3. 研究プロジェクトに費やした研究資金

まず、研究プロジェクトに費やした研究資金を見る。科学者サーベイでは人件費として、当該プロジェクト実施のために雇用していた研究者(プロジェクトのための特任研究者)や研究支援者の人件費のみを計測している。また、設備の整備費については、もっぱら当該研究プロジェクトのために整備した設備は「研究資金」に含め、そうでない場合(既存の設備を利用したなど)は除外されている。

図1に研究プロジェクトに費やした研究資金の分布を示す。研究資金の分布は対数正規分布に概ね従っているため、ここでは対数スケールで値を示している。通常群に注目すると、大学等の物理科学系における研究資金の中央値は500万円、生命科学系における中央値は800万円となっている。高被引用度論文産出群が費やしている研究費は中央値で通常群の4~5倍となっており、高被引用度論文産出群の方が多くの研究資金を費やしている傾向がみられる。

図1 研究プロジェクトに費やした研究資金



注1: 赤の箱ひげ図が高被引用度論文産出群、青の箱ひげ図が通常群に対応している。箱の左端が第一四分位値、右端が第三四分位値に対応する。また、線記号は中央値、四角記号は平均値を示す。ひげの左端は5パーセンタイル、右端は95パーセンタイルを示す。

注2: 研究費の分布は対数正規分布に従っているため、ここでは対数表示で値を示している。

### 3. 説明変数

研究プロジェクトの支出はさまざまな要因によって変化すると考えられる。本報告では要因として、プロジェクトのサイズ、研究チームの構成、研究施設・設備やデータベース・インターネットの利用状況、研究マネジメント、研究成果の商業化を考える。それぞれに対応する説明変数として、科学者サーベイから得られたデータを用いる。以下では、分析に用いた説明変数と、説明変数と研究プロジェクトの支出との関係についての仮説をまとめる。

#### ① プロジェクトのサイズ

研究プロジェクトの支出はプロジェクトにかかわった研究者の数や期間に依存すると考えられる。そこで、プロジェクトのサイズについての説明変数として、つぎの 2 変数を導入する。

- 研究に費やした人月 [*lg\_manmonths; numeric, log*]
- 研究プロジェクトの年数 [*lg\_duration; numeric, log*]

科学者サーベイで調査対象とした研究プロジェクトの中には、調査実施時点で終了していないものもある。その際は、実質的に研究を開始した年から 2010 年(調査実施時点)までに経過した年数を用いた。これらの説明変数は、いずれも研究プロジェクトの支出の増加に寄与すると考えられる。

#### ② 研究チームの構成

研究チームは、博士・修士・学部学生、ポストドクター、技能者、教授などの教員など多様な人材から構成されている。科学者サーベイにおける人件費は、研究プロジェクトのために雇用された研究者について計測されている。したがって、人件費についてはポストドクター等の雇用にかかわるものが主に計測されており、若手研究者の参画が、研究プロジェクトの支出の増加に寄与すると考えられる。そこで、研究チームの構成についての説明変数として、つぎの 4 変数を導入した。ポストドクター数、博士・修士・学部学生数、シニア研究者数の和が、調査対象論文の著者数に対応する。

- ポストドクター数 [*posdoc\_authors; numeric*]
- 博士・修士・学部学生数 [*other\_yng\_authors; numeric*]
- シニア研究者数 [*senior\_authors; numeric*]
- 技能者数 [*technicians; numeric*]

#### ③ 研究施設・設備やデータベース・インターネットの利用状況

研究施設・設備は研究プロジェクトを実施する上で、重要な要素である[2]。例えばスーパーコンピュータの発展

により、これまでは不可能であった大規模かつ高度なシミュレーションが可能になり、走査型プローブ顕微鏡の発明によってナノテクノロジーの発展が促された。本報告では、研究施設・設備の利用状況の説明変数として、つぎの 2 つを導入した。

