DE GRUYTER
OPEN

Research Paper

Minimum Representative Size in Comparing Research Performance of Universities: the Case of Medicine Faculties in Romania

Xiaoling Liu[1], Mihai Păunescu[2], Viorel Proteasa[3], Jinshan Wu[1†]

[1]School of Systems Science, Beijing Normal University, Beijing, 100875, P. R. China
[2]Department of Political Sciences, The National University of Political Studies and Public Administration, Bucharest, Romania
[3]Department of Political Sciences, The West University of Timişoara, Timişoara, Romania

Citation: Xiaoling Liu, Mihai Păunescu, Viorel Proteasa, Jinshan Wu (2018). Minimum representative size in comparing research performance of universities: the case of medicine faculties in Romania.

Vol. 3 No. 3, 2018
pp 32–42
DOI: 10.2478/jdis-2018-0013

Received: Jun. 27, 2018
Revised: Aug. 11, 2018
Accepted: Aug. 27, 2018

Abstract

Purpose: The main goal of this study is to provide reliable comparison of performance in higher education. In this respect, we use scientometric measures associated with faculties of medicine in the six health studies universities in Romania.

Design/methodology/approach: The method to estimate the minimum necessary size, proposed in in Shen et al. (2017), is applied in this article. We collected data from the Scopus data-base for the academics of the departments of medicine within the six health studies universities in Romania during the 2009 to 2014. And two kind of statistic treatments based on that method are implemented, pair-wise comparison and one-to-the-rest comparison. All the results of these comparisons are shown.

Findings: According to the results: We deem that Cluj and Tg. Mureş have the superior and inferior performance respectively, since their reasonably small value of the minimum representative size, in either of the kinds of comparison, whichever indexes of citations, *h*-index, or *g*-index is used. we can not reliably distinguish differences among the rest of the faculties, since the quite large value of their minimum representative size.

Research limitations: There is only six faculties of medicine in health studies universities in Romania are analyzed.

Practical implications: Our methods of comparison play an important role in ranking data sets associated with different collective units, such as faculties, universities, institutions, based on some aggregate scores like mean and totality.

Originality/value: We applied the minimum representative size to a new emprical context—that of the departments of medicine in the health studies universities in Romania.

Keywords Research evaluation; Minimum representative size; Bootstrap sampling; Medicine departments

† Corresponding author: Jinshan Wu (Email: jinshanw@bnu.edu.cn).

1 Introduction

The evaluation of departments or universities has become common place for nowadays academia. Different indicators are used for such purposes, while research is emphasized in many cases (Hazelkorn, 2011). Scientific productivity can be measured using different indicators, such as the number of published articles, the number of received citations, *h*-index (Egghe, 2008; Hirsch, 2005; Molinari & Molinari, 2008), *g*-index (Egghe, 2006; Tol, 2008). Whole departments, higher education institutions or even cities and countries are assessed in these terms. In practice, often data are aggregated and compared using the mean value or the total value of the aggregated sets. For instance, Academic Ranking of World Universities (ARWU, 2018), QS World University Rankings (QS, 2016), and CWTS Leiden Ranking (CWTS, 2018) all use either the mean or total value type of a specific indicator. Such approaches can be criticized for reliability, as the arithmetic mean may not be an appropriate gauge for the performance of a collective set, when the sample values in the set display a highly skewed distribution. When the distribution of the data set deviates significantly from a normal function the samples are no longer within a narrow region centered at the arithmetic mean. Often for a set of papers in a journal, a set of researchers in a university, it is often the case that the distribution is skewed (Lotka, 1926; Seglen, 1992). In such a case, the high variation within the populations to be compared obscures the inter-population variation, thus rendering the comparison often unreliable.

