

# Article and Journal-Level Metrics in Massive Research Evaluation Exercises: The Italian Case

Marco Malgarini<sup>1</sup>, Carmela Anna Nappi<sup>1</sup> and Roberto Torrini<sup>1</sup>

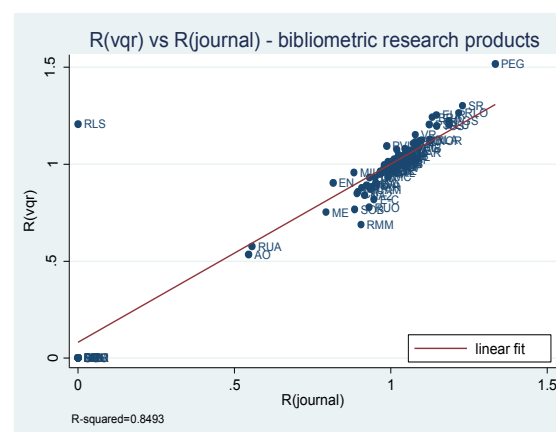
{marco.malgarini, carmelaanna.nappi, roberto.torrini}@anvur.it  
ANVUR, Via Ippolito Nievo 35, 00153, Rome. (Italy)

## Introduction

Article level metrics are usually the preferred choice for research evaluation. However, for recent articles they may be integrated or substituted considering some measure of journal impact (Abramo et al., 2012). The use of journal level metrics is also often considered as particularly appealing for administrative purposes, because of their readily availability, easiness to use and comprehensibility (Bordons et al., 2002). On the other hand, the IF is often criticized on the grounds of its possible biases and lack of methodological consistency (Vanclay, 2012). The aim of our paper is to provide evidence about the effects of the use of journal level metrics on the results of a massive research evaluation exercise like the one that has been performed in Italy with reference to the period 2004-2010 (VQR 2004-2010, see Ancaiani et al., 2015). More specifically, in the following we evaluate the effects of the use of the impact factor (IF) on the ranking of Italian Universities at the aggregate level, at the area level and for individual researchers.

## Effect of the use of the Impact Factor at the University level

In order to assess the impact of the use of IF, we calculate two different indicators of research quality, denoted as  $R_{VQR}$  and  $R_{IF}$ . The former is based on the rules used for the VQR, and the latter uses only the Impact Factor in order to evaluate the articles; the analysis is limited to the research products evaluated only with bibliometrics. We then rank the 93 Italian Universities on the basis of those indicators, finding that the Spearman correlation index among the two rankings is equal to 0.92; moreover, the  $R^2$  of a regression of  $R_{VQR}$  over  $R_{IF}$  and a constant is equal to 0.85. Hence, the analysis at the aggregate level shows that the final ranking of Italian Universities based on journal metric alone is very close to that obtained with the VQR algorithm (see also Figure 1).



**Figure 1 – The relationship among University evaluation performed with different metric.**

## Effect of the use of the Impact Factor at the Area level

However, it is well possible that the relationship is weaker when we are interested in ranking Universities in each scientific area. In order to shed light on this issue, we repeat the analysis for the 14 areas considered in the VQR (Table 1). Correlation between the two rankings is still above 0.8 in all the Research Areas except for Chemistry. The Spearman correlations among rankings are significant at 5% level in all the research areas.

Table 2 reports the coefficients of the regressions of  $R_{VQR}$  on  $R_{IF}$  (beta) and a constant (alpha); the table also reports the  $R^2$  of the regression (column 3) and the standard deviation (column 4) normalized with respect to the average value of  $R_{VQR}$  in each Area. Standard deviation is pretty low if compared to the average value of  $R$  (around 7%) in the Areas of Mathematics, Physics and Industrial Engineering, while in Earth Science, Medicine and Biology the normalized standard deviations grow to 17% of the average level of  $R$  in those areas. Similarly, the areas with a low normalized standard deviation are also those with a higher  $R^2$  and vice-versa. Hence, results confirm that the two evaluation methods bring very similar results also at the area level.

**Table 1. Spearman Correlation between Rankings obtained with VQR bibliometric rules and Journal metric (\* indicates statistical significance at 5%).**

Research Area	Spearman	# Univ.
Mathematics	0.926*	64
Physics	0.825*	65
Chemistry	0.654*	60
Earth Science	0.724*	46
Biology	0.861*	66
Medicine	0.701*	58
Veterinary Sciences	0.876*	50
Construction engineering	0.720*	54
Industrial engineering	0.769*	67
Psychology	0.764*	61

**Table 2. Sensitivity of research evaluation to the use of the Journal Impact Factor at the area level.**

	(1)	(2)	(3)	(4)
Research Area	$\alpha$	$\beta$	$R^2$	St. dv.
Mathematics	-0.055	1.039***	0.921	0.058
Physics	-0.13**	1.124***	0.847	0.060
Chemistry	-0.029	0.998***	0.706	0.100
Earth Science	0.180	0.815***	0.478	0.170
Biology	-0.142	1.132***	0.720	0.168
Medicine	0.083	0.894***	0.340	0.167
Veterinary Sciences	-0.004	1.016***	0.787	0.125
Construction engineering	0.186*	0.813***	0.532	0.100
Industrial engineering	-0.014	1.004***	0.675	0.070
Psychology	0.0778	0.916***	0.744	0.155

#### Effect of the use of the Impact Factor at the individual level

We finally look at how the use of the IF influences evaluation results for each individual researcher. In this case, we regress individual scores obtained using either citations or the Impact Factor. Results of the estimation are reported in Table 3.

The relationship among the results obtained with the two different metrics is now rather weak: the  $R^2$  of the regression is equal to 0.18 for the whole sample, varying between 0.20 and 1.156 in each year. The constant of the regression is rather high, while the beta coefficient associated with the IF is much lower than in previous estimates. Hence, at the individual level using alternatively only the citations or only the impact factor would imply a rather different outcome.

**Table 3. Citations vs Journal Metric scores at individual level.**

	Coefficient							
	Whole sample	2004	2005	2006	2007	2008	2009	2010
IF	0.488 ***	0.525 ***	0.531 ***	0.521 ***	0.507 ***	0.487 ***	0.507 ***	0.383 ***
Constant	0.280 ***	0.247 ***	0.232 ***	0.254 ***	0.282 ***	0.301 ***	0.233 ***	0.374 ***
#obs.	76,15	9,23	9,77	10,24	10,88	11,56	12,15	12,31
$R^2$	0.184	0.201	0.197	0.202	0.197	0.194	0.186	0.156

#### Conclusions

Overall, results may be considered as supportive of the idea of using two different bibliometric indicators for assessing research quality: on one hand, the use of the IF is not found to bias in a significant way University rankings, both at the aggregate and at the Area level; on the other hand, at the individual level, citations and IF evaluation are found to be rather different, pointing to the need of integrating the two different information in order to obtain a more robust measure of research quality for each individual researcher.

#### References

- Abramo G., D'Angelo C.A., & Costa F. (2012). Citations versus journal impact factor as proxy of quality: could the latter ever be preferable? *Scientometrics*, 84(3), 821-833.
- Ancaiani A. et al. (2015). Evaluating scientific research in Italy: the 2004-2010 Research evaluation exercise. Forthcoming in *Research Evaluation*.
- Bordons, M., Fernández, M. T., & Gomez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance. *Scientometrics*, 53(2), 195-206.
- Vanclay J.K. (2012). Impact factor: outdated artefact or stepping-stone to journal certification? *Scientometrics*, 92(2), 211-238.