

Lord Stern's review of the Research Excellence Framework

The British Society for Immunology (BSI) is the largest immunology society in Europe. We are a learned society representing the interests of members working in academia, clinical medicine, and industry. Our main objective is to promote and support excellence in research, scholarship and clinical practice in immunology for the benefit of human and animal health. As an organisation, we derive no institutional benefit from the REF.

Summary of Key Points

- We have some concern that the definition of impact disadvantages valuable basic science and that this could have negative consequences for the UK and its research base. We would encourage a broader definition of impact, one that more reliably captures fundamental research, and discourage any move to attach greater weighting to the impact portion of the REF score.
- It is important that the REF accurately capture output that is achieved through a consortium of authors and that "substantial contribution" of individual authors is appropriately measured. We would encourage more guidance on how this is achieved in the assessment process.
- Peer review is the gold standard of academic assessment and has an important role to play, in combination with metrics, in assessing research output. New metrics, such as ratio-metrics, may be a useful addition to the exercise. In contrast, altmetrics are yet to reach a stage where they are a reliable indicator of research quality and we would not encourage their use until further research is applied to disentangle their value and applicability across a broader range of research outputs.
- We would strongly encourage the provision of detailed feedback for individual researchers so as to build more transparency into the assessment exercise. Our members have reported that they would find this useful.
- It is somewhat regrettable that REF is now widely used for HR purposes as a staff appraisal tool. This unfairly disadvantages researchers who work collaboratively or have a smaller number of high impact papers. The "transfer window" created as a result of universities attempting to "poach" staff who can increase their scores is also an unsettling influence.
- Additional data such as conference activity and contributions to external examination could comprise additional information which would support research funders in driving academic excellence.
- Much of the immunology ongoing in UK institutions is highly collaborative by its very nature. The REF has influenced this to a degree, including through encouraging collaboration with the private sector through the assessment of impact. However, there may be disadvantages in "forcing" these relationships further as opposed to allowing them to arise organically.

- The perverse outcomes in university recruitment practice that arise through the REF may be resolved through ensuring that the staff selected have been in place for at least 3 years prior to the REF census date. Alternatively, fractional selection may also have a role to play, where staff are selected at random for analysis.
- Reviews, especially peer reviewed reviews, continue to have an important place in the academic literature and should be included as part of the assessment.
- There is a risk that if universities assign QR funding along existing strengths, as researched by REF performance, inequalities both within and between institutions can be exacerbated. However, the REF exercise does have utility in informing investment decisions such that institutions can invest where they feel they will achieve the best value for money.
- There may be merit in exploring the benefits of submitting data at the level of the department/faculty. This could help mitigate negative associations around individual reporting, especially protecting early career researchers, and also has the advantage of communicating the importance of team science.

Questions

1. What changes to existing processes could more efficiently or more accurately assess the outputs, impacts and contexts of research in order to allocate QR? Should the definition of impact be broadened or refined? Is there scope for more or different use of metrics in any areas?

- We believe there would be considerable merit in broadening the definition of impact. We have heard representation from immunologists that the existing characterisation disadvantages fundamental research and is skewed towards translational science (where it is easier to evidence material impacts). The proscriptive definition and requirements risk devaluing the basic research upon which strands of outstanding applied immunology are based.
- Consequently there is some concern that the present definition may influence a shift in research funding and activity away from underlying areas of basic science. These concerns are backed by the findings of a BiS report highlighting that publicly funded pure basic research declined from a peak of 62% in 2005/06 to just 35% in 2011/12.¹
- We would therefore resist attempts to increase the weighting attached to impact and the bearing it has on the overall REF score unless the definition was broadened. We acknowledge the value of assessing impact in terms of understanding the benefit research brings to wider society. We also recognise the strengthening political emphasis around the contribution of science to the national economy and its impact beyond academia. Nevertheless, we feel that increasing the weighting without significantly amending the definition to better take account of basic research would further incentivise universities to encourage short-term commercially-oriented research at the expense of creative blue-sky activities. The consequences for such short-terminism will be a massive negative impact on creative innovative fundamental research in the UK, which is the foundation upon which applied/translational research is built. In the medium to long term, this will render the UK much less competitive in the knowledge economy-based enterprise. We feel it is important to recognise and encourage a broad research mix of fundamental and applied science. The REF has a role to play in this.
- Ensuring the REF accurately captures output accomplished through a consortium of multiple authors, often collaborating across multiple institutions, and with a multi-disciplinary mix of expertise (sometimes including partners in industry) is important. Assigning “substantial contribution” in these circumstances can be difficult and our members would appreciate clearer guidance on the evaluation process. There is an argument for including a form of weighting to establish an author’s involvement such that majority contributions are more reliably captured. However, would need to be counterbalanced in such a way so as to not discourage collaboration between researchers.
- On metrics, there is recognition that appropriate use of quantitative data – coupled with the necessary infrastructure to capture information – could play a useful role in streamlining the REF process, reducing cost and administration. Enhanced use of metrics, such as grant income

or citations for example, would also provide a more objective and transparent assessment of research quality. There may also be merit in exploring the use of “ratio” metrics (such as output/impact per lab/grant received) so as to mitigate the output difference between big and small labs.

