

Changes in Methodology for Assessing Performance of Research Organisations and Influence of Such Changes on Researchers' Behaviour¹

Luboš Marek² | *University of Economics, Prague, Czech Republic*

Stanislava Hronová³ | *University of Economics, Prague, Czech Republic*

Richard Hindls⁴ | *University of Economics, Prague, Czech Republic*

Abstract

Assessing quality of research results on an international scale is a basis for evaluating the level of scientific activities pursued in research organisations. In the past 15 years, significant changes have occurred in the Czech Republic in research management and, in particular, the methodology of assessing research results. The methodology of assessment and its modifications should always be focused on increasing quality of research results; the rules of assessment have their effects on researchers' behaviour. This paper studies a question of whether the changes applied to the methodology of assessing research results in the Czech Republic have supported higher quality research results, i.e., results published in high-quality international journals. The authors have developed their own statistical test to measure significance of such changes, as well as other statistical tests of hypotheses. The main source is represented by the results of assessing public universities in the Czech Republic according to "Methodology for assessing results of research organisations" in 2010 and 2013. Our tests have not proven any statistically significant differences in the numbers of papers published in the journals monitored in the Web of Science and Scopus databases.

Keywords

Assessment methodology, test of significance of the changes, public universities

JEL code

C12, I20

¹ This article was processed with contributions from long-term institutional support of research activities by the Faculty of Informatics and Statistics, University of Economics, Prague.

² Faculty of Informatics and Statistics, Department of Statistics and Probability, W. Churchill Square 4, 130 67 Prague 3, Czech Republic. E-mail: marek@vse.cz.

³ Faculty of Informatics and Statistics, Department of Economic Statistics, W. Churchill Square 4, 130 67 Prague 3, Czech Republic. E-mail: hronova@vse.cz.

⁴ Faculty of Informatics and Statistics, Department of Statistics and Probability, W. Churchill Square 4, 130 67 Prague 3, Czech Republic. E-mail: hindls@vse.cz.

INTRODUCTION

Assessing research results undoubtedly has motivational effects. Emphasis on a higher quality of basic-research results should be manifested in a higher number of results passing internationally recognised criteria of evaluation, i.e., papers published in journals with nonzero impact factor values monitored by the Web of Science and Scopus databases. The motivational effects of assessment in research organisations in the Czech Republic have been widely discussed. A question arises: have qualitative changes in this methodology really led to a focus on better publication output in recent years?

1 ASSESSMENT OF R&D RESULTS IN THE CZECH REPUBLIC

Systemic evaluation of research results on the basis of strictly specified rules and procedures in the Czech Republic dates back to 2004. The reason for introducing the evaluation system based on a methodology approved by the government were the stagnating and, in certain fields, even decreasing numbers of R&D results in the Czech Republic while the expenses incurred on R&D from the state budget were increasing. The position of the Czech Republic in the international comparison was worsening. The goal was to motivate researchers to higher quality and quantity of research results via allocating the means provided to the research organisation from the state budget on the basis of the assessment results. The first "Methodology of assessment of R&D results" (hereinafter the "Methodology") was approved in 2004; the currently valid methodology is called Methodology 2017+.⁵

The basic general rules for the assessment (creating the database containing the information about the R&D results, definitions of the output types – papers, books, patents, applied results, evaluation for the five most recent years, and evaluation of R&D efficiency) were set out in the National R&D policy of the Czech Republic in the period of 2004–2008.⁶ As a future plan, the policy mentions a relationship between the assessment results and allocation of financial means to research organisations. This methodology has been a set of measures and tools to assess R&D results. The formulation was changed every year, which fact has always been criticised by research organisations. However, the changes were implied by the effort to rectify the most serious errors and shortcomings of the preceding version. Problems concerning the concept of the methodology for assessing the R&D results were fully manifested when the assessment results were applied according to Methodology 2008; it was the first time when part of the means from the state budget was allocated according to the assessment results.

1.1 Development of assessment methodology for research organisations in the Czech Republic

The evolution of the Methodology in the Czech Republic can, from the viewpoint of principles, be divided into four stages. The first stage (Methodology 2004 – Methodology 2009) represents the beginnings of the assessment principles (unfortunately, sometimes by trial-and-error) and the modification of the rules every year. The second stage came with Methodology 2010, whose validity was first approved for two years (2010 and 2011), and later extended to 2012. Methodology 2010 brought a number of modifications directed at respecting specific features of different fields, but it did not rectify the fundamental shortcoming of all previous methodologies, namely, the focus on quantity. The third stage is represented by Methodology 2013 (valid for the period of 2013–2016); it brought a fundamental change of combining bibliometric parameters with peer-review, and a different assessment for applied research results. The last stage, Methodology 2017+, should be a transition from evaluation of mere results to that

⁵ The official name "Methodology of assessment of R&D results" was valid in the period of 2004–2009. In 2010, the name was changed to "Methodology of assessing results of research organisations and results of completed projects". For the sake of brevity, we will use the simplified name Methodology, or Methodology with the year of validity, i.e., Methodology 2004, Methodology 2005, etc.

