TYPE：Research Progress

Multidimensional assessment of R&D performance: evidence from the pilot evaluation exercise of Russian public research institutions

Konstantin Fursov [1] Gleb Kuzmin [2]

[1] *ksfursov@hse.ru*
National Research University Higher School of Economics (Russia)

[2] *gkuzmin@hse.ru*
National Research University Higher School of Economics (Russia)

Abstract

Research evaluation recently became a widely disseminated exercise aimed in the end of the day at improving the cost efficiency of public funding of national R&D sectors. In November 2013, the Government of the Russian Federation initiated a national evaluation exercise of public research institutions (PRIs) to provide information basis for development of S&T policies aimed at increasing effectiveness and strengthening the role of R&D performing institutions in economic and social development. The aim of this paper is that of providing an approach for multidimensional assessment of R&D performance based on quantitative data derived from the national evaluation exercise, specifically looking at its applicability and limitations for further analysis and preliminary differentiation of PRIs as well as for use in policymaking.

Conference Topic
Science policy and research assessment

Introduction

Research evaluation of public institutes is considered important for measuring the performance of R&D sector, first of all, in terms of improving the cost efficiency of public funding. Various studies (e.g. see Luwel et al., 1999; Senker 2001; Coccia, 2004; Abramo & Angelo, 2015; Ancaiani et al., 2015) show a growing interest in evaluating national scientific performance with the use of different metrics and models.

In Russia the first round of the national evaluation exercise that covered around 400 PRIs subordinated to the national science academies and federal agencies was implemented in 2009-2011. This initiative involved a wide range of indicators (R&D expenditure, personnel, research equipment, publications, citations, journal impact-factors, patents and other IPRs, cooperation, innovation infrastructure, income and profitability, etc.), but its results did not lead to any substantial change neither in policy instruments nor in the institutional structure of the national R&D sector. As for the use of performance-based indicators for evaluating individual researchers, there was no common framework established at the national level, but a number of research institutions and universities adopted certain practices within their human resource development strategies.

In 2013, a new round of research evaluations was launched taking into account the lessons learnt from the previous round. The distinctive features of this latest initiative included (1) an open and interagency approach in selecting criteria for evaluation; (2) combination of quantitative and qualitative (peer-review based) assessment procedures applicable, if necessary, also at the level of research teams; (3) grouping of R&D institutions into “reference groups” on the basis of respective research fields and basic functions; (4) a 5-year evaluation period. A list of indicators was revisited and comprised both conventional inputs (R&D expenditure and personnel) as well as research equipment and novel output categories. In particular, there were added different types of publications (journal articles, conference proceedings, books and book chapters), citations and impact-factors; results of inventive work like designs, blueprints, patents and other IPRs; and financial results such as income from technology transfer or S&T services.

While this exercise has not yet been finished and its results have not been implemented for decision-making, it allowed constructing an advanced model of knowledge production function taking into account different types of R&D outputs. Furthermore, the exercise stimulated 'Agora processes' (Barré, 2001, 2005) in the scientific community around the use of S&T indicators for research evaluation and motivated academics to accept expert roles in forthcoming peer-review procedures.

The aim of this paper is that of providing an approach for multidimensional assessment of R&D performance based on quantitative data, specifically looking at its applicability and limitations for further analysis and preliminary differentiation of organizations as well as for use in policymaking.

Data sources and methods

Data for this study is derived from the Federal System for Monitoring of R&D Performance maintained and updated by the Ministry of Education and Science of the Russian Federation (http://www.sciencemon.ru/) in September 2016. Reference year is 2015. Primary information was collected from institutions performing R&D in line with the guidelines explaining definitions of key terms and methodology for reporting key variables. The dataset available included information on different characteristics of R&D performance for the 1625 institutions including PRIs, universities and research divisions of state companies. For further analysis to provide higher homogeneity of objects under assessment only PRIs were selected. Additionally arithmetic and logical controls were conducted for key variables to construct a final sample covering 815 observations. Research outputs considered for further analysis were journal articles, conference proceedings and book chapters indexed in Web of Science; blueprints and designs, registered intellectual property rights (IPRs) and others. Additionally a number of financial results such as income from technology transfer, S&T services and contractual works provided were taken into account.

