

**Did submission rules affect the submission sizes of the units to the UK's Research
Excellence Framework in 2014?**

Mehmet Pinar

Emre Unlu

Abstract

The Research Excellence Framework (REF) follows a set of submission rules. Here we analyse whether the submission rules of the impact element arguably shaped the submission sizes of the submitting units – a group of academics researching on a specific subject area in higher education providers – to the REF in 2014. The number of impact case studies required was determined by the number of full-time equivalent (FTE) staff members submitted by the unit. We argue that units that did not have an extra impact case study or units' perception of lower quality of additional impact case study made some units to lower their submission sizes. We show that there were proportionately more submissions with a size just below the threshold FTE – a threshold used to identify the numbers of impact case studies– than those just above the threshold, suggesting that some units arguably decreased their size to return fewer impact studies.

Keywords: Research Excellence Framework, impact element, submission rules, gaming, metrics, research assessment

This is authors' accepted manuscript and this article is published in *Perspectives: Policy and Practice in Higher Education*, and can be accessed via:
<https://doi.org/10.1080/13603108.2022.2122623>

Business School, Edge Hill University, St Helens Road, Ormskirk, United Kingdom

1. Introduction

Research assessment exercises have been widely used across the globe to evaluate the quality of the research (Zacharewicz et al., 2019; Pinar and Horne, 2022). Research Excellence Framework (REF) of the United Kingdom (UK) is one of the many research evaluations and was firstly introduced in 2014. Yet, the research exercises in the UK go back to the Research Selectivity Exercise, which was implemented in 1986 (see Shattock, 2012 for discussion of research exercises in the UK). Recently, the next round of REF was carried out in 2021 by four UK funding bodies (i.e., Research England, Scottish Funding Council, Higher Education Funding Council for Wales and Department for the Economy in Northern Ireland) to provide accountability for public money, to establish reputational yardsticks, and to inform the selective allocation of research funding (REF, 2021).

Like any assessment, the research assessments in the UK were not immune to the ‘game-playing’ based on the rules of the ‘game’. Murphy (2017) examined the rules for submitting research outputs to the REF and highlighted that higher education institutes (HEIs) might have acted tactically in REF2014 by ‘cherry-picking’ both staff and research outputs for submission and recruiting staff with existing high-quality papers. Examining the economics and econometrics submissions to the assessment periods between 1992 and 2014, Johnston and Reeves (2017) found that units that performed below the university expectations were not submitted in the following submission periods to increase the institution’s overall reputation. Furthermore, the early- and mid-career researchers in library and information science also believed that the introduction of impact to REF resulted in ‘game playing’ (Marcella et al., 2018).

The REF in 2014 (REF2014) assessed the quality of research outputs, impacts of the research on the economy, society and/or culture, and the research environment. One of the main differences of the REF from its predecessors was the inclusion of the impact element as

part of the evaluation. Four UK funding bodies introduced the impact element in the new assessment and argued that excellent research should have an impact on society and the economy, and they encouraged the HEIs to produce impactful research (REF, 2010). However, the inclusion of this element has received many criticisms. Many academics were discontent with the inclusion of the impact element in the research assessment exercise as academics view this element as an infringement on a scholarly way of life and fundamentally harmful to the production of new knowledge (Watermeyer, 2012; Watermeyer, 2016; Weinstein et al., 2021). In a recent paper, Pinar and Unlu (2020a) showed that the inclusion of the impact element in the REF2014 increased the research income gaps between subjects and HEIs, and they argued that the differences in impact quality across submissions were relatively higher. Even though there were criticisms of the impact element, the impact element was kept in the REF in 2021 (REF2021), and its importance in the funding allocation was increased from 20% to 25% (REF, 2018).

Similar to any element of the REF, the inclusion of the impact may have led to ‘game playing’. Even though some impact case study examples were provided, most academics were not familiar with the impact element (Manville et al. 2015) and lacked public engagement (Chikoore et al., 2016). Even the evaluators had difficulty assessing the impact element (Menville et al., 2015). Based on the arguments mentioned above, we argue that the non-existence or availability of relatively weaker impact case studies may have led submitting units (i.e., a group of academics working on a specific subject area of research in higher education providers) to return relatively lower numbers of full-time equivalent (FTE) staff members to the REF2014 since the numbers of FTE staff members used as a proxy to determine the number of impact case studies submitted by units.

2. Research Excellence Framework in 2014

The first REF cycle assessed the quality of the research outputs, quality of research outputs (output hereafter), impacts of the research on the economy, society and/or culture (impact hereafter), and the research environment (environment hereafter). Each submitting unit was required to submit up to four outputs per staff member, submit an environment data consisting of information on i) research doctoral degrees awarded each year in the period 1 August 2008 to 31 July 2013; ii) the amounts and sources of external research income for each year in the period 1 August 2008 to 31 July 2013; and iii) the amount of research income-in-kind for each year in the period 1 August 2008 to 31 July 2013; and environment template detailing the research environment (see Thorpe et al. (2018) for analysis of environment templates; Pinar and Unlu (2020b) for analysis of research environment data). Finally, submitting units provided impact templates and impact case studies. Impact templates offered information about the unit's approach to impact during the assessment period (1 January 2008 to 31 July 2013). On the other hand, impact case studies provided details of the societal and economic impact that occurred during the assessment period (1 January 2008 to 31 July 2013) underpinned by excellent research published from 1 January 1993 to 31 December 2013 (REF, 2011). Table 1 provides the number of impact case studies required from units based on the number of FTE staff members returned.