⁶ Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=5580>>.

of each research organisation as a whole (not only on the basis of the results). Ideological theses of this Methodology have been published, and its full implementation is expected in the years 2019 and 2020.

The principles of the first Methodology of assessment of R&D results⁷ (the so-called Methodology 2004) were very simple⁸ and completely insufficient for assessing the quality of scientific results (all types of results were valued identically, by one point). The results of this assessment should lead to classifying all research organisations into three categories according to efficiency of the means incurred on R&D (above-average, average, and below-average). Consequently, the allocation of financial means in future years should have been related to that classification. However, this stage was not implemented due to the lack of a criterion of efficiency and the disputed assignment of the same values to all types of output.

Methodology 2005⁹ (for assessing the results achieved in the period of 2000–2004) was just a more accurate update of Methodology 2004. It newly distinguished between different types of results – a paper in a journal with nonzero impact factor, a paper in another type of professional journal, a professional book, a chapter in a book, a contribution to conference proceedings, a patent, and an applied research result – and a higher number of points is always assigned to a publication in a world language. An index was set up for comparing the number of points assigned for the results achieved with the R&D means allocated to the given organisation from the state budget. On the basis of this index, all organisations were classified into four colour-coded categories. Such classification according to the efficiency level should have a positive/negative impact on the amount of means allocated from the state budget in future years. However, the results of this assessment turned out to be very disputable and a system for future allocation of financial means was not implemented.

In the introduction to Methodology 2006¹⁰ (for assessing the results achieved in the period of 2001–2005), it is said on page 2 that "applications of principles given in Methodologies 2004 and 2005 did not bring the expected effects and, despite the ever-increasing R&D expenses from the state budget, many fields of science in the Czech Republic lag behind even more". Hence "SR index" (a ratio between the number of points obtained for results and the amount of the R&D means allocated to the given organisation from the state budget)¹¹ was defined as an indicator of efficiency. Similar to Methodology 2005, research organisations were again classified into four colour-coded categories according to their efficiency levels. Another modification in this Methodology was concerned with increasing the number of types of applied research results and increase of their point valuation compared to basic research results. This approach was criticised and later led to an "inflation" of these types of results.

Methodology 2007¹² (for assessing the results achieved in the period of 2002–2006) was an update of Methodology 2006 and emphasis was again – even if disputably – put on the efficiency level expressed by the SR index. For the first time, this Methodology admitted verbal descriptions of the results and points to social sciences and humanities were assigned differently from other sciences.

Methodology 2008¹³ (for assessing the results achieved in the period of 2003–2007) brought many more changes. The SR index was abandoned with respect to results of research organisations (but remained for assessing results of completed programmes); and only humanities were set aside for assessment and social sciences were returned back to the other sciences. A group of Czech journals was defined so that only papers in those selected journals would pass for the assessment, and contributions

⁷ Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=18750>>.

⁸ Just for comparison: Methodology 2004 was a six-page text; Methodology 2013 was 59 pages.

⁹ Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=18751>>.

¹⁰ Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=21846>>.

¹¹ SR index = index of the state budget; state budget = "Státní Rozpočet" in the Czech language.

¹² Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=31543>>.

¹³ Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=503762>>.

to proceedings were restricted to ISI Proceedings (today CPCI). This Methodology was for the first time used to allocate the financial means for 2010 pursuant to the amended Act No. 130/2002 Coll. The intention to do so had already been announced in Methodology 2004; nevertheless, the direct calculations of financial R&D allocations from the state budget on the basis of obtained numbers of points a surge of disagreement came from research institutions.

Prior to approval of Methodology 2009¹ (for assessing the results achieved in the period of 2004–2008), there was a very sharp debate in the academic sphere about "what now" – the gradual improvements of the Methodology had not removed its basic shortcomings (motivation to quantity, not quality of results; no differentiation by fields; no peer-review; etc.). The resulting changes were, however, minor; e.g., a category of prestigious journals was introduced (Nature, Science) with a high assignment of points for results published in them.