Preliminary correlation analysis demonstrated strong significant cohesion between basic research and publication output. At the same time, as seen from the table 1, organizations implementing applied research and experimental development are likely to be the main providers of S&T services and have weak linkages to IPR protection. In other words organizations with identical balance between different types of research activities may have different results and vice versa. Therefore, they cannot be included into a single and homogenous reference group and compared. Similarly, it could be shown that other characteristics such as type of institution, legal status of an organization or its size are weak to differentiate R&D outputs.

Table 1. Pearson correlation matrix

Variable		*Publications*		*IPRs*		*S&T services*	
		*abs**	*norm**	*abs*	*norm*	*abs*	*norm*
R&D expenditure		-					
Basic research	value	0.76**		0.06		0.64**	
	% of total		0.14**		-0.20**		-0.19**
Applied research	value	0.10**		0.39**		0.61**	
	% of total		-0.13**		0.15**		0.15**
Experimental	value	0.18**		0.23**		0.59**	

development	% of total		-0.05		0.14**		0.13**

* abs = absolute values, and norm = normalised values; ** (Sig. = 0.99)

In order to construct the reference groups for PRIs a two-step algorithm was implemented. At the first step, research organizations were divided into groups by field of science (the first level of the OECD FOS classification (OECD, 2002) in line with the specific research areas mentioned by the organizations. As a result, an organization may refer to more than one field of science. At the second step, within each field of science, a research profile of an organization describing its orientation towards one or several social functions, i.e. knowledge generation, technology development or provision of S&T services, and consequently corresponding with certain types of research outputs was identified. Herewith a research profile was determined with the use of the following key indicators:

A. The number of papers indexed in Web of Science per 100 researchers.

B. Number of IPRs registered in the Russian Federation or abroad as well as the number of issued design documentation per 100 researchers.

C. Income from contractual R&D, S&T services provided and technology transfer per total R&D personnel of an organization.

As seen from the above, to reduce scale effects each indicator (A – C) reflecting certain type of outputs is divided by the number of employees, mainly involved in its acquisition. While researchers tend to be mostly engaged in publications and IPR production, financial results require involvement of different categories of R&D personnel.

Consequently, an organization refers to a particular research profile reflecting its bent for certain research function and corresponding output, if one or more of the indicators mentioned above (A – C) is not zero and equal or exceeds the median value for the field of science. It allows identification up to eight research profiles (see figure 1). An intersection of field of science and a research profile constitutes a reference group.