<Insert Table 1 approximately here>

The REF 2014 consisted of 36 units, and sub-panels of subject experts assessed each unit (see e.g., REF, 2012 for detailed definitions of units and panels in the REF 2014). The outputs were evaluated in their 'originality, significance and rigour'. The impact element was evaluated in terms of its 'reach and significance' for the economy, society and/or culture underpinned by excellent research. Finally, the research environment was evaluated in terms of its 'vitality and sustainability', including its contribution to the vitality and sustainability of the broader discipline or research base (REF, 2011). Outputs, environment template and data,

and impact templates and case studies were rated by the experts based on the five categories: four-star (world-leading research); three-star (internationally excellent research); two-star (internationally recognised research); one-star (nationally recognised research) and unclassified if the research falls below the standard of nationally recognised. Finally, the output, impact, and environment elements were given 65%, 20% and 15% importance to obtain the overall quality profile of submitting units.

Based on the REF2014 results obtained in each element, the four UK funding bodies have distributed quality-related research (QR) funding to the universities. Research England (2021) provides four stages followed in allocating the funding across universities:

- Stage 1: Mainstream QR budget split into three sub-profile pots. 65%, 20% and 15% of the total funding is distributed in output, impact and environment pots, respectively.
- Stage 2: Each sub-profile pot is distributed between the four main panels: Panel A (medicine, health and life sciences); Panel B (physical sciences, engineering and mathematics); Panel C (social sciences); and Panel D (arts and humanities). The total in each pot is divided in proportion to the volume of research in each panel that met or exceeded the three-star quality level in the REF, weighted to reflect the relative costs of research in different subjects.
- Stages 3 and 4: The funding in each main panel is distributed between the unit of assessments (UOAs) and higher education providers. The allocation of funds is proportional to the volume of research activity reaching the REF's three- and four-star quality levels, multiplied by quality and cost weights.

High-cost laboratory and clinical subjects are given a cost weight of 1.6, intermediate-cost subjects are given a cost weight of 1.3, and other subjects (primarily considering the social sciences subjects) are given a cost weight of 1. Finally, the research activity rated as four-star

and three-star were given quality weights of 4 and 1, respectively, and research activity rated less than three-star was given zero quality weights. Therefore, the research activity rated as world-leading (four-star) is allocated four times QR funding compared to the research activity rated internationally excellent (three-star). The research activity rated less than three-star is allocated no QR funding. The interested readers could refer to Kelly (2016) and Pinar (2020) that examined the relationship between REF results and funding allocation across English universities.

3. Research hypothesis and data

Based on the submission rules in REF2014 and the impact element being the ‘novel’ part of the REF, many units and assessors were unfamiliar with the impact element. Therefore, we argue that units that did not have an extra impact case study or units’ perception of lower quality of additional impact case study made some units return lower numbers of FTE staff members to avoid extra impact case study submission to the REF. In other words, we expect that there would be proportionately more submissions with a size just below the threshold FTE than those just above the threshold, suggesting that some units arguably decreased their size to return fewer impact studies. The threshold FTEs here refer to the FTE levels used to identify the number of impact case studies returned by the submitting units (see Table 1 for the threshold levels).

<Insert Table 1 approximately here>

Let us provide a hypothetical example to discuss why we expect more submissions just below a given threshold compared to submissions just above the threshold. Consider a unit that plans to return 26 staff members to the REF. Then this unit would submit specific numbers of outputs, environment data and template, impact case study template, and four impact case studies. Before making the final submissions, the units conduct internal and external evaluations of different submission elements. We expect that the critical aspect in decisions

about the number of FTE returned in each unit was the estimated quality of the impact case studies. The decision-makers on the submission of the REF returns were unaware of the actual scores obtained by potential impact case studies. Therefore, there was uncertainty associated with the impact case study ratings. Consider that the estimated ratings of this unit's three impact case studies were four-star, and one impact case study's rating was estimated to be three-star. Given that the four-star research activity was rewarded four times more funding than the three-star one, the unit may decide to exclude 2 FTE from their submission, resulting in the submission of the three impact case studies estimated to be rated four-star. 2 FTE from this submission could be returned in another unit that does not have similar constraints (i.e., their allocation in another unit will not result in submitting an additional impact case study) or could be dropped out from the REF submission. The exclusion of 'two weakest' FTE staff members from this unit would not result in significant changes in the environment template and data and output returns; however, excluding 2 FTE staff members from this unit would generate more QR funding and higher REF scores. Therefore, based on the above arguments, we set the hypothesis as follows:

Hypothesis: The number of submissions just below a threshold FTE is significantly higher than the number of submissions just above the threshold FTE.