Methodology 2010 and 2011² (for assessing the results achieved in the period of 2005–2009, or rather 2006–2010) under a new name of "Methodology of assessing results of research organisations and results of competed projects" tried to cope with the most glaring problems in the assessment process. That is why a chapter on allocation of financial means was, for the first time, included into the Methodology; in that chapter, an idea occurred that the means should be divided by fields and the points should be corrected with respect to the numbers of results. Results published in the journals monitored in the Scopus and ERIH databases were newly added to the results to be assessed.

Methodology 2012³ (for assessing the results achieved in the period of 2007–2011) was, in principle, an extension of Methodology 2010–2011. Only the chapter on the allocation of financial means on the basis of the assigned numbers of points was modified (made more specific).

Preparations of Methodology 2013⁴ had taken a lot of time; this Methodology introduced fundamentally different methods for result evaluation. In addition to bibliometric evaluation, exclusively applied to that date, peer-review evaluation of papers and books was to be applied, as well as evaluation of selected excellent results. Panels of reviewers were set up, in which experts from abroad also participated. The methods for assessing applied research results were also modified, but this concept was criticised. This evaluation process was originally planned to take place every year, which was too demanding; this fact led to delays and degradation of the originally good idea. This Methodology should have been valid for the period of 2013–2015, but it was later extended to 2016 (that is, results were evaluated for the periods 2008–2012, 2009–2013, 2010–2014, and 2011–2015).

Despite many year-to-year modifications, each Methodology was just a tool for calculating money from obtained points.⁵ A comprehensive system for assessing research had been and still is missing, which would view a research organisation regarding not only the results achieved, but also other aspects of activities pursued in R&D. This approach should be implemented in the new "Methodology for assessing research organisations and targeted-support programmes in research, development and innovations"⁶ (Methodology 2017+), whose roles are to be introduced in the period of 2017–2019; beginning 2020, the comprehensive assessment should be carried out in five-year cycles, not every year (the annual assessments turned out to be impossible to implement).

1.2 Motivational effects of Methodology

All the year-to-year modifications of the Methodology were motivated by the effort to improve the assessment of results and respond to criticism from research organisations. This criticism was mainly aimed

¹⁴ Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=532412>>.

¹⁵ Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=566918>>.

¹⁶ Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=650022>>.

¹⁷ Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=685899>>.

¹⁸ For this reason, this methodology is often called the "coffee grinder" in the Czech academic environment.

¹⁹ Cf. <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=799796>>.

at the lack of concept in creation of the Methodology (this lack was namely proved by the year-to-year changes), no regard to specific aspects of each field, subjective and erratic character of the point values assigned to individual results, and preferring quantity to quality. The last-mentioned aspect was the reason for the inflation of low-quality results and non-ethical behaviour of certain research organisations, which led to the necessity of sanctions for wrongly reported results.²⁰ The direct relationship between the assessment results and allocation of financial means was also criticised, because this relationship had negative impacts on management of some research organisations.

On the other hand, there was a positive effect of the mere fact that a methodology was created to implement the outcome of the discussion about possibilities in assessment of results achieved by research organisations. The awareness that research activities must be assessed was important. However, it is disputable whether the Methodology modifications always brought the expected impacts on increasing not only quantity, but also quality of research activities.

We asked whether qualitative changes in Methodology 2013 as compared with Methodology 2010 were reflected in a better quality of research results. For the purposes of this study, we deem high-quality results papers published in the Web-of-Science-monitored journals (denoted by Jwos) and the Scopus-monitored journals (denoted by Jsc). Papers published in such journals undergo an independent review process according to international standards and can, therefore, be viewed as a certain indicator of good quality of research activities.²¹ If Methodology 2013 was to bring a new approach to result evaluation and motivate researchers to focus on high-quality results, numbers of the Jwos and Jsc papers should be higher within assessment according to Methodology 2013 (for the period of 2008–2012) than those according to Methodology 2010 (for the period of 2005–2009).²² Even though it is clear that there is a two-year overlap, newer data could not be used due to the requirements for comparability of results – the assessments in 2014 and 2015 follow the principles of Methodology 2013, but only numbers of points assigned to the so-called assessment pillars are public, not the numbers of results by the type (papers, books, etc.). The results of the assessment which should have been made in 2016 are yet not available. Due to the incomparability of the assessment reports, neither older data (assessment according to Methodology 2009 and older) could be used.

