



# Nondeterministic ranking of university departments<sup>☆</sup>



Andrea Bonaccorsi<sup>a</sup>, Tindaro Cicero<sup>b,\*</sup>

<sup>a</sup> DESTEC, School of Engineering University of Pisa Largo, Lucio Lazzarino 2, 56125 Pisa, Italy

<sup>b</sup> ANVUR Italian Agency for the Evaluation of Universities and Research Institutes, Via Ippolito Nievo 35, 00153 Rome, Italy

## ARTICLE INFO

### Article history:

Received 3 October 2014

Received in revised form 8 January 2016

Accepted 10 January 2016

### Keywords:

Departments ranking

Ranking

Research quality

National research assessment

## ABSTRACT

Rankings in higher education are largely used to summarize a huge amount of information into easily understandable numbers. They are also used by governments in order to allocate funding. Nevertheless, they are often criticized. One stream of criticism refers to the fact that rankings build up an ordinal order by considering only the mean of the distribution of indicators and not their variability. Using the micro-data from the Italian evaluation of the quality of research (VQR, Valutazione della Qualità della Ricerca), we examine whether difference in performance between departments with different position in the ranking are distinguishable from random effects. We obtain a robust clustering of departments in a limited number of groups. The number of groups is in the range 3–7, while in most cases it is 4–6. The implications of these findings for evaluation and research policy are explored.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

Rankings of universities or university departments are largely used by the media, public opinion, and governments in order to examine the position of individual institutions or raise issues about the national performance. They are considered useful because they summarize a huge amount of information into easily understandable numbers (Hazelkorn, 2011; O'Connell, 2013). Rankings are however heavily criticized by scholars in social sciences, ranging from statistics to sociology and political sciences.

One stream of criticism refers to the fact that rankings build up an ordinal order by considering only the *mean* of the distribution of indicators, disregarding other moments of the distribution. Whatever the metrics adopted, each university or department receives a score that is the result of the activity of all its members, or at least its active part. However, the distribution of results of all members is entirely lost and only a single score enters into the composite indicator, hence the final ranking.

We follow this line of criticism by examining the relation between individual scores of quality of research and average scores assigned to departments. We make use of micro-data from the largest evaluation exercise ever carried out, i.e. the evaluation of the quality of research (in the rest of the paper: VQR, Valutazione della Qualità della Ricerca), completed in July 2013 by the Italian Agency for the Evaluation of Universities and Research Institutes (ANVUR).

VQR has produced a ranking of departments and a ranking of universities, following a mandatory provision of the national Italian legislation. At the same time, it has published all data aggregated at the level of departments and scientific areas,

<sup>☆</sup> *Disclaimer:* This paper was prepared or accomplished by Tindaro Cicero and Andrea Bonaccorsi in their personal capacity. The opinions expressed in this article are the authors' own and do not reflect the view of the ANVUR.

\* Corresponding author. Tel.: +39 06 58317321.

E-mail address: [Tindaro.cicero@anvur.it](mailto:Tindaro.cicero@anvur.it) (T. Cicero).

allowing actors in the research system to reconstruct possibly different rankings, for example by changing the criteria or weights of aggregation. It has recognized that, while the raw indicators associated to individual publications should be considered reliable, rankings of departments depend crucially on the definitions of the perimeter of comparison and on the treatment of the heterogeneity within the department. We bring this recognition a step further, by using anonymized micro-data on *individual* publications of *individual* researchers, and examining whether departments rankings are statistically robust.

The paper is organized as follows. Section 2 introduces the VQR and explains the source of data. Section 3 review the current debate on university rankings. Section 4 introduces the statistical approach to rankings, as discussed by Lubrano (2009). Section 5 discusses the potential and limitations of the method suggested by Lubrano with respect to the Italian data and applies the statistical analysis of rankings. Section 6 discusses the results while Section 7 concludes.