In Shen et al. (2017), we asked when the arithmetic mean of a sample set, or some other kind of average score, be used to represent the set, and under which conditions a comparison based on such measures of central tendency can be reliable? When does the comparison of arithmetic means indicate a grouping of academics, such as a department or a university, performs better than the one it is compared to? In order to answer this question, we proposed a definition of the minimum representative size κ as a parameter which characterizes a pair of data sets which are to be compared. The method is described in details, including the analytical demonstration, in Shen et al. (2017). In a nut-shell, κ represents the size of the smallest sample of randomly extracted points within the data set whose variance of averages is less than or equal to the variance between the data sets to be compared.

Given, for example, two sets of data, (1) and (2), whose averages are in an ordinal relation e.g. the average of data set (1) is higher than that of data set (2), what is the probability that a random sample from the first data set has a higher average than a random sample from the second? If the two data sets are skewed, the probability is often not high. For the very purpose of increasing this probability when it is possible to increase it, we introduced in Shen et al. (2017) the definition

of κ and instead of drawing one random sample, we turn to draw κ_i random samples from the corresponding set i and compare the average of those κ_1 and κ_2 samples. The value of κ depends on the variance of the data set: the smaller the variance, the smaller κ. If κ is comparable to the size of the data set, or even larger—and this happens when the set has very large variance compared to other sets, then the average is totally an unreliable indicator for comparing the two data sets. Therefore, the minimum representative size κ serves as an indicator of consistence of the set, which can be used as a supplementary indicator to be computed before using the average as a basis for comparing two data sets with a skewed distribution which deviates significantly from normality.

An alternative approach is to truncate the distribution and to consider a particular segment when comparing a set of populations. That particular segment has the property of smaller variation and thus renders the comparison between populations more reliable. In this case, theoretical arguments for choosing a specific segment as being representative of the entire population are strongly required. We pursued this stream of research in Proteasa et al. (2017).

In this article, we apply the former analytical approach in an empirical context: we calculate the minimum representative size for six medicine departments in Romania in order to allow for reliable comparisons between them as collective units.

2 The Method

Let us denote a skewed data set j, which may be total number of papers, received citations, h-index, g-index of each researchers in a university, as $\{s_j\}$. Due to the skewness of the distribution of $\{s_j\}$, the mean $S_j = \langle s_j \rangle = \frac{\Sigma_p s_j^p}{\Sigma_p}$ does not display the properties which are typical for a normal distribution. For example, an academic from a university with such a distribution has a higher probability to have an individual score s_j which is at a significant distance from the average of the data set S_j. Such skewed data sets display also higher frequencies of extremely large moments of the distribution function and a higher degree of overlap when compared in pairs. Thus:

$$\Pr\left(s_i > s_j \mid S_i > S_j\right) = \int_{-\infty}^{\infty} dx_j \left(\int_{x_j}^{\infty} dx_i \rho_i\left(x_i\right) \rho_j\left(x_j\right) \right) << 1, \quad (1)$$

where $\rho_i(x_i)$ is the distribution with mean S_i in data set i. This means that the probability of a random sample from the data set with a higher mean to have a higher value than that from the data set with a lower mean can be very low. However, when moving from an individual sample to large enough K_i and K_j samples from

data sets i and j respectively, the odds change significantly and the probability that a random sample of K_i researchers from university i has higher mean score than a random sample of K_j random researchers from university j, is high, as indicated in the equation below:

$$\Pr\left(G_i\left(K_i\right) > G_j\left(K_j\right) \mid S_i > S_j\right) >> 0 \approx 1, \tag{2}$$

where, $G_i(K_i)$ is the mean of the K_i samples from the data set i,

$$G_i\left(K_i\right) = \frac{1}{K_i}\sum_{p=1}^{K_i} x_i^p \tag{3}$$

We also know that the average of $G(K)$, $\mu(G(K))$ is close to S (the mean of data set), and the variance of $G(K)$ decreases with K.