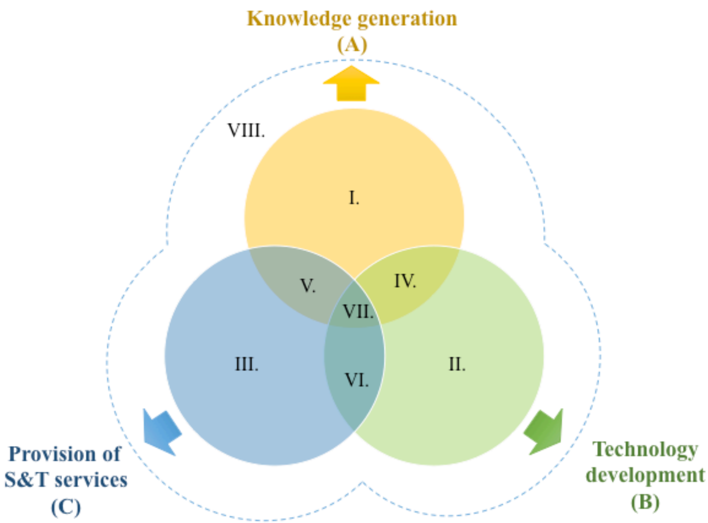

Figure 1. Research profiles of organisations

For further differentiation of organisations within a reference group, a second median value for the same set of core performance indicators of a research profile in the relevant field of science was taken. For instance, an indicator A will be core for the research profiles I, IV, V and VII. Then, for research profiles IV and V indicators B and C will also be core. The research profile VIII should be specifically mentioned. It includes organizations demonstrating lower level of research productivity by all indicators. At the same time, it may have some other characteristics, not included to the initial selection criteria.

According to the requirements mentioned in the related legal documents regulating overall evaluation procedure (see: http://www.sciencemon.ru/legal/acts/postanovlenie-pravitelstva-rf--ot-1-noyabrya-2013-979/), the leadership within the reference group is distributed as follows:

- 1st category (leading institutions) = median + 25%
- 3rd category (loosing research functions) = median – 25%
- 2nd category (stable research organisations) = others

General distribution of organisation by reference groups and performance categories is provided below as a possible solution obtained from the available sample and implementation of the above-mentioned criterion. To distinguish between 2nd and 3rd performance categories within the VIII research profile it was suggested to use additional criterion measured as a share of private funding in the intramural R&D expenditure. The role of private financing was discussed in (Coccia, 2004). In other words if an organization demonstrates poor performance level, but not "at public expense" (the share of private funding in total R&D expenditure is equal or above the median level within the relevant field of science) it falls into the 2nd category. Otherwise, it is associated with institutions loosing their research functions and requires further peer-review based expertise.

Preliminary results

Achieved sample allowed identification of eight research profiles in six fields of science, which provides adequate representation of reference groups and thereby – the statistical significance of the derived threshold values. In the present structure natural sciences, medical and agricultural sciences as well as technology studies are well represented. Somewhat lower concentration of organizations is seen in in the social sciences and humanities (table 2).

Each research profile is characterized by a unique configuration of resources and results that can not apply the same evaluation criteria and thresholds to the whole population of organizations. In particular, the resulting distribution naturally include confluent profiles of technology developers and service providers (VI) in the field of humanities. Similar misbalances can be seen in other fields.

Table 1. Model distribution of PRIs by reference groups and performance categories

Field of Science	Category	Total	Research profile							
			I.	II.	III.	IV.	V.	VI.	VII.	VIII.
1. Natural Sciences	Total	455	50	55	29	43	69	64	66	79
	1-st category		16	22	12	6	9	11	4	X
	2-nd category		24	17	17	33	60	51	61	28
	3-rd category		10	16	0	4	0	2	1	51
2. Engineering and Technology	Total	158	23	25	16	12	21	19	23	19
	1-st category		8	12	5	1	3	2	0	X
	2-nd category		12	7	11	10	18	17	23	8
	3-rd category		3	6	0	1	0	0	0	11
3. Medical and Health Sciences	Total	124	19	18	13	11	16	17	16	14
	1-st category		7	8	5	2	1	1	0	X
	2-nd category		10	3	7	9	15	16	16	5
	3-rd category		2	7	1	0	0	0	0	9
4. Agricultural Sciences	Total	266	21	34	22	27	39	26	46	51
	1-st category		8	11	7	5	8	2	3	X
	2-nd category		7	15	13	17	29	21	41	25
	3-rd category		6	8	2	5	2	3	2	26
5. Social Sciences	Total	92	13	4	14	8	10	7	15	21
	1-st category		5	2	4	1	1	0	0	X
	2-nd category		4	0	6	7	9	7	14	3
	3-rd category		4	2	4	0	0	0	1	18
6. Humanities	Total	61	9	3	9	1	14	1	7	17
	1-st category		3	1	2	0	2	0	0	X
	2-nd category		3	1	7	1	12	1	7	6
	3-rd category		3	1	0	0	0	0	0	11