- 研究チームで保有する最先端の設備・施設の利用の有無 [*adv\_equip\_owned; binary*]
- 外部の最先端実験設備・施設(加速器、スーパーコンピュータ、望遠鏡など外部の設備・施設)の利用の有無 [*adv\_equip\_ext; binary*]

研究プロジェクトの支出との関係については、前者については支出の増加に寄与すること予想されるが、後者についてはその効果は明らかではない。例えば大型放射光施設(SPring-8)については、論文等により研究成果を公表する場合、ビーム使用料は無料となる。

これらに加えて、データベースやインターネットの利用状況についての説明変数も導入した。説明変数は、それぞれを利用した際に 1、利用していない場合は 0 となるダミー変数である。

- 国内外の最先端の研究情報(ジャーナルで発表される前の情報)へのアクセス [*pre\_pub; binary*]
- インターネットを通じた遠隔地の研究者の研究への参加 [*remote\_resear; binary*]
- 最先端のデータベース(ゲノム、材料など、論文は除く) [*materials; binary*]
- 論文データベース [*literature; binary*]

#### ④ 研究マネジメント

研究チームの構成や研究施設・設備等の利用状況に加えて、研究マネジメントも研究プロジェクトの支出に影響すると考えられる。そこで、研究プロジェクトの知識の蓄積、研究の効率化、研究者コミュニティとの交流にかかわる以下の説明変数を導入した。これらの研究マネジメントの実施の有無の研究プロジェクトの支出への影響は明らかではない。説明変数は、研究マネジメントを実施した際に 1、実施していない場合は 0 となるダミー変数である。

(研究プロジェクトの知識の蓄積)

- ラボノートや実験ノートなどへの研究過程の記録 [*archivnotes; binary*]
- 研究成果のデータベースへの蓄積 [*database; binary*]

(研究の効率化)

- アウトソーシングを含めた作業分担を通じた研究の効率化・高速化 [*divisoim\_res; binary*]
- 研究チームで保有している実験設備の継続的な改善 [*imprv\_facilities; binary*]

- 計算・シミュレーションプログラムの改善  
[imprv\_comput; binary]
- (研究者コミュニティとの交流)
- 学会発表を通じた情報の共有・研究の評価  
[inf\_sharing\_assess\_conf; binary]
  - 新分野開拓のための研究者コミュニティの確立  
[dev\_res\_comm; binary]

#### ⑤ 研究成果の商業化

研究プロジェクトの活動は、科学的な知識の獲得にとどまらず、そこから得られた研究成果の商業化も含む。基礎研究の成果を応用や開発に結び付けるための活動は、研究プロジェクトに追加の支出をもたらすことが予想される。そこで、研究成果の商業化に関する以下の3つの説明変数を導入した。説明変数は、これらの活動がある場合は1、活動がない場合は0となるダミー変数である。

- 特許の出願の有無[patent\_app; binary]
- 標準化の有無[startup; binary]
- スタートアップ企業の設立の有無[standard; binary]

#### 4. 推定方法

被説明変数を研究プロジェクトの支出[lg\_fund\_project; numeric, log]として線形重回帰分析を行った。全部で6ケースについて推定を行った。それぞれのケースの内容を表1に示した。なお、推定の際には、分野や研究プロジェクトの手法(実験・観察、数値計算・シミュレーション、理論計算、新規の実験方法や新しい実験器具・設備・施設の開発)の影響を制御した。

また、高被引用度論文産出群と通常群の違いが何に起因するかを分析する目的で、説明変数として研究プロジェクトについてのダミー変数[citedness; binary]を導入し、他の説明変数の導入によって、このダミー変数の回帰係数がどの程度変化するかを確認した。ダミー変数は高被引用度論文産出群において1、通常群において0となる。