Equation (2) can be solved by bootstrap sampling (Wasserman, 2004) (see Shen et al. (2017) for further details). For data sets i and j with $S_i > S_j$, starting from $K_i = 1$ and $K_j = 1$, bootstrap sampling unfolds as follows:

1. For given values of K_i and K_j, L=4000 sets of random samples are generated from data sets i and j, respectively. Following this, $G_i^l\left(K_i\right)$ and $G_j^l(K_j)$ are calculated for each l=1, 2,…, L.
2. Pair matching is performed on $G_i^m\left(K_i\right)$ and $G_j^n(K_j)$ and the percentage of $G_i^m\left(K_i\right) > G_j^n(K_j)$ is calculated. The following conditions are imposed:

$$\begin{cases} \Pr\left(G_i\left(K_i\right) > G_j\left(K_j\right) \mid S_i > S_j\right) \geq 0.9 \\ \dfrac{\sigma_i^2}{\kappa_i} = \dfrac{\sigma_j^2}{\kappa_j} \end{cases}, \tag{4}$$

3. If (4) conditions are satisfied, then the pairs values $(\kappa_i, \kappa_j)=(K_i, K_j)$ represent what Shen et al. (2017) termed ***the minimum representative size*** of the pair of data sets i and j. If (4) conditions are not satisfied, then K_i or K_j are increased and he sequence is repeated starting with step 1.

If κ_i is less than the size of data set, i.e. $\kappa_i < N_i$ and $\kappa_j < N_j$, then we may say the averages $G_i(\kappa_i)$ and $G_j(\kappa_j)$ can be reliably compared. The smaller κ, the more consistent the distribution within the data set. Values of κ_i which are greater than N_i indicate that the average is not a reliable measure to describe the central tendency of the data set i.

We increase K_i and K_j according to the ratio between their variances, $\frac{K_i}{K_j} = \frac{\sigma_i^2}{\sigma_j^2}$, instead of making $K_i = K_j$ since we believe that the set with larger variance should be responsible to reduce its variance more by using larger sample size.

3 Results

3.1 Data

In this article, we study 3374 academics from the departments of medicine within the six health studies universities in Romania: Cluj, Bucharest, Timişoara, Iaşi, Craiova and Tg. Mureş.

The personnel lists were compiled from public sources: websites and reports in 2014. We collected publication data from the Scopus data-base for the population of academics we established. We collected information regarding publications, citations, and Hirsch's *h*-index. We included single and co-authored papers, as well as their respective citations for a five year interval: from 2009 to 2014. We excluded authors' self-citations. The data were collected between November 2014 and May 2015. Full counting was used. We disambiguated manually authors' name and affiliation. We included all papers from authors with multiple affiliation, provided they were published in the 2009–2014 time window. We included in the population all the academics which were considered to belong to health disciplines, according to an official categorization of teaching personnel in health studies (MS, 2009). We excluded only a small minority of the population we compiled e.g. English teachers employed in medicine departments. We identified a set of publication with an outstanding number of citations whose character is different from a standard research article: guidelines, medical procedure recommendations, and definitions. We excluded these articles as well, based on a systematic word search. We then recalculated the number of received citations, the *h*-index and the *g*-index for each academic in the data base. The data base was used previously for the analyses presented in Proteasa et al. (2017). A description of the populations we study is presented in Table 1.

Table 1. **Basic statistics.** For each university, its name, number of academics(N), mean score($\langle\bullet\rangle$), standard variance (σ) of the corresponding index, are shown

University	N	Citation		*h*-index		*g*-index	
		$\langle c\rangle$	σ	$\langle h\rangle$	σ	$\langle g\rangle$	σ
Cluj	596	24.75	124.66	1.69	2.23	2.53	4.03
Bucharest	1119	18.41	80.81	1.45	1.99	2.15	3.56
Timişoara	505	15.93	57.61	1.42	2.03	1.93	3.16
Iaşi	547	15.86	75.22	1.54	1.90	2.06	3.08
Craiova	280	14.19	40.95	1.50	1.91	1.98	2.90
Tg. Mureş	327	5.40	22.19	0.83	1.28	1.08	2.04