We obtain the submission data from the REF2014 web page (<https://www.ref.ac.uk/2014/>) to examine whether the frequency of submissions just below the threshold is significantly different than the frequency of submissions just above the threshold.

4. Analysis

To examine whether such game culture exists, we counted the number of submissions with a size just below the FTE threshold levels (i.e., the number of submissions with a size ranging between 14 and 14.99, 24 and 24.99, 34 and 34.99, and so on). Similarly, we also obtained the number of submissions with a size just above the FTE thresholds (i.e., 15 and

15.99, 25 and 25.99, 35 and 35.99, and so on). Based on the hypothesis, we expect that the number of submissions just below a threshold FTE is significantly higher than that of submissions just above the threshold FTE.

Figure 1 presents the numbers of submissions that were just below (above) the threshold of submissions made to the four main panels: Panel A: medicine, health, and life sciences (consisting of UoAs 1-6); Panel B: physical sciences, engineering, and mathematics (consisting of UoAs 7-15); Panel C: social sciences (consisting of UoAs 16-26); and Panel D: arts and humanities (consisting of UoAs 27-36). In all the panels, we observe that numbers of units with submission sizes just below the threshold levels were disproportionately higher than those with submission sizes just above the threshold levels (i.e., 86 vs. 5, 129 vs. 4, 94 vs. 22, and 78 vs. 19 in panels A, B, C and D, respectively). Overall, 33%, 28%, 16% and 14% of the total submissions made to panels B, A, C and D were just below the FTE thresholds, respectively. This finding suggests that this type of game playing was observed more frequently in panels A and B than in panels C and D.¹ However, irrespective of the panel, we observe that large numbers of submissions were clustered just below the FTE thresholds.

<Insert Figure 1 approximately here>

Figures 2, 3, 4 and 5 also offer the same information as Figure 1 for the UoAs in panels A, B, C and D, respectively. Again, a similar behaviour (i.e., large numbers of submissions with a size just below the threshold) is observed across different UOAs. In particular, more than 30% of the submissions in General Engineering (UOA15), Physics (UOA9), Earth Systems and Environmental Sciences (UOA7), Clinical Medicine (UOA1), Civil and Construction Engineering (UOA14), Biological Sciences (UOA5), Allied Health Professions, Dentistry, Nursing and Pharmacy (UOA3), Chemistry (UOA8), Mathematical Sciences

¹ The percentages are obtained by dividing the numbers of submissions with a size just below the threshold with the total number of submissions, and multiplied by 100.

(UOA10) had submission sizes just below the threshold FTE levels. On the other hand, the tendency of submissions with a size just below the FTE threshold was relatively low in social sciences and arts and humanities fields (i.e., panels C and D, respectively). For instance, only 9%, 10% and 11% of the submissions in the Law (UOA20), Philosophy (UOA32), and Geography, Environmental Studies and Archaeology (UOA17) had sizes just below the FTE thresholds, respectively.

<Insert Figures 2, 3, 4 and 5 approximately here>

We also carried out a Student's t-test to examine whether there were significantly more submissions below the threshold levels or not. If the thresholds set for determining the number of impact studies returned did not play any role, then we would expect that the submissions clustered around 15, 25, 35, and so on, would be random. Hence, the number of submissions just below (above) the threshold would not have any statistical difference. To carry out the t-test, we only used the submissions with submission sizes ranging between 14 and 15.99, 24 and 25.99, 34 and 35.99, and so on, and then subtracted 10, 20, 30, and so on, respectively, from these submissions, so that we could analyse the whole set of submissions clustered around the threshold levels. Finally, we used the Student's t-test to examine whether the average value of the submissions clustered around the threshold levels is significantly different from 5. Appendix Tables A1 and A2 present the detailed test statistics and the respective significance levels for the panels and UoAs, respectively. We find that the submissions were clustered significantly just below the threshold at the 1% level for all the panels. Furthermore, with the exceptions of the Sociology (UOA23), Anthropology and Development Studies (UOA24) and Theology and Religious Studies (UOA33) units, there were significantly more submissions just below the threshold levels for all the units.

We argue that most units submitting just below the threshold have potential improvements in their overall REF scores and experienced higher subject rankings by not

returning additional impact case studies. If the quality of the additional impact case study obtained a low rating, the overall scores of these units would have been lower and they would have generated a relatively lower amount of QR funding. Units might have had some strategies to have a submission size below the thresholds. One possibility is that some staff members are potentially excluded from their submissions, which would have implications for these staff members. These staff members were potentially moved to the teaching-only contracts before the REF census date (i.e., the date for selecting staff members to be returned to REF). Therefore, their job descriptions are changed, and their research independence would be limited as they would not be allowed to carry out research activities that would enable them to be part of the REF submission. Another possibility is that some units did strategic hiring. If they did not expect that they could fulfill the number of impact cases, they might prioritise the researchers with an established research agenda when recruiting new staff members. This might create disadvantages for early career researchers in the academic job market. Strategic hiring may also imply that the units might prefer part-time or temporary contracts rather than offering permanent contracts (such as hiring associate tutors). Therefore, we argue that this type of game playing may limit research roles for some academic members. Overall, we do not argue that some institutions are unfairly funded due to game playing as they follow the guidelines and act based on these rules, but submission rules of the exercise may lead to unintended consequences such as excluding some academics from the REF submission.