表1 推定を行った6つのケース

ケース	説明変数
1	プロジェクトのサイズ
2	プロジェクトのサイズ+研究チームの構成
3	プロジェクトのサイズ+研究施設・設備やデータベースやインターネットの利用状況
4	プロジェクトのサイズ+研究マネジメント
5	プロジェクトのサイズ+研究成果の商業化
6	全ての説明変数

#### 5. 推定結果

6つのケースについての推定結果を表2に示す。まず、プロジェクトのサイズに注目する。全てのケースで、研究に費やした人月、研究プロジェクトの年数の回帰係数が正で統計的に有意となっている。ケース6をみると、研究に費やした人月が1%増加すると、研究プロジェクトの支出が0.55%増加することが分かる。

研究チームの構成をみると、ケース2と6のいずれでも、ポストドクター数と技能者数の回帰係数が正で統計的に有意となった。つまり、ポストドクター数と技能者数が研究プロジェクトに参画することで、研究プロジェクトの支出は増加する。ただし、ポストドクターと技能者では回帰係数が異なり、ポストドクターが1名増加する方が、技能者が1名増加するよりも、研究プロジェクトの支出の増加が大きいことが分かる。

研究施設・設備等の利用状況をみると、ケース3では研究チームで保有する最先端の設備・施設、国内外の最先端の研究情報(ジャーナルで発表される前の情報)へのアクセス、論文データベースを利用した研究プロジェクトで支出が増加するとの推定結果が得られた。他の説明変数も含めたケース6では、研究チームで保有する最先端の設備・施設のみ回帰係数が正で統計的に有意である。統計的には有意となっていないが、外部の最先端の設備・施設については回帰係数が負となっている。

研究マネジメントについてはケース4と6のいずれでも、アウトソーシングを含めた作業分担を通じた研究の効率化・高速化、研究チームで保有している実験設備の継続的な改善、計算・シミュレーションプログラムの改善、新分野開拓のための研究者コミュニティの確立の回帰係数が正で統計的に有意となっている。

研究成果の商業化では、特許出願の有無とスタートアップ企業設立の有無の回帰係数が正で統計的に有意である。つまり、特許出願やスタートアップ企業の設立を行った研究プロジェクトの方が、研究プロジェクトの支出が大きくなる。

最後に研究プロジェクトについてのダミー変数をみる。このダミー変数の回帰係数は、いずれのケースにおいても正であり統計的にも有意となっている。ケース1とケース6を比べると回帰係数が減少していることから、さまざまな要因を考慮することで、高被引用度論文産出群と通常群における研究プロジェクト支出額の違いがある程度、説明可能であることが分かる。しかし、ケース6においても研究プロジェクトについてのダミー変数の回帰係数の値が大きいためであることから、推定で用いた説明変数以外の要因が、高被引用度論文産出群と通常群における研究プロジェクト支出額の違いを生んでいることが推察される。



## 6. まとめ

本報告では、研究プロジェクトにかかわるさまざまな活動の状況と研究プロジェクトで費やした研究資金の関係に注目し、高被引用度論文産出群と通常群の支出額の違いが何に起因するかの理解を試みた。

線形重回帰分析の結果から、研究プロジェクトの支出を増加させる要因として、ポストドクターや技能者の研究プロジェクトへの参画、最先端の設備・施設の研究チームでの保有、アウトソーシング等による研究の効率化・高速化、研究者コミュニティの確立、特許出願やスタートアップ企業の設立が統計的にも有意であることが明らかになった。

これらの結果から、高被引用度論文産出群では通常群と比べて、

- 研究プロジェクトの生産性や質を向上させるための活動(ポストドクター等の研究プロジェクトへの参画、最先端の設備・施設の研究チームでの保有など)
- 幅広い研究成果を生み出すための活動(研究者コミュニ

ニティの確立、特許出願やスタートアップ企業の設立など)