- However, we believe no system of quantitative assessment can yet match the fidelity of judgement achieved through peer review. Peer review is the foundation on which scientific assessment is based and commands widespread respect as a sound and effective mechanism for assessing scholarly output. This is despite inherent flaws around the subjective nature of peer review, the fact that it is slow, inefficient and expensive, and its capacity to lead to scoring inconsistencies between different Units of Assessment (UoAs).
- However no one system is perfect and a combination of metrics and peer review is perhaps the most effective means of achieving a reliable indicator of output quality. We would therefore agree with the findings of the Independent Review of the Role of Metrics in Research Assessment that a “variable geometry of expert judgement, quantitative and qualitative indicators” is the best approach to the assessment of research outcomes.²
- The use of altmetrics and their potential value as a complementary tool in evaluating research output is gaining momentum. Indeed, evidence has shown a positive correlation between social media based altmetrics and bibliometric indicators.³ However, we believe such metrics are not yet at a stage where they are a reliable indicator of research quality. They depend on whether the subject of the paper is something that the public are concerned about, or which is easily understood. Other factors, such as promotion from university press offices for example, can also influence altmetric outputs. Further research is therefore required to disentangle the value and applicability of altmetric data across a broader range of research outputs.

Key points

- We have some concern that the definition of impact disadvantages valuable basic science and that this could have negative consequences for the UK and its research base. We would encourage a broader definition of impact, one that more reliably captures fundamental research, and discourage any move to attach greater weighting to the impact portion of the REF score.
- It is important that the REF accurately capture output that is achieved through a consortium of authors and that “substantial contribution” of individual authors is appropriately gauged. We would encourage more guidance on how this is achieved.
- Peer review is the gold standard of academic assessment and has an important role to play, in combination with metrics, in assessing research output. New metrics, such as ratio-metrics, may be a useful addition to the exercise. In contrast, altmetrics are yet to reach a stage where they are a reliable indicator of research quality and we would not encourage their use until further research is applied to disentangle their value and applicability across a broader range of research outputs.

2. If REF is mainly a tool to allocate QR at institutional level, what is the benefit of organising an exercise over as many Units of Assessment as in REF 2014, or in having returns linking outputs in particular investigators? Would there be advantages in reporting on some dimensions of the REF (e.g. impact and/or environment) at a more aggregate or institutional level?

- There is some merit to organising the REF over so many UoAs. For example, it provides institutions with detailed information about their research output across a broad range of disciplines. This in turn enables universities to benchmark their research performance and compare it on both a national and international scale. An improved understanding of the research strengths within an institution has utility from a strategic perspective but can also be used by organisations to communicate their strengths to external audiences, including students, potential collaborative partners in the public and private sectors, and prospective new staff.
- However, maintaining scoring consistency is a challenge across so many UoAs. We have heard concern that the scoring of outputs across the UoAs can vary considerably. This perception is not encouraged by the opaque nature of the assessment process and we would strongly encourage moves to build more transparency into the assessment exercise. Panels offering detailed individual feedback would be an important way of achieving this, for example.
- Aggregating data is useful in some contexts, primarily because it is easier to understand and interpret. However, aggregating output quality comes with a loss of granularity that offsets the advantages of a detailed dataset which can more clearly identify pockets of excellence within an institution. Moreover, reporting at an aggregate level may prompt some to query why it is necessary to expend significant time and money to collate data at the level of the individual.

Key points

- We would strongly encourage the provision of more detailed feedback for individual researchers so as to build more transparency into the assessment exercise. Our members have reported that they would find this very useful.

3. What use is made of the information gathered through REF in decision making and strategic planning in your organisation? What information could be more useful? Does REF information duplicate or take priority over other management information?

- We are aware that the REF can serve as an important tool in informing decision making at universities, especially on research strategy. For example, REF data can be used to evidence

investment decisions in specific areas of research or prompt reviews on how research is managed or performed both at the departmental and institutional levels.

- An unwelcome aspect of the REF has been its use as a staff appraisal tool. We have heard concern that scores have influenced management decisions regarding the promotion or dismissal of individuals within or from an organisation, respectively. This unfairly disadvantages researchers who work collaboratively or have a smaller number of high impact papers. Similarly, REF has been a factor in the strategic recruitment of staff by universities looking to “poach” staff who could increase their scores.

Key points

- It is somewhat regrettable that REF is now widely used for HR purposes as a staff appraisal tool. This unfairly disadvantages researchers who work collaboratively or have a smaller number of high impact papers. The “transfer window” created as a result of universities attempting to “poach” staff who can increase their scores is also an unsettling influence.

4. What data should REF collect to be of greater support to Government and research funders in driving research excellence and productivity?

- As previously stated we believe it is important to recognise excellence in a broad research base that acknowledges the essential contribution of fundamental research in addition to translational and commercial activities.
- However, there are other contributions that researchers make to the academic community which are important drivers of research excellence and productivity. These include, for example, contributions to conferences (organisation and presentation) and external examination (e.g. for PhDs).