Different fields of science demonstrate different functional roles. In particular, in the field of medical sciences most of the organizations appear in the group of technology developers (II and IV profiles, mostly first and second performance categories), agricultural sciences also dominated by the developers as well (specifically in the first category). Social sciences are represented mostly in the VIII profile that is characterized by a high number of ulterior (unexpressed) results mostly due to the high number of libraries and infrastructure elements historically holding status of research organizations. The latter is also true for the technical and technological and natural science areas having a number of pilot plants, experimental stations, etc. The second reason for over-representation of the VIII profile is multiple research orientation by fields of science and inability within the existing framework to divide measurable types of outputs by corresponding fields of science, which is one of the limitations of the proposed approach.

Discussion and further research

The proposed approach for reference group identification and assessment could be easily replicated in different samples. However, it has several significant limitations.

The results of the described distribution are largely determined by the boundary values set for the each reference group. Changing these values may lead to significant changes in the whole picture. In the proposed model the thresholds are set automatically, depending on the behavior of all scientific organizations in a sample. Accordingly, any change in the sample can make a difference in the assessment for an individual organization. Such changes may be both productive (e.g. increased performance indicators of some organizations over time can worsen the situation for others) or counterproductive (for instance, through the introduction of unusual objects in the sample for which the proposed indicators would not be relevant).

Secondly, the method does not exclude the possibility for an organization to change its category due to minimal change in one of the performance indicators. In other words, when two organizations in the same field of science demonstrate about the same level of performance and one of them lies on the threshold separating research profiles, then they may be included in different categories that is logically incorrect, taking into account their almost complete identity. Thus, the organizations holding leadership positions in the same pure profiles (I to III) may lose their advantage in relation to one of the mixed profiles (IV to VII), i.e. improving value of an additional indicator, it may lose its leading position by another one.

Third, the current approach does not allow further differentiation of R&D outputs by field of science. Therefore, performance is measured by the overall output of an organisation that may significantly distort the assessment results in case of its multidisciplinary specialization. Further approach might include development of approaches for bibliometric data verification and application of fractional counting for more accurate assessments.

In the end, it should be mentioned that, such an exersise is rather declarative and should be considered only as a preliminary and rough assessment the results of which though may be used for further detailed analysis and peer-review bases evaluation. Further research is aimed at overcoming the described constraints and searching for an optimized solution.

References

Abramo, G., & D'Angelo, C. A. (2015). The VQR, Italy's second national research assessment: Methodological failures and ranking distortions. Journal of the Association for Information Science and Technology, 66(11), 2202-2214.

Ancaiani, A., Anfossi, A. F., Barbara, A., Benedetto, S., Blasi, B., Carletti, V., ... & Costantini, M. (2015). Evaluating scientific research in Italy: The 2004–10 research evaluation exercise. Research Evaluation, 24(3), 242-255.

Barré, R., 2001. The Agora model of innovation systems: S&T indicators for a democratic knowledge society. Research Evaluation 10, 13–18.

Barré, R., 2005. S&T indicators for policy making in a changing science-society relationship, in: Handbook of Quantitative Science and Technology Research. Springer, 115–131.

Coccia, M., 2004. New models for measuring the R&D performance and identifying the productivity of public research institutes. R&D Management 34, 267–280.

Luwel, M., Noyons, C.M. and Moed, F., 1999. Bibliometric assessment of research performance in Flanders: policy background and implications. R&D Management, 29, 2, 33–141.

OECD. 2002. Frascati Manual: Proposed Standard Practice for Surveys on Research and Experimental Development. OECD, 67.

Senker, J., 2001. Changing organisation of public sector research in Europe – implications for benchmarking human resources in RTD. Paper prepared for Human resources in RTD session of the contribution of European socio-economic research to the benchmarking of RTD policies in Europe. Conference, Brussels, March 15–16.