が、盛んに行われている事が、高被引用度論文産出群の支出が大きい理由と推察される。

ただし、これらの要因では説明できない部分も大きい。本報告では、多くの説明変数についてダミー変数を用いたが、科学者サーベイでは特許出願数など一部のデータについては実数も計測している。これらの数値データを用いることで、ケースの説明力が向上する可能性がある。また、さまざまな要因を考慮しても高被引用度論文産出群と通常群の差が生まれる要因として、マタイ効果の存在も考えられる。

今後、さらなる分析を進めることで、研究プロジェクトにかかわる活動と研究プロジェクトで費やした研究資金の関係の理解を深めていく予定である。

表2 各ケースの推定結果(研究プロジェクトの支出[ $lg\_fund\_project$ ; numeric, log]が被説明変数)

説明変数	ケース1	ケース2	ケース3	ケース4	ケース5	ケース6	
	b/t	b/t	b/t	b/t	b/t	b/t	
プロジェクトのサイズ	lg_manmonths	0.66 [17.19]***	0.62 [15.73]***	0.64 [16.62]***	0.63 [16.70]***	0.63 [16.74]***	0.55 [14.47]***
	lg_duration	0.42 [6.13]***	0.42 [6.06]***	0.41 [6.00]***	0.41 [6.15]***	0.40 [5.98]***	0.40 [5.99]***
研究チームの構成	senior_authors		0.00 [0.25]				0.00 [0.34]
	posdoc_authors		0.08 [2.83]***				0.08 [2.91]***
	other_yng_authors		0.00 [0.12]				0.01 [0.50]
	technicians		0.05 [3.24]***				0.05 [3.31]***
研究施設・設備等の利用状況	adv_equip_owned			0.43 [4.39]***			0.30 [3.10]***
	adv_equip_ext			0.04 [0.36]			-0.04 [-0.31]
	pre_pub			0.27 [2.09]**			0.15 [1.18]
	remote_resear			-0.05 [-0.46]			0.03 [0.22]
	materials			0.15 [1.59]			0.06 [0.60]
	literature			0.18 [1.79]*			0.07 [0.72]
	archivnotes				-0.09 [-0.68]		-0.15 [-1.17]
研究マネジメント	database				0.07 [0.70]		-0.05 [-0.51]
	division_res				0.38 [3.50]***		0.23 [2.09]**
	imprv_facilities				0.51 [4.56]***		0.40 [3.52]***
	imprv_comput				0.31 [2.57]**		0.30 [2.48]**
	inf_sharing_assess_conf				-0.03 [-0.20]		-0.01 [-0.06]
	dev_res_comm				0.36 [3.61]***		0.29 [2.85]***
	patent_app					0.77 [7.12]***	0.76 [7.05]***
研究成果の商業化	startup					0.27 [0.95]	0.31 [1.09]
	standard					0.39 [2.23]**	0.36 [2.04]**
	project_type						
プロジェクトの種類	citedness	1.16 [11.44]***	1.06 [10.23]***	1.06 [10.37]***	1.06 [10.53]***	1.05 [10.42]***	0.82 [8.04]***
	R-squared	0.445	0.456	0.463	0.477	0.474	0.519
	Adj-R-squared	0.438	0.448	0.454	0.468	0.467	0.505
	N	1253	1234	1249	1252	1243	1221

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

注1: 科学者サーベイで、回答者の対象論文投稿時の所属が大学等であり、対象論文の分野分類が物理科学系、生命科学系であるデータを用いて推定を行った。

注2: 推定の際、分野や研究プロジェクトの手法の影響を制御した。

## 参考文献

[1] 長岡 貞男、伊神 正貴、江藤 学、伊地知 寛博、2010、「科学における知識生産プロセスの研究 -日本の研究者を対象とした大規模調査からの基礎的発見事実-」、文部科学省科学技術

政策研究所 調査資料-191、2010年11月

[2] Stephan, P., 2010, "The Economics of Science," in Hall, B.H. and N. Rosenberg (eds.), Handbook of The Economics of Innovation, Elsevier