2 METHODS OF ANALYSIS AND THE DATA USED

In order to verify the hypothesis that the changes in Methodology 2013 brought a fundamental change in quality, manifested by increased numbers of Jwos and Jsc results, we will apply our originally developed statistical test of significance of the changes, as well as standard hypothesis testing. The variables of interest are the numbers of papers published by twenty Czech public universities in the Web-of-Science-monitored (Jwos) and Scopus-monitored (Jsc) journals in two different periods of time and according to different methodologies for assessing research organisations' results – namely, M2010 (Methodology 2010, period of 2005–2009) and M2013 (Methodology 2013, period of 2008–2012). The source of the data was the R&D Information System of the Czech Republic.

To form a basic idea of the character of the data to be processed, we reviewed descriptive statistics, which may indicate some differences. The values of the descriptive statistics are shown in Table 1. The input data (numbers of articles Jwos and Jsc by universities) related to the partial calculations in the significance test are given in Table 1A in the Appendix.

²⁰ Nevertheless, this effort did not have the desired effect; the sanctions for incorrectly claimed results were only applied once.

²¹ We are aware that this assumption is not exactly true: it is clear that not all fields have papers in professional journals as their main output, and not all journals monitored by these databases are particular about the high professional quality.

²² Therefore our analysis takes in account only bibliometric data.

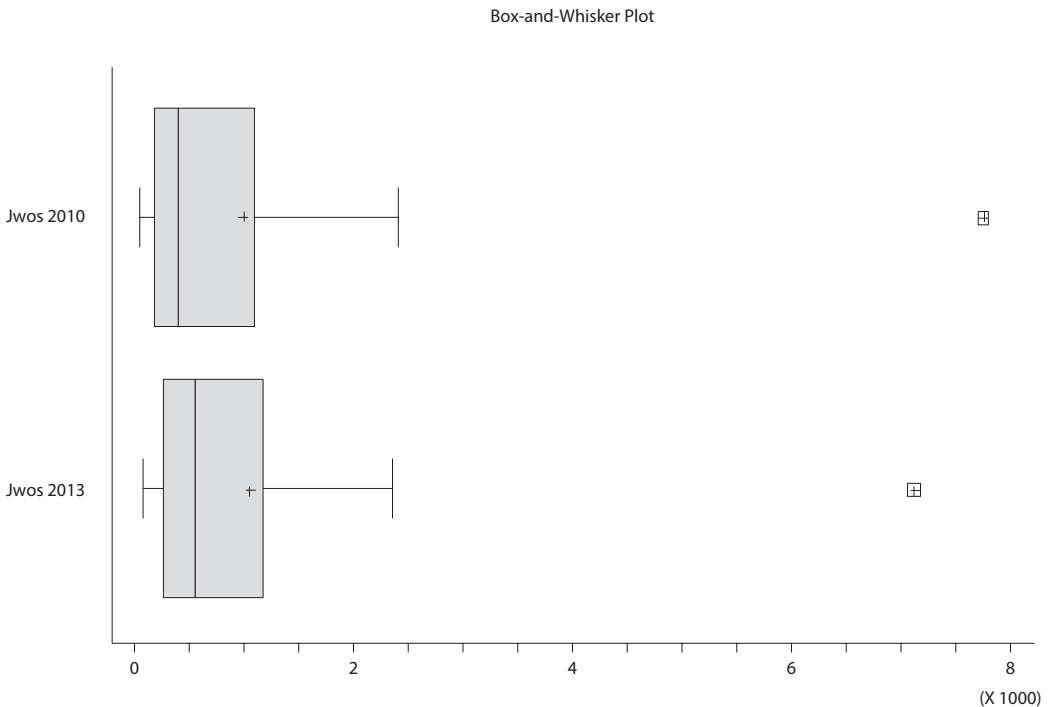
Table 1 Values of selected descriptive characteristics for Jwos and Jsc

	Jwos		Jsc	
	M2010	M2013	M2010	M2013
Average	982.3	1 052.5	445.9	529.3
Standard deviation	1 705.1	1 556.4	881.1	716.4
Coefficient of variation (%)	173.6	147.9	197.6	135.4
Minimum	30.0	60.0	48.0	58.0
Maximum	7 751.0	7 117.0	3 936.0	3 182.0