There are potential policy actions that could be carried out by four funding bodies and HEIs to avoid this type of game playing in future research assessments. First, since most academics were unfamiliar with the impact element (Manville et al. 2015) and lacked public engagement (Chikoore et al., 2016), four funding bodies could provide training sessions at the HEIs to familiarise academics with the impact element. These training sessions could

include assessors and impact case study authors. The evaluators can explain why some impact case studies were rated well. On the other hand, academics with good impact case studies could provide their experience with the academics. Secondly, funding bodies may change the threshold values set for the number of impact case studies. Since this type of clustering below the threshold FTE levels occurred for smaller units, the funding bodies may expand the band of the FTEs for returning the same number of impact studies at the lower levels of FTE. This would then decrease this type of game playing. Thirdly, HEIs could also provide some incentives for academics that engage with impactful research by providing additional research money and time for staff members to engage with impactful research.

5. Conclusions

Given that the impact evaluation was first introduced in the REF2014 and that some units did not engage with the impact element, submitting teams and HEIs had reputational and monetary reasons to act tactically to return fewer FTE staff members to submit fewer impact case study. This paper analysed whether the threshold FTE levels set to determine the number of impact case studies to be returned to the REF2014 made units to ‘play the game by the rules’ and act strategically. Our analysis shows that a high percentage of the submissions had a size just below the FTE thresholds in most of the UOAs in panels A and B. Except for Agriculture, Veterinary and Food Science (UOA6) and Aeronautical, Mechanical, Chemical and Manufacturing Engineering (UOA12) units, more than 20% of the submissions in other units in panels A and B had a size just below the threshold FTE. On the other hand, except Law (UOA20) unit, 10% to 20% of the total submissions in a given UOA in panels C and D had a size just below the threshold. Overall, the strategy of having a submission size just below the FTE threshold was more present in UOAs of panels A and B compared to those in panels C and D, suggesting that units in panels A and B acted strategically to increase their likelihood of receiving more QR funding as well as their overall reputation.

In this paper, we used submission data for the REF2014; however, the REF in 2021 (REF2021) took place, and the results of the REF2021 is released in May 2022. Compared to the REF2014, there were some changes in the rules of the REF2021. Some of the key changes relevant to this paper's analysis are as follows. First, the weight attached to the impact element was increased from 20% to 25%, and the weight given to the output element was decreased from 65% to 60%, and the weight given to the research environment was left to be 15% (see paragraph 51 of REF, 2019). Secondly, HEIs were expected to return all staff "with significant responsibility for research" to the REF2021 (see paragraph 51 of REF, 2019). Finally, the number of case studies required from a submitting unit was determined by the submission size. Four UK funding bodies changed the FTE intervals that determine the number of impact case studies required from the submitting units (see paragraph 309 of REF, 2019). The units that had a size up to 19.99 FTE were required to submit two impact case studies. If submitting unit had a size ranging from 20 to 34.99, from 35 to 49.99, from 50 to 64.99, from 65 to 79.99, from 80 to 94.99, from 95 to 109.9, from 110 to 159.99, they were required to submit 3, 4, 5, 6, 7, 8 and 9 impact studies, respectively. Finally, any submission that had a size of 160 or more was required to submit ten impact case studies, and a further case study was required per additional 50 FTE. In other words, the minimum FTE threshold for the REF2021 was increased to 19.99 from 14.99 in REF2014. The intervals of the bands that require the same number of impact case studies increased from 10 to 15 for the cases up to 8 impact case studies.

Based on the changes made between REF2014 and REF2021, Wilsdon (2017) argued that the so-called 'universal REF' removes the selectivity card played so frequently in REF2014 (i.e., choosing the numbers and staff to be returned), but suggested that the new rules do allow some scope for institutional brinkmanship. We also think that the possibility of this type of game playing may be limited in the REF2021. Given the range of FTE intervals for the submission of the same number of impact case studies increased, we argue that the

maneuvering possibility of the units to submit relatively lower FTE is reduced. Furthermore, units are less likely to change their submission size because of the ‘universal REF’. Finally, the units had more know-how experience of impact element since the REF2014 (e.g., how to carry out impactful research, how to evidence the impact, and how to engage with the end-users, etc.) and therefore, we believe that the engagement of the staff members with the impact studies may have increased between REF2014 and REF2021.