Key points

- Additional data such as conference activity and contributions to external examination could comprise additional information which would support research funders in driving academic excellence.

5. How might the REF be further refined or used by Government to incentivise constructive and creative behaviours such as promoting interdisciplinary research, collaboration between universities, and/or collaboration between universities and other public or private sector bodies?

- Much of the immunology ongoing in UK institutions is highly collaborative and interdisciplinary already, a state which has arisen largely organically. There is also evidence the REF process in itself has encouraged collaboration within institutions. It builds an institutional awareness of the organisation's research culture, encourages researchers to consider how their research portfolio fits within the wider activities of the institution (and indeed, thanks to impact analysis, wider society), and in doing so enables the identification of potential synergies along recognised research strengths. This is similarly true for research collaboration outside the organisation, which may develop as a result of the reputational benefits which are derived through strong REF scores in a particular area, or has been driven, in the case of collaboration with industry, through the push for greater "impact".
- However, there is no escaping the fact that the REF is ultimately an exercise that creates rank orderings. This breeds competition, and as stated previously, we would support further guidance on the factors which qualify individual's contribution levels. We are also in favour of some form of grading to assess contribution. Clarity over the assessment of multi-author outputs would be of significant benefit to the REF process.

Key points

- Much of the immunology ongoing in UK institutions is highly collaborative by its very nature. The REF has influenced this to a degree, including through encouraging collaboration with the private sector through the assessment of impact. However, there may be disadvantages in "forcing" these relationships further as opposed to allowing them to arise organically.

6. In your view how does the REF process influence, positively or negatively, the choices of individual researchers and/or higher education institutions? What are the reasons for this and what are the effects? How do such effects of the REF compare with effects of other drivers in the system (e.g. success for individuals in international career markets, or for universities in global rankings)? What suggestions would you have to restrict gaming in the system?

- The REF incentivises universities to make judgements on staff based on their potential to return high scoring outputs. Whether or not this is widespread, there is certainly a perception amongst many researchers that the REF can make or break careers. This is especially true for younger academics who may feel pressurised by the sense that their career depends on their ability to return high REF scores.
- This, coupled with the onerous requirements of the exercise, can mean that the REF exercise has a negative perception amongst research staff.
- We are also aware that the REF plays a role in recruitment practice, with institutions targeting investigators likely to enhance potential REF returns. This can have perverse outcomes in encouraging short-term or part-time hires in an attempt to game the system. One potential

solution around this would be to ensure staff selected for the REF have been in place for at least 3 years prior to the REF census date or by introducing fractional sampling, where all staff from a department submit data but only a subsection would be selected for assessment.

- The emphasis around demonstrating impact can also have negative consequences. It penalises researchers undertaking creative and theoretical work and pushes them towards activities that fit more closely with the definition of “reach and significance” (and are therefore more likely to return a higher score). This may mean investigators undertaking fundamental research are appraised more critically by their institution because they generate less income.

Key points

- The perverse outcomes in university recruitment practice that arise through the REF may be resolved through ensuring that the staff selected have been in place for at least 3 years prior to the REF census date. Alternatively, fractional selection may also have a role to play, where staff are selected at random for analysis.

7. In your view how does the REF process influence the development of academic disciplines or impact upon other areas of scholarly activity relative to other factors? What changes would create or sustain positive influences in the future?

- The REF can have an influence over publication strategy, most notably with review articles. These are not counted within the assessment and so there is no motivation to publish review articles in the context of the REF. Reviews, especially peer reviewed reviews, continue to have an important place in the academic literature and it would be harmful for the scientific community as a whole if such publications were widely disregarded.

Key points

- Reviews, especially peer reviewed reviews, continue to have an important place in the academic literature and should be included as part of the assessment.

8. How can the REF better address the future plans of institutions and how they will utilise QR funding obtained through the exercise?

- The REF allows for a more informed approach to strategic planning and this can help universities ensure the best ‘value for money’ in the allocation of their QR funding. However, there is a danger that by assigning funds based on retrospective performance existing

inequalities between – and sometimes within institutions – are exacerbated if funds are simply allocated along existing strengths.

Key points

- There is a risk that if universities assign QR funding along existing strengths, as researched by REF performance, inequalities both within and between institutions can be exacerbated. However, the REF exercise does have utility in informing investment decisions such that institutions can invest where they feel they will achieve the best value for money.

9. Are there additional issues you would like to bring to the attention of the Review?

- It may be worth exploring the possibility of submitting data at the level of the department/faculty as opposed to the individual. This would mitigate the negative associations around individual reporting of REF data while still allowing universities to identify key research strengths. It also has the advantage of communicating the importance of team science.

Key points

- There may be merit in exploring the benefits of submitting data at the level of the department/faculty. This could help mitigate negative associations around individual reporting, especially protecting Early Career Researchers, and also has the advantage of communicating the importance of team science.

References

¹ [Department for Business, Innovation and Skills \(2013\). SET Statistics 2013.](#)

² [Wisldon et al. \(2015\). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management.](#)

³ [Priem J, Piwowar, H, and Hemminger, B. \(2012\). Altmetrics in the Wild: Using Social Media to Explore Scholarly Impact](#)