Source: <www.rvvi.cz>, authors' own results

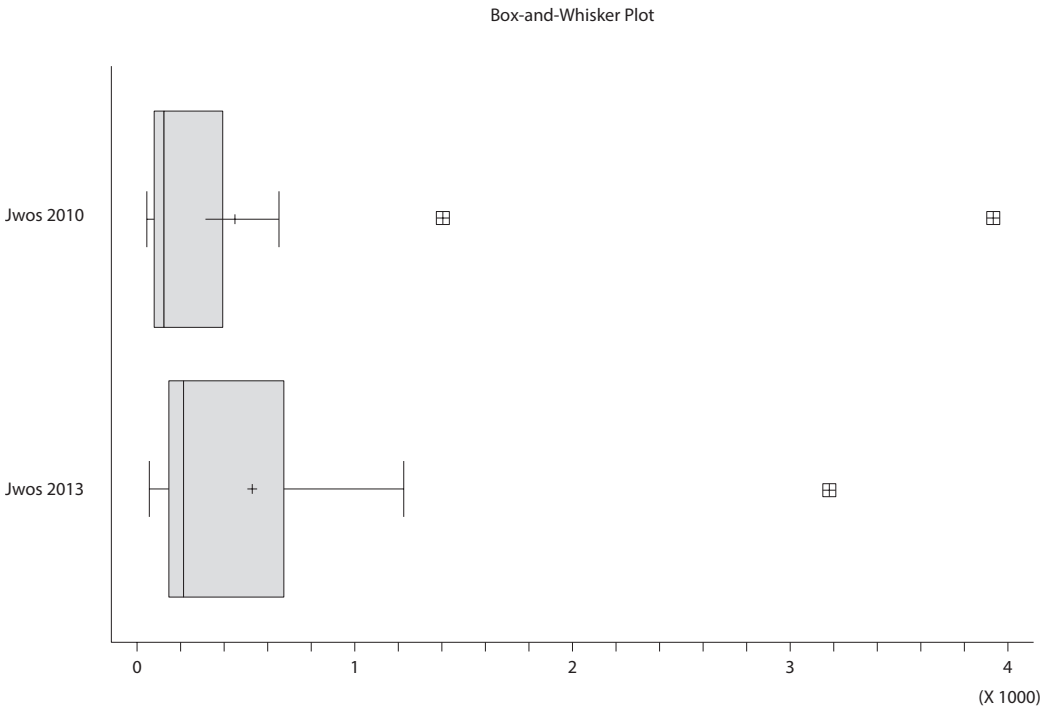
This preview is confirmed by box-and-whisker plots. Regarding numbers of papers published in the Web-of-Science-monitored journals, no substantial differences are observed between the 2005–2009 and 2008–2012 periods (cf. Figure 1).

Figure 1 Box-and-whisker plot for Jwos



Source: <www.rvvi.cz>, authors' own results

On the other hand, changes in both levels and variability values are clearly seen for the numbers of papers published in the Scopus-monitored journals (cf. Figure 2).

Figure 2 Box-and-whisker plot for Jsc

Source: <www.rwi.cz>, authors' own results

Selected statistical hypothesis tests will clarify whether the changes in the numbers of papers published in the Scopus-monitored journals (Jsc) outbalance the lack of such changes in the numbers of papers published in the Web-of-Science-monitored journals (Jwos), and whether, consequently, the overall changes in the numbers of papers published in internationally renowned journals (according to Methodology 2013 as compared with Methodology 2010) can be considered statistically significant.

2.1 Test for significance of changes

A test for measuring the significance of changes from one situation to another was proposed by two of the authors of the present paper (cf. Hindls and Hronova, 2007). It has turned out that this test is very well capable of identifying significance of changes from one situation in time to another. Here we try to establish a change in two variables (Jwos and Jsc publication numbers) for several units (public universities in the Czech Republic except for universities of arts) in two periods of time (according to M2010 and M2013). The hypothesis to be tested states that the numbers of the Jwos and Jsc publications were the same in both time periods of interest, while the alternative hypothesis denies the tested one in the sense that the numbers of the Jwos and Jsc publications in the period of 2008–2012 (i.e., according to M2013) was statistically significantly higher than in the period of 2005–2009 (i.e., according to M2010). An advantage of this test is the fact that, unlike the standard tests (cf. Section 2.2) it measures the overall significance of the changes in both variables at the same time.

Starting points and notation

Let us denote the first surveyed characteristic of two-criterion evaluation as x (number of articles Jwos), and the second one as y (number of articles Jsc). Further, we introduce the symbols 1 and 2 for the corresponding evaluation methodologies. We thus employ the following symbols:

- x_{1i} for the number of Jwos articles of i -th public university, $i = 1, 2, \dots, n$, (in this case $n = 20$) in the first period (according to M2010);
- x_{2i} for the number of Jwos articles of i -th public university, $i = 1, 2, \dots, n$, (in this case $n = 20$) in the second period (according to M2013);
- y_{1i} for the number of Jsc articles of i -th public university, $i = 1, 2, \dots, n$, (in this case $n = 20$) in the first period (according to M2010);
- y_{2i} for the number of Jsc articles of i -th public university, $i = 1, 2, \dots, n$, (in this case $n = 20$) in the second period (according to M2013).