Even though the game playing may be limited in the REF2021 compared to the REF2014, we still think that units may have continued to act tactically to return fewer FTE staff members to avoid additional impact case studies if they consider this additional impact case study to be rated low. Therefore, we still expect to see a similar type of clustering of submissions with a size just below the new thresholds set in the REF2021. There are various reasons for this possibility. Firstly, units may have done this by altering the contracts of the staff members to decrease the number of staff members “with significant responsibility for research”. Secondly, the unit might have followed a hiring strategy based on the impact case studies available to them during the assessment period of the REF2021 and may have ensured that they ended up with a submission size just below the threshold. Thirdly, potentially HEIs could return some staff members, not in their ‘original’ units – a unit in which there is a lack of impact case studies - but in other units in which there are available ‘good’ impact case studies. In sum, even though the game playing of this nature may be limited, units had some tools that they may have used to return submissions with sizes just below the thresholds.

A future study could analyse the REF2021 submission data to examine whether large numbers of submissions clustered just below the new FTE thresholds or not in the REF2021. Furthermore, this analysis could also compare the intensity of such clustering in the REF2021 with the REF2014 to explore whether the new rules in the REF2021 decreased the intensity of such behaviour.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

<Insert Appendix Tables A1 and A2 approximately here>

References

Chikoore, L., Proberts, S., Fry, J., and C. Creaser. 2016. "How are UK Academics Engaging the Public with their Research? A Cross-Disciplinary Perspective." *Higher Education Quarterly* 70 (2): 145-169.

Johnston, J. A. Reeves. 2017. "Assessing research performance in UK universities using the case of the economics and econometrics unit of assessment in the 1992–2014 research evaluation exercises." *Research Evaluation* 26 (1): 28-40.

Kelly, A. 2016. "Funding in English Universities and its relationship to the Research Excellence Framework." *British Educational Research Journal* 42 (4): 665-681.

Manville, C., Guthrie, S., Henham, M.-L., Garrod, B., Sousa, S., Kirtley, A., Castle-Clarke, S., and T. Ling. 2015. "Assessing impact submissions for REF 2014: An evaluation."

www.rand.org/content/dam/rand/pubs/research_reports/RR1000/RR1032/RAND_RR1032.pdf

f

Marcella, R., Lockerbie, H., Bloice, L., Hood, C., and F. Barton. 2018. "The effects of the research excellence framework research impact agenda on early- and mid-career researchers in library and information science." *Journal of Information Science* 44 (5): 608-618.

Murphy, T. 2017. "Revisiting the research excellence framework: ensuring quality in REF2021 or new challenges ahead?" *Perspectives, Policy and Practice in Higher Education* 21 (1): 34-39.

Pinar, M. 2020. "It is not all about performance: Importance of the funding formula in the allocation of performance-based research funding in England." *Research Evaluation* 29 (1): 100-119.

Pinar, M., and Horne, T.J. 2022. "Assessing research excellence: Evaluating the Research Excellence Framework." *Research Evaluation* 31 (2): 173-187.

Pinar, M., and E. Unlu. 2020a. "Evaluating the potential effect of the increased importance of the impact component in the Research Excellence Framework of the UK." *British Educational Research Journal* 46 (1): 140-160.

Pinar, M., and E. Unlu. 2020b. "Determinants of quality of research environment: An assessment of the environment submissions in the UK's Research Excellence Framework in 2014." *Research Evaluation* 29 (3): 231-244.

Research England. 2021. "Research England: how we fund higher education institutions. Supporting healthy and dynamic research and knowledge exchange in HEPs."

<https://www.ukri.org/wp-content/uploads/2021/08/RE-06082021-RE-How-we-fund-HEPs-FINAL.pdf>

REF. 2010. "Research Excellence Framework impact pilot exercise: Findings of the expert panels. A report to the UK higher education funding bodies by the chairs of the impact pilot panels."

https://www.ref.ac.uk/2014/media/ref/content/pub/researchexcellenceframeworkimpactpiloxercisefindingsoftheexpertpanels/re01_10.pdf

REF. 2011. "Assessment framework and guidance on submissions." REF 02.2011 July 2011.

<https://www.ref.ac.uk/2014/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/GOS%20including%20addendum.pdf>

REF. 2012. “Panel criteria and working methods.” REF 01.2012 January 2012. Available via:
https://www.ref.ac.uk/2014/media/ref/content/pub/panelcriteriaandworkingmethods/01_12.pdf

[f](#)

REF. 2018. “Draft guidance on submissions.” https://www.ref.ac.uk/media/1016/draft-guidance-on-submissions-ref-2018_1.pdf

REF. 2019. “Guidance on submissions.” https://www.ref.ac.uk/media/1092/ref-2019_01-guidance-on-submissions.pdf

REF. 2021. “What is the REF?” <https://www.ref.ac.uk/about/what-is-the-ref/>

Shattock, M. 2012. *Making Policy in British Higher Education 1945–2011*. Berkshire: McGraw-Hill.