We illustrate the skewness of the six distribution in Fig.1, where we plotted the distributions of total citations of each of the academics affiliated to the six universities. A visual inspection of the plots reveals that the first quartile and minimum citation

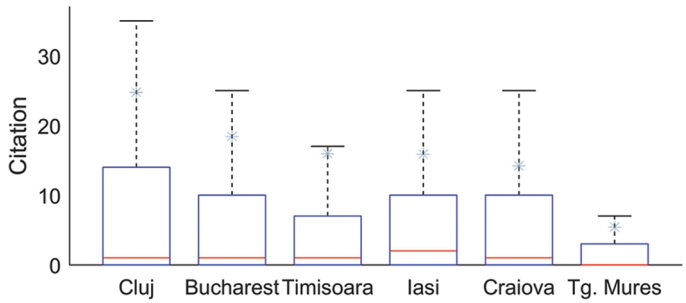

Figure 1. **Box-plot for the citation distribution.** The citation distributions of six universities are shown in box-plot. The five horizontal lines in each box represent the maximum, third quartile, median, first quartile and minimum citation (the last two lines can not be seen in these figures since they are to close to zero.), respectively. The star markers represent the mean of the citations. There is a huge difference between mean and median in each distribution.

of the six universities are all close to zero thus cannot be seen in the figure, since a considerable number of academics received no citations in the time window during which citations were collected. The six distributions are skewed and present a high degree of overlapping, as indicated by their large standard variance and by the large difference between medians and means.

In the following section we will engage with two research questions, conceptualized in the section dedicated to the outlining the method: (1) Is the mean a reliable measure of central tendency for the purpose of establishing a hierarchy of the six medical schools, which can account for the quality of the academics affiliated with them? (2) How can the six medical schools be compared using the minimum representative size?

3.2 Pairwise comparison

A first statistic treatment we perform includes pairwise comparison of the six medicine faculties. In this respect, we used distributions of citations(c), values of the h-index(h) and, respectively, the g-index(g). Thus, we calculate the minimum representative size, i.e. the minimum size of a sample of representative academics, κ for each pair of medicine faculties, according to Eq.4. We use notation $\kappa(I)_c^r$ to denote the minimum number of representative academics of faculty \$r\$ when compared with faculty c, based on the index I. For example, κ for Tg. Mureş and Cluj on g-index is $\kappa(g)_{CL}^{MU} = 5$ and $\kappa(g)_{MU}^{CL} = 22$ respectively. The considerable difference of mean g-index between these two universities may be the main reason of the small values of the pair of κ. $\kappa(g)_{MU}^{CL} = 22$ is larger partially due to the relatively higher heterogeneity of the g-index distribution of Cluj. The value of $\kappa(I)_c^r$ is determined by the overlap between distributions r and c and also by the variances of r and c.

Research Paper

The distributions of the Bootstrap average $G(\kappa)$, are shown in Fig.2(b). As a comparison, the distributions of individual g-index are shown in Fig.2(a), where there is a larger overlap between the two universities though we can clearly see the difference between their average scores. In Fig.2(b) the overlap is visibly smaller: the standard deviations are much smaller for $G(\kappa)$, while mean values did not change.

The complete results of pairwise comparison based on the three indexes, citation, *h*-index, and *g*-index are shown in Fig.3(a), (b), and (c), respectively. The size of each circle at position (r, c) is the ratio of mean value of the corresponding index I for university *r* and *c*, i.e. $\frac{s_r^I}{s_c^I}$. Their magnitude can be assessed against the gray circles along the diagonal, which serve as references and amount to size 1. The color of the other circles represents the value of $\kappa(I)_c^r$. The darker the color, the greater the value of κ. For example, the size of circle at the upper right corner of Fig.3(c) corresponds to $\frac{s_{CL}^g}{s_{MU}^g} = 2.34$ and the color of the same circle corresponds to $\kappa(g)_{MU}^{CL} = 22$. The size of the circle at the lower left corner of Fig.3(c) corresponds to $\frac{s_{MU}^g}{s_{CL}^g} = 0.43$, while the color of the same circle corresponds to $\kappa(g)_{CL}^{MU} = 5$.