The formulation for the test of significance of the changes in researchers' attitude over time using two-criterion evaluation

Let us denote:

- by K_1 the mean value of the aggregate two-criterion evaluation in the first period (the so-called *mean space localisation*); and
- by K_2 the mean value of the aggregate two-criterion evaluation in the second period (the so-called *mean space localisation*).

We test the null hypothesis H_0 about equality of the mean localisation in the space coordinate system, i.e.,

$$H_0: K_1 = K_2$$

against an alternative hypothesis:

$$H_1: K_1 < K_2$$

We will use the following statistic as the test criterion:

$$T = \frac{\sum_{i=1}^n k_i}{n} \sqrt{\frac{\sum_{i=1}^n \left[k_i - \frac{\sum_{i=1}^n k_i}{n} \right]^2}{n(n-1)}}$$

where:

$$k_i = \text{sign}\{y_{2i}^2 + x_{2i}^2 - y_{1i}^2 - x_{1i}^2\} \sqrt{(x_{2i} - x_{1i})^2 + (y_{2i} - y_{1i})^2},$$

$$\frac{\sum_{i=1}^n k_i}{n} \text{ is the point estimator of the statistic } K_2 - k_1.$$

It can be proved that this T statistic has Student's distribution $t [n - 1]$ under the validity of the tested hypothesis H_0 .

Comment: Using the sign {...} operator, the orientation of the aggregate "space" change (" \pm ") is determined for the level of the two-criterion value in the second (later) period in comparison with the first one. This operator thus expresses whether the i -th space localisation (i.e., the localisation of the i -th university) in the 2nd period (i.e., M2013) has moved nearer to (" $-$ ") or farther from (" $+$ ") the origin of coordinates $[0; 0]$, in comparison with the 1st period (i.e., M2010). For example, when the i -th space

localisation has moved farther from the centre, then the sign "+" expresses that the aggregate number (i.e., for both surveyed characteristics together) of the i -th university has been improved (it is a kind of a "geometric" summary of the surveyed characteristics "number of Jwos" and "number of Jsc"). The power of the test is sharply increased when not using the sign operator. The probability of the type II error β would be lower and the risk of the type I error α would be higher.

The critical region of the test W is defined by the following inequalities:

$$W = \{ T \geq t_{1-\alpha}[n-1] \text{ for left-sided alternative} \},$$

where: $t_{1-\alpha}[n-1]$ is the quantile of Student's t -distribution.

According to the result and using the significance level $\alpha = 0.05$, we draw a conclusion about the statistical significance of the time change in researchers' attitudes on the basis of results of the two-criterion evaluation.

The value of the test criterion, $T = 1.555$, does not exceed the critical level $t_{0.95}[19] = 2.093$. We can therefore observe that, at the 5% level of significance, the changes in the assessment methodology did not cause a statistically significant change in researchers' behaviour in the sense of their stronger focus on the results published in the Web-of-Science- and Scopus-monitored databases. The input data and partial calculations are shown in Table 1A.²³

2.2 Tests for equality of expectation values

We have applied additional tests to verify (or reject) hypotheses formulated in compliance with this particular problem. Namely, a parametric test for equality of averages, and the Mann-Whitney (Wilcoxon) median test have been carried out. Unlike our own test mentioned above, these standard tests only deal with equality of expectation values for each of the variables of interest, i.e., separately for Jwos and Jsc. All tests were carried out at the 5% significance level.

Let us first review the outcome of the t -test concerning the equality of averages.²⁴ Hypothesis H_0 states that the values of the average numbers of papers were the same in both time periods of interest; the alternative hypothesis, H_1 , denies H_0 in the sense that the average number of papers in the period of 2005–2009 (Methodology 2010) is smaller than that in the period of 2008–2012 (Methodology 2013).

Concerning the equality of averages for Jwos, the value of the test criterion is $t = -0.136$, and P -value = 0.446. Hence, the difference between the numbers of Jwos papers in the 2008–2012 and 2005–2009 periods cannot be viewed as statistically significant at the selected significance level. We have applied the test without knowing the variance values for the samples, but assuming that they are equal to each other. We have further tested this equality of variance values by F -test²⁵ (the value of the test criterion $F = 1.200$, and P -value = 0.695; the hypothesis that the variance values are equal to each other cannot be rejected at the selected significance level).