Thorpe, A., Craig, R., Tourish, D., Hadikin, G., and S. Batistic. (2018). “‘Environment’ Submissions in the UK’s Research Excellence Framework 2014.” *British Journal of Management* 29: 571-587.

Watermeyer, R. 2012. “Issues in the Articulation of ‘Impact’: the Response of UK Academics to ‘Impact’ as a new Measure of Research Assessment.” *Studies in Higher Education* 39 (2): 359-377.

Watermeyer, R. 2016. “Impact in the REF: Issues and Obstacles.” *Studies in Higher Education* 41(2): 199-214.

Weinstein, N., Chubb, J.A., Haddock, G., and J.R. Wilsdon. 2021. “A conducive environment? The role of need support in the higher education workplace and its effect on academics’ experiences of research assessment in the UK.” *Higher Education Quarterly* 75 (1): 146-160.

Wilsdon, J. 2017. “The new REF rules could end game-playing – but only if we let them. Centre for Global Higher Education.” <https://www.researchcghe.org/blog/2017-11-30-the-new-ref-rules-could-end-game-playing-but-only-if-we-let-them/>

Zacharewicz, T., Lepori, B., Reale, E., and K. Jonkers. 2019. "Performance-Based Research Funding in EU Member States—A Comparative Assessment." *Science and Public Policy* 46: 105-115.

Table and Figures

Table 1: Number of case studies required in submissions	
Number of staff submitted (FTE)	Required number of case studies
Up to 14.99	2
15 – 24.99	3
25 – 34.99	4
35 – 44.99	5
45 or more	6, plus 1 further case study per additional 10 FTE
Obtained from REF (2011).	

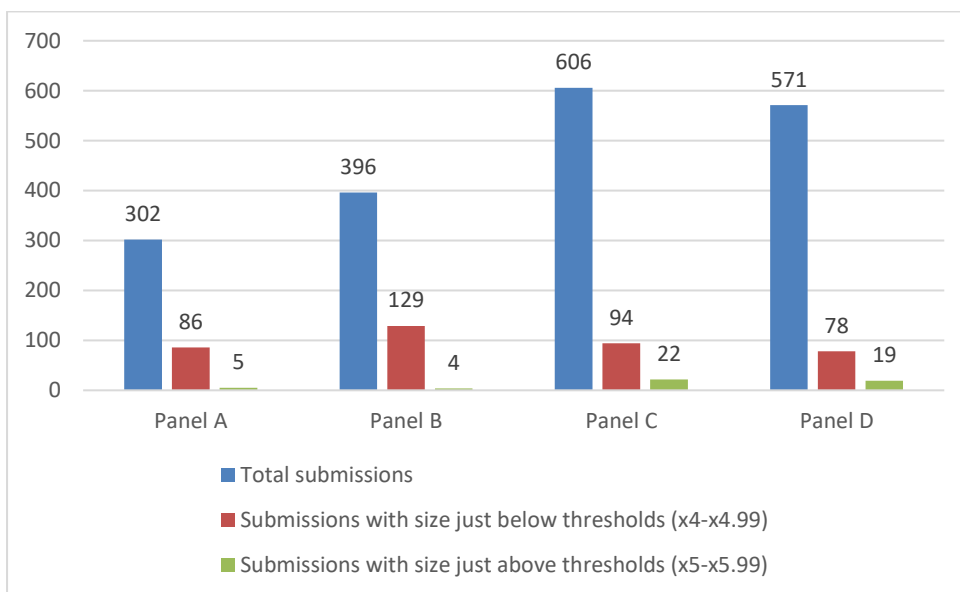


Figure 1. Total number of submissions, submissions with size just below and just above the threshold for panels A, B, C and D

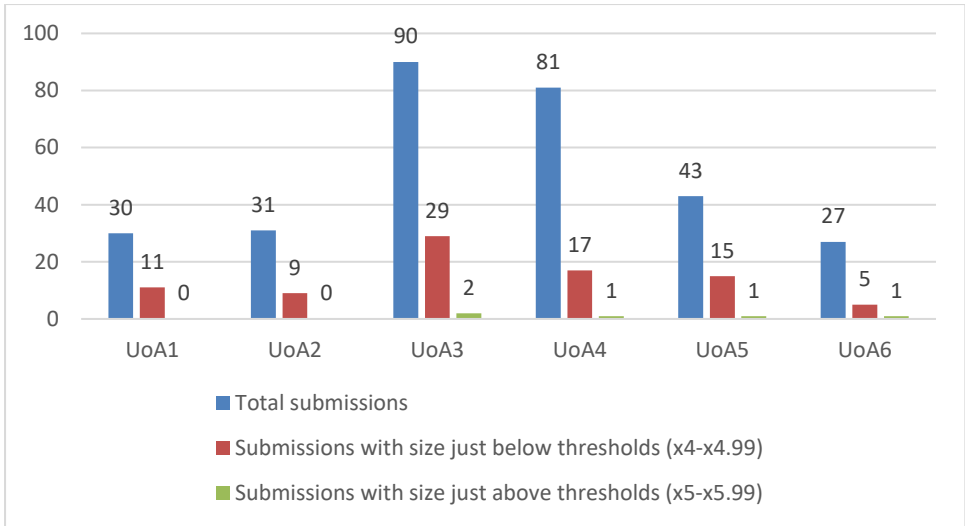


Figure 2. Total number of submissions, submissions with size just below and just above the threshold for UoAs in panel A