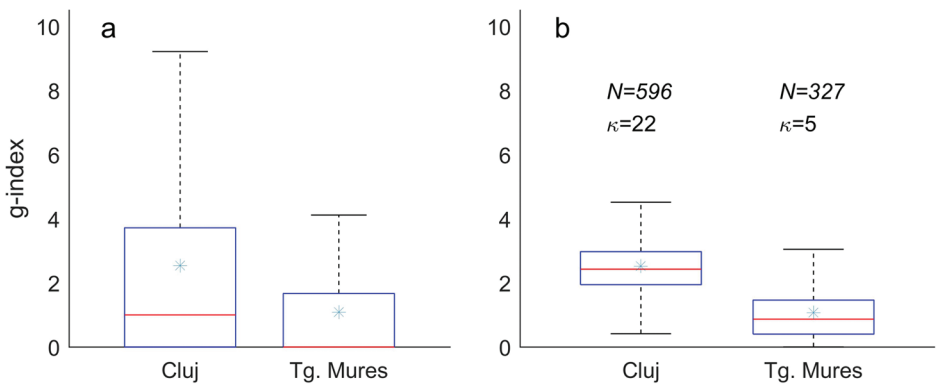

Figure 2. **Box-plot for the g-index distribution of Cluj and Tg. Mureş**. (a) The original distribution of Cluj and Tg. Mureş. The five horizontal lines in each box represent the maximum, third quartile, median, first quartile and minimum *g*-index, respectively. The star markers represent the mean *g*-index. (b) Distribution of the Bootstrap *κ*-sample average of *κ*-index, with the minimum representative sizes of the two sets when the two sets are pair-wisely compared.

A visual inspection of Fig.3 reveals that most of the circles corresponding to pairwise comparisons between faculties are leaning towards the dark ends of the spectra. We interpret this observation as proof of the incapacity of the mean scores

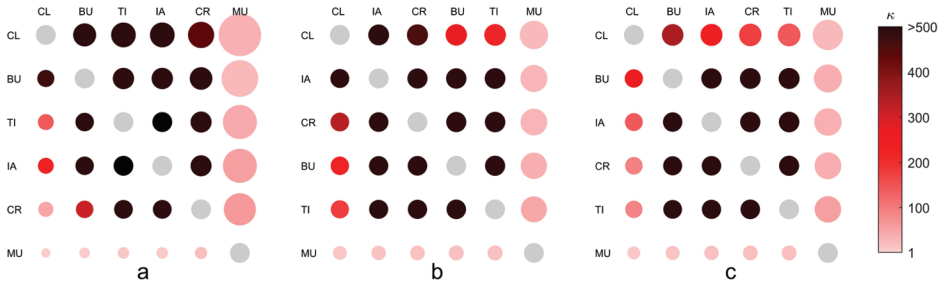

Figure 3. **Heatmap of the pairwise compared.** (a) Faculties are sorted in descending order by average citations. The darkness of each circle represents the value of $\kappa(I)_c^r$ for the pair formed by the faculty on row *r* and the one on column *c*. The size of the circles represents the ratio of the mean citation for the faculties corresponding to row *r* and column *c*. The gray circles on the diagonal are reference circles with size 1. (b) Results are calculated based on *h*-index. (c) Results are calculated based on *g*-index.

to account for the differences between the six faculties, when compared in pairs. More than this, some of the values κ takes are greater than size of the faculty, e.g. $\kappa(c)_{IA}^{BU} = 3188$, $\kappa(c)_{BU}^{IA} = 2762$, but $N_{BU} = 1119$ and $N_{IA} = 547$. In this particular case, the comparison of the two faculties based on their average is unreliable.