Similar conclusions has been made for Jsc, for which the value of the test criterion is $t = -0.328$, and P -value = 0.372. Hence, the difference between the numbers of Jsc papers in the 2008–2012 and 2005–2009 periods again cannot be taken for statistically significant at the selected significance level. We have applied the test without knowing the variance values for the samples, but assuming that they are equal to each other. We have again further tested this equality of variance values by F -test with the value of the test criterion $F = 1.512$, and P -value = 0.375; the hypothesis that the variance values are equal to each other cannot be rejected at the selected significance level).

²³ The results of this test for significance of changes are usually easy to display in graphical form; however, for the data processed here the graphical presentation would be unclear due to the necessity to display outlying observations. That is why we only present this tabular form of the results.

²⁴ Cf. e.g., Hindls et al. (2007).

²⁵ Cf. e.g., Hindls et al. (2007).

²⁶ Cf. Blatná (1996).

We have also applied a nonparametric test, namely, the Mann-Whitney (Wilcoxon) median test.²⁶ Hypothesis H_0 states that the medians of the numbers of papers were the same in both time periods of interest; the alternative hypothesis, H_1 , denies H_0 in the sense that the median of the number of papers in the period of 2005–2009 (Methodology 2010) is smaller than that in the period of 2008–2012 (Methodology 2013).

Concerning the equality of medians for J_{wos} , the value of the test criterion is $W = 224.0$, and P -value = 0.262. Hence, the difference between the numbers of Jsc papers in the 2008–2012 and 2005–2009 periods cannot be taken for statistically significant at the selected significance level.

The same test for Jsc comes to the same conclusions. The value of the test criterion is $W = 259.5$, and P -value = 0.055. The difference between the numbers of Jsc papers in the 2008–2012 and 2005–2009 periods again cannot be taken for statistically significant at the selected significance level.

CONCLUSION

Changes in the methodology for assessing results of research organisations have been present in the academic environment in the Czech Republic since 2004. The original effort to simply keep records of research results has been replaced with different forms of and rules for assessing the results. Importance of year-to-year changes in such rules was regarded just marginally in the beginning, but later such changes were viewed negatively by research organisations (in particular, public universities). Since 2009, financial means from the state budget for long-term conceptual development of research organisations have been allocated according to the assessment results. Never-ending changes in the methodology were motivated by an effort to respond to the quickly changing environment in research organisations (which very quickly adapted themselves to the methodology rules), certain negative phenomena occurring in applying the methodology to management of research organisations, and – of course – to justified criticism. Methodology 2013 was the latest version of the methodology according to which assessment of research organisations' results was actually carried out and completed. This version of methodology brought qualitatively new aspects in assessment of results achieved by both basic and applied research. It was the last version of the methodology which took that approach; the currently valid Methodology 2017+ views assessment of results (i.e., bibliometric assessment) as one of five modules to be applied within assessing activities of research organisations.

A question arose whether the important qualitative changes in Methodology 2013 as compared with Methodology 2010 were positively reflected in the behaviour of public universities, namely, whether they caused an increase in the numbers of papers published in internationally renowned journals, i.e., monitored by the Web of Science and Scopus databases. Our own test for significance of changes and other tests regarding the equality of levels were applied to verification of that assumption. None of the tests have proved that the changes in the assessment methodology would lead to statistically significant changes in researchers' behaviour in the sense of a stronger focus on results published in Web-of-Science- and Scopus-monitored journals in the period of 2008–2012 (Methodology 2013) as compared with 2005–2009 (Methodology 2010).

References

- ARNOLD, E. et al. *Metodika hodnocení ve výzkumu a vývoji a zásady financování [R&D Evaluation Methodology and Funding Principles]*. Prague: MŠMT, 2015.
- BLATNÁ, D. *Neparametrické metody [Non-parametric methods]*. Prague: University of Economic, 1996.
- HINDLS, R. AND HRONOVÁ, S. How Much Are Changes in Attitudes Significant over Time? In: *ISI 2007*, Lisbon: International Statistical Institute, 2007.