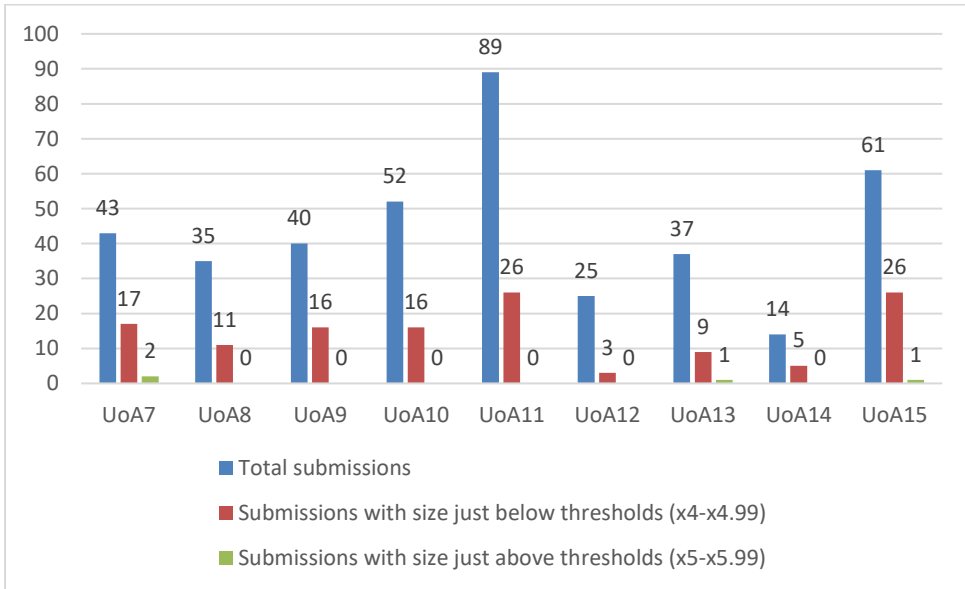


Figure 3. Total number of submissions, submissions with size just below and just above the threshold for UoAs in panel B

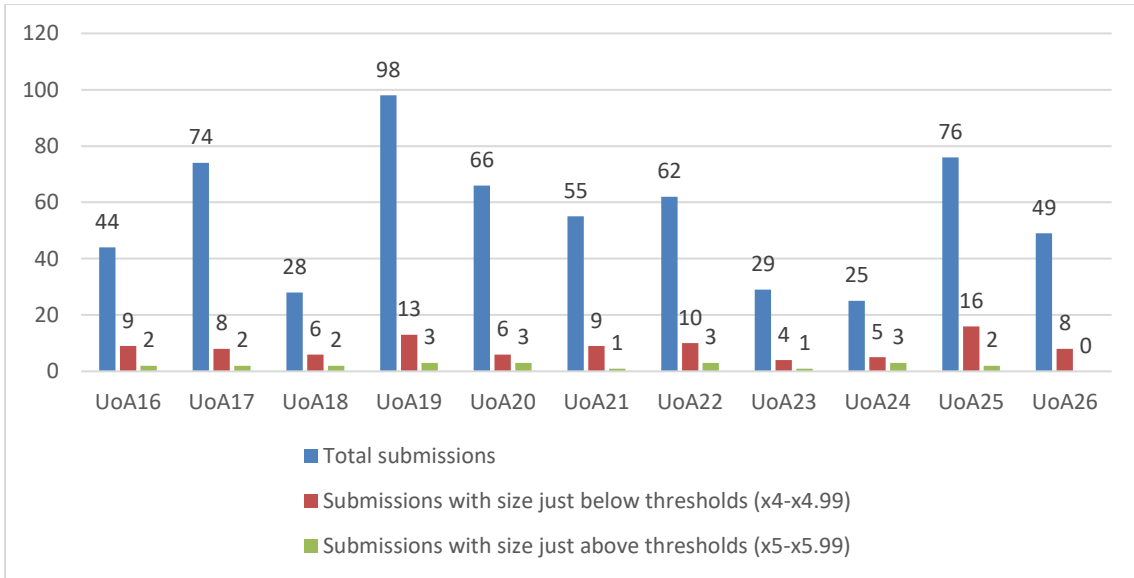


Figure 4. Total number of submissions, submissions with size just below and just above the threshold for UoAs in panel C