3.3 Group comparison

In a second statistical treatment, we compare one faculty *u* against the rest of the faculties, taken as a whole i.e. the reunion of their populations of academics. In other words, we compare the academics within one faculty with the rest of the academics as if they are from a single virtual faculty *rest*. This approach allows us to perform pairwise comparisons as we did in the previous sub-section, between faculty *u* and *rest* in order to compute the values of $\kappa(I)_{rest}^{u}$, for each *I* index. The results we obtained are presented in Table 2, below. The values of $\kappa(I)_{u}^{rest}$ are not included in the table, as they do not serve for the description of the social phenomenon we study.

The faculty from Cluj has the highest mean score in all the comparisons we performed, regardless of the index. The corresponding values of the minimum representative size, κ_{rest}^{CL}, are not small compared to the size of the population, but remain lower than it. At the other extreme, the faculty based in Tg. Mureş has the lowest mean score and smallest values of κ_{rest}^{MU} in all the comparisons. For the other four universities, the values of κ_{rest} are all exceeding their number of academics. We interpret these results as proof of the fact that it is reliable to compare the faculty based in Cluj against the rest of the faculties, and, in fact, it exhibits superior values of all the three indexes we used. At the same time, it is also reliable to argue that the faculty from Tg. Mureş exhibits inferior performance compared to the rest of

Table 2. **Values of κ_{rest}.** For each university, its name, mean score of the *rest* university ($\langle c\rangle_{rest}$, $\langle h\rangle_{rest}$, $\langle g\rangle_{rest}$), and minimum number of representative academics (κ_{rest}), calculated by the corresponding index, are shown.

University	N	Citation		*h*-index		*g*-index	
		$\langle c\rangle_{rest}$,	σ	$\langle h\rangle_{rest}$	σ	$\langle g\rangle_{rest}$	σ
Cluj	596	15.50	480	1.40	183	1.95	145
Bucharest	1119	16:50	5283	1:44	$> 10^4$	2.00	1664
Timişoara	505	17:35	4900	1.45	9066	2.08	1378
Iaşi	547	17:38	7187	1.43	893	2.06	$> 10^4$
Craiova	280	17:40	477	1.42	3375	2.06	3634
Tg. Mureş	327	18:39	2	1.51	12	2.16	13

the faculties, under all the three indexes we used as a basis of comparison. Last but not least, our results from both the pair-wise comparison and the one-to-the-rest comparison indicate that it is not reliable to distinguish differences in the performance of the other four faculties (Bucharest, Timişoara, Iaşi, and Craiova), irrespective of the index which is used as a basis for calculation.

4 Summary

In a nut-shell, this contribution consists in applying the minimum representative size, a methodology developed in Shen et al. (2017), to a new empirical context—that of the faculties of medicine in the health studies universities in Romania, previously studied by Proteasa et al. (2017). The "quality" of the academics affiliated to the six faculties located in Cluj, Bucharest, Timişoara, Iaşi, Craiova, and Tg. Mureş is measured by the total citations received by each academic, and the respective values of the *h*-index and *g*-index. We performed pair-wise comparison and one-to-the-rest comparison. We found that, when the population of academics from Cluj is compared to the others, in either of the two methods of comparison, the minimum representative size is reasonably small. We interpret this finding as a reliable indication of superior performance, in relation to all three indexes used in this article. We also find that when the population of academics affiliated to Tg. Mureş is compared to the others, in either of the two methods of comparison, the minimum representative size is quite small, thus it is reliable to say that its performance is inferior to the other faculties, on the three measures used in this work. For the rest of the faculties we investigated in this work, we cannot reliably distinguish differences among them, since their minimum representative size are all quite large, sometimes even bigger than their own sizes.