- HINDLS, R. AND HRONOVÁ, S. Odras ekonomického vývoje vybraných zemí ve struktuře výdajů na konečnou spotřebu [Reflection of Economic Development of Selected Countries in the Structure of Final Consumption Expenditure]. *Politická ekonomie*, 2012, Vol. 60, No. 4, pp. 425-442.
- HINDLS, R., HRONOVÁ, S., SEGER, J., FISHER, J. *Statistika pro ekonomy [Statistics for economists]*, 8th Ed. Prague: Professional Publishing, 2007.
- JURAJDA, Š., KOZUBEK S., MUNICH, D., ŠKODA, S. *Mezinárodní srovnání kvality publikačního výkonu vědních oborů v České republice [International comparison of publication performance of sciences in the Czech Republic]*. Prague: CERGE-EI, Studie 12/2015.
- MUNICH, D. AND ŠKODA, S. *Světové srovnání českých a slovenských časopisů podle indikátorů Impact Factor (IF) a Article Influence Score (AIS) [Worldwide comparison of Czech and Slovak journals according to the Impact Factor (IF) and Article Influence Score (AIS) indicators]*. Prague: CERGE-EI, Studie 19/2016.
- VANĚČEK, J., FAŤUN, M., PAZOUR, M. *Srovnávací studie vybraných metodik hodnocení výzkumu a vývoje [A comparative study of selected methodologies for assessment R&D]*. Prague: Technologické centrum AV ČR, 2008.
- Závěrečná zpráva mezinárodního auditu výzkumu, vývoje a inovací v České republice [Final report of international audit of research, development and innovations in the Czech Republic]*. Prague: MŠMT, 2012.

APPENDIX

Table 1A Number of articles and results

University	Jwos	Jwos	Jsc	Jsc	Jsc	x_{2i}^2	y_{1i}^2	y_{2i}^2	h_i	V_i	k_i	$(k_i - k_{avr})^2$		
	M2010	M2013	M2010	M2013	M2013									
	x_{1i}	x_{2i}	y_{1i}	y_{2i}	y_{3i}	x_{1i}^2	x_{2i}^2	y_{1i}^2	y_{2i}^2	y_{3i}^2	h_i	V_i	k_i	$(k_i - k_{avr})^2$
Czech Technical University, Prague	1 426	1 685	325	499		2 033 476	2 839 225	105 625	249 001		312	949 125	312	39 823
Czech Univ. of Life Sciences, Prague	469	646	350	515		219 961	417 316	122 500	265 225		242	340 080	242	16 774
USB, České Budějovice	827	948	189	206		683 929	898 704	35 721	42 436		122	221 490	122	95
Masaryk University, Brno	2 413	2 351	1 406	1 226		5 822 569	5 527 201	1 976 836	1 503 076		190	-769 128	-190	91 713
Mendel University, Brno	341	471	507	992		116 281	221 841	257 049	984 064		502	832 575	502	151 833
University of Ostrava	183	328	96	207		33 489	107 584	9 216	42 849		183	107 728	183	4 920
Silesian University, Opava	125	154	80	58		15 625	23 716	6 400	3 364		36	5 055	36	5 786
Technical University, Liberec	162	202	77	186		26 244	40 804	5 929	34 596		116	43 227	116	13
University of Pardubice	778	740	168	211		605 284	547 600	28 224	44 521		57	-41 387	-57	28 848
University of Hradec Králové	30	60	48	148		900	3 600	2 304	21 904		104	22 300	104	65
J. E. Purkyně University, Ústí nad Labem	103	147	56	67		10 609	21 609	3 136	4 489		45	12 353	45	4 504
Charles University, Prague	7 751	7 117	3 936	3 182		60 078 001	50 651 689	15 492 096	10 125 124		985	-14 793 284	-985	1 204 701
Palacký University, Olomouc	1 350	1 760	654	1 009		1 822 500	3 097 600	427 716	1 018 081		542	1 865 465	542	184 788
Tomáš Bata University, Zlín	249	374	61	291		62 001	139 876	3 721	84 681		262	158 835	262	22 293
University of VPS, Brno	553	549	109	108		305 809	301 401	11 881	11 664		4	-4 625	-4	13 592
Technical University, Ostrava	267	527	103	379		71 289	277 729	10 609	143 641		379	339 472	379	71 137
University of Economics, Prague	159	202	67	105		25 281	40 804	4 489	11 025		57	22 059	57	3 034
Univ. of Chemistry and Technology, Prague	1 410	1 386	125	143		1 988 100	1 920 996	15 625	20 449		30	-62 280	-30	20 296
University of Technology, Brno	702	941	438	835		492 804	885 481	191 844	697 225		463	898 058	463	123 149
West Bohemia University, Pilsen	348	462	123	219		121 104	213 444	15 129	47 961		149	125 172	149	1 338
Total	19 646	21 050	8 918	10 586		74 535 256	68 178 220	18 726 050	15 355 376		4 783	-9 727 710	2 249	1 988 702
$k_{0.95}(19)$	2 093													
T	1 555													