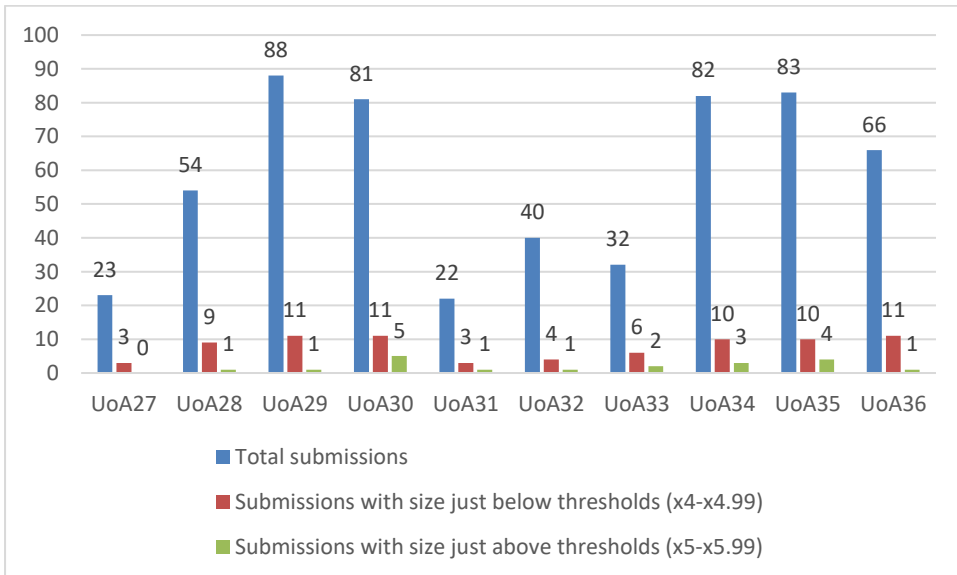


Figure 5. Total number of submissions, submissions with size just below and just above the threshold for UoAs in panel D

Table A1. Test statistics for panels				
Panel	Panel name	Average size (x4-x5.99)	Test statistic	Significance
Panel A	Medicine, health and life sciences	4.504	-11.827	1%
Panel B	Physical sciences, engineering and mathematics	4.492	-14.564	1%
Panel C	Social sciences	4.636	-7.006	1%
Panel D	Arts and humanities	4.621	-7.147	1%

Notes: The submissions with submission sizes ranging between 14 and 15.99, 24 and 25.99, 34 and 35.99, and so on, are considered. 10, 20, 30, and so on are subtracted from these submission sizes, respectively. Average size presents the average submission size for these submissions.

UoA no	UoA name	Average size (x4-x5.99)	Test statistic	Significance
1	Clinical Medicine	4.308	-13.842	1%
2	Public Health, Health Services and Primary Care	4.503	-4.655	1%
3	Allied Health Professions, Dentistry, Nursing and Pharmacy	4.543	-6.130	1%
4	Psychology, Psychiatry and Neuroscience	4.566	-4.040	1%
5	Biological Sciences	4.499	-4.666	1%
6	Agriculture, Veterinary and Food Science	4.487	-2.455	5%
7	Earth Systems and Environmental Sciences	4.734	-2.544	5%
8	Chemistry	4.405	-5.869	1%
9	Physics	4.566	-5.418	1%
10	Mathematical Sciences	4.378	-7.045	1%
11	Computer Science and Informatics	4.472	-8.252	1%
12	Aeronautical, Mechanical, Chemical and Manufacturing Engineering	4.380	-4.881	5%
13	Electrical and Electronic Engineering, Metallurgy and Materials	4.392	-3.801	1%
14	Civil and Construction Engineering	4.450	-3.939	1%
15	General Engineering	4.458	-5.895	1%
16	Architecture, Built Environment and Planning	4.664	-1.903	5%
17	Geography, Environmental Studies and Archaeology	4.812	-1.506	10%
18	Economics and Econometrics	4.575	-2.979	5%
19	Business and Management Studies	4.649	-2.408	5%
20	Law	4.667	-1.456	10%
21	Politics and International Studies	4.485	-3.013	1%
22	Social Work and Social Policy	4.726	-1.639	10%
23	Sociology	4.804	-0.696	NS
24	Anthropology and Development Studies	5.014	0.053	NS
25	Education	4.481	-4.268	1%
26	Sport and Exercise Sciences, Leisure and Tourism	4.294	-5.794	1%
27	Area Studies	4.180	-4.556	5%
28	Modern Languages and Linguistics	4.518	-3.439	1%
29	English Language and Literature	4.549	-3.132	1%
30	History	4.580	-3.246	1%
31	Classics	4.300	-2.941	5%
32	Philosophy	4.502	-2.329	5%
33	Theology and Religious Studies	4.801	-0.862	NS
34	Art and Design: History, Practice and Theory	4.796	-1.293	10%
35	Music, Drama, Dance and Performing Arts	4.773	-1.381	10%
36	Communication, Cultural and Media Studies, Library and Information Management	4.613	-3.264	1%

Notes: The submissions with submission sizes ranging between 14 and 15.99, 24 and 25.99, 34 and 35.99, and so on, are considered. 10, 20, 30, and so on are subtracted from these submission sizes, respectively. Average size presents the average submission size for these submissions. NS represents no significance.