One might think that these results which substantiate that the faculties located in Cluj and Tg. Mureş are quite different from the rest, while the others are rather similar is trivial. It can be argued that a similar conclusion can be reached by a

simple comparison of the mean scores in Table 1. We emphasize that the method we unfolded in this article which builds on the concept of the minimum representative size (κ), especially the relation between κ and the size of the whole population N, represents a validation, in this case, of the falsifiable hypothesis which can be derived from a simple comparison of the averages. When $\kappa \ll N$ the hypothesis is validated and such a comparison is reasonable, while when $\kappa \sim N$ or even $\kappa > N$, then the hypothesis is invalidated, and the hierarchy of the averages represents more a numerical artifact, than a substantial property of the distributions, thus such a comparison is unreliable. That case is exemplified through the four faculties whose corresponding minimum representative size exceeded the size of their distributions. To conclude, the minimum representative size relative to the size of the population proved to be a useful and reliable ancillary indicator to the mean scores.

We consider our findings are particularly relevant in situations when aggregate scores are computed for the purpose of ranking data sets associated with different collective units, such as faculties, universities, journals etc. Whenever one wants to distinguish the performance of two collective units, the minimum representative size of pair-wise comparison should be calculated first as an indication of the reliability of the comparison of the means. When κ is small compared to size of the set, then the mean can be seen as a good representative value of a Bootstrapped κ number of samples from the set. Whenever one is interested in comparing one particular collective unit against a group of similar units, such as the medicine faculties in health studies universities in Romania, then κ calculated by one-to-the-rest comparison would be appropriate. In both cases, those groups whose κ are comparable or even bigger than their group sizes should be discarded from such comparisons. A small κ is an indication of the consistency of a data set, which is an attribute that, we consider, should be assessed before ranking collective units consisting from individuals with different performance levels.

Reference

ARWU (2018). Ranking Methodology of Academic Ranking of World Universities—2017. Retrieved from: http://www.shanghairanking.com/ARWU-Methodology-2017.html.

CWTS (2018). CWTS Leiden Ranking. Retrieved from: http://www.leidenranking.com/information/indicators.

Egghe, L. (2006). Theory and practise of the g-index. Scientometrics, 69, 131–152.

Egghe, L. (2008). Modelling successive h-indices. Scientometrics, 77, 377–387.

Hazelkorn, E. (2011). Rankings and the reshaping of higher education: the battle for world-class excellence. Houndmills Basingstoke Hampshire; New York: Palgrave Macmillan.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102, 16569–16572.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences, 16, 317–323.

Molinari, J. F., & Molinari, A. (2008). A new methodology for ranking scientific institutions. Scientometrics. Retrieved from: https://akademiai.com/doi/abs/10.1007/s11192-007-1853-2.

MS (2009). Ordin nr. 1509/2008 privind aprobarea Nomenclatorului de specialităţi medicale, medico-dentare şi farmaceutice pentru reeauă de asistenă medicală. Retrieved from: https://www.eshg.org/fileadmin/www.eshg.org/documents/countries/Roman.

Proteasa, V., Păunescu, M., & Miroiu, A. (2017). An extension of the characteristics scales and scores: Isolating interinstitutional from intra-institutional variance in highly skewed real population distributions. In 16th International Conference on Scientometrics & Informetrics, ISSI 2017 (pp. 1192–1203).

QS (2016). QS World University Rankings. Retrieved from: https://www.topuniversities.com/qs-world-university-rankings/methodol.

Seglen, P. O. (1992). The skewness of science. Journal of the American Society for Information Science; New York, N.Y., 43.

Shen, Z., Yang, L., Di, Z., & Wu, J. (2017). How large is large enough? In 16th International Conference on Scientometrics & Informetrics, ISSI 2017 (pp.302–313).

Tol, R. S. J. (2008). A rational, successive *g*-index applied to economics departments in Ireland. Journal of Informetrics, 2, 149–155.

Wasserman, L. (2004). All of Statistics. Springer New York.