## 2. The ranking of departments in the Italian evaluation of quality of research (VQR)

The VQR is the largest research assessment exercise ever carried out. The results of this exercise have been published (in Italian) in July 2013 and are currently under translation into English for a wider circulation. In 2015 a new exercise, covering the period 2011–2014, has been launched, confirming that research assessment is a permanent feature of the institutional landscape.

Following a Ministry of Research decree in 2011, VQR has been managed by the newly created Agency for the Evaluation of Universities and Research Institutes (ANVUR), an independent public institution in charge of research evaluation and higher education quality assurance. VQR has covered the period 2004–2010, involving all researchers of all Italian universities (including private universities and distance universities) and Public Research Organizations (PROs). Other research organizations not subject to evaluation by law have also applied for voluntary evaluation.

Researchers at universities have been asked to submit three research products; researchers at PROs have submitted six products. Research products include publications (articles, books, chapters, proceedings, critical editions, translations and commentaries) but also patents, design, drawings, performances, exhibitions, artifacts, prototypes, databases, software, and maps. 95 universities, 12 PROs and 26 voluntary research consortia and other research organizations have been submitted to evaluation.

VQR has followed a common set of rules that have been dictated by the Ministry of Research. Research products should be evaluated in terms of their relevance (value added for the advancement of knowledge and of science in general), originality/innovativeness (contribution to the state of the art in the field) and internationalization. While the metrics for evaluation could differ across disciplines, all products should be eventually rated into four merit classes (A, B, C and D), which are defined in terms of the position in the overall distribution of merit, as follows: A (Excellent): more than 80%, B (Good): 60–80%, C (Fair): 50–60%, D (Limited): less than 50%. For each class the Ministerial decree introduce a score, as follows: 1 (Excellent), 0.80 (Good), 0.50 (Fair), and Zero (Limited). The scores can also be negative: –1 for not admissible products, –2 for plagiarism, –0.5 for each product missing with respect to the expected number of products (three for university researchers, six for PROs, respectively). Consequently, university researchers have a maximum score of 3, if all their products are considered excellent.

ANVUR nominated in 2011 a panel of 450 experts, divided in 14 disciplinary areas<sup>1</sup>. Expert panels have debated and published evaluation criteria, following a mix of bibliometric approaches and informed peer review. Bibliometric criteria have been based on a combination between number of citations (citations received by an article from the date of publication until end 2010, expressed as quantile of the world distribution of citations for articles in the same journal subject category) and journal impact score (using various measures available from JCR and/or Scopus).

All disciplinary panels in science, technology and medicine, including psychology, opted for bibliometric methods as the dominant methodology. However, in all disciplines a sample of papers have been evaluated with both bibliometric and peer review methods, in order to compare the results.

On the contrary, disciplinary panels in architecture, humanities and social sciences have adopted peer review as the main approach, in most cases at 100%. In the case of peer review, each research product has been submitted to two external referees, chosen independently and without mutual information by members of the panel. Scores of referees have then been reconciled within consensus groups; in case of severe disagreement a third referee has been recruited. Overall, more than 14,000 external referees have been mobilized.

Scores of research products have then been aggregated at the level of disciplines and departments. These data have been published at the lowest possible level of aggregation, which in most cases goes down to more than 300 fine-grained disciplines (SSD, Settori Scientifico Disciplinari). Only when the number of researchers in a university was smaller than four, data have been aggregated at a higher level, in order to preserve statistical confidentiality. Individual scores have been communicated via mail to all researchers but have neither been disclosed, nor transferred internally to the Ministry or anyone else, but kept strictly confidential.

<sup>1</sup> The Italian Research System is composed by 370 SDSs, grouped in 14 CUN (Italian National University) areas; every researcher belongs to one and only one SDS. See [http://www.cun.it/media/116411/settori\\_scientifico\\_disciplinari\\_english.pdf](http://www.cun.it/media/116411/settori_scientifico_disciplinari_english.pdf) (accessed November 20, 2015).

Download English Version:

<https://daneshyari.com/en/article/10358345>

Download Persian Version:

<https://daneshyari.com/article/10358345>

[Daneshyari.com](https://daneshyari.com)