

A Stakeholder Response to the Joint Funding Bodies' Invitation to Contribute to the Review of Research Assessment compiled by the Wellcome Trust

Introduction

1. The Wellcome Trust ("the Trust") is an independent, biomedical-research-funding charity established under the will of Sir Henry Wellcome in 1936. The Trust is the largest non-Governmental funder of biomedical research in the world. In the financial year ended 30 September 2002, the Trust's expenditure was more than £500 million, the majority of which was in the UK. This balance of funding is likely to continue whilst the excellence of the UK science base is maintained and we believe research assessment and the reward of excellence is key to this.
2. We therefore welcome the opportunity to contribute to the Review of Research Assessment. This response was produced through consultation with a variety of Trust stakeholders, including Trust staff, members of the Trust's Board of Governors and a number of senior academics from our expert panels. The consultation took the form of a workshop, following the suggested format presented in the Review team's invitation document. **As such, it does not represent the Trust's corporate view, but the collected ideas and concerns of a number of our key stakeholders.**
3. This response details general issues and concerns about research assessment, thoughts on the individual assessment methods and issues surrounding the cross-cutting themes.

The RAE general comments

4. The starting point for discussions was that the Research Assessment Exercise had increased the quality of research in the UK. However, a number of themes emerged which should be considered in the development of any future research assessment exercise:
 - *Avoiding perverse drivers.* It was thought that the current system incorporated a number of perverse drivers, which should be avoided in the development of new techniques. Most importantly, it was felt that the current RAE encouraged increased volume of research activity at the expense of investment in research infrastructure. The 5 year cycle of the RAE was also believed to encourage short-term research strategies, deterring people from more complex long-term research challenges. It was also perceived that the RAE had promoted the reduction in number of UK based journals, as impact factors were greater for USA journals and reduced the influence of UK academic associations.
 - *Reducing game-playing.* It was recognised that institutions were becoming increasingly sophisticated in their submissions to the RAE. The discretion afforded to institutions was being exploited to ensure the best outcome, particularly in relation to the number of individuals submitted and allocation of individuals to particular units of assessment (UoA). Our discussants felt that assessment methods should use objective output measures where possible to provide an honest indication of quality and productivity.
 - *Consistency.* The activities of individual assessment panels were thought to have been very different and the subsequent ratings were not felt to be comparable across UoA. It was perceived that the criteria used to assess publications had varied considerably. For example, it was felt that some panels were more strongly influenced by impact factors than others.
 - *Reducing frequency of assessment.* It was generally felt that a 5-year cycle was too short for successive research assessments. A longer period between assessments would be more cost effective, increase stability and allow long-term research strategies to be developed.
 - *Protecting existing excellence.* It was noted that the current RAE does not protect or reward small centres of excellence within larger groups of less excellent people and has often lead to the demise of such groups. The problems following the 2001 RAE were highlighted where institutions have seen dramatic changes in their levels of funding with none or only minor changes in their research quality in the absence of growing resources.
5. A number of fundamental issues concerning how any future research assessment process might be operationalised were also discussed, most importantly:

- *Assessing teams vs assessing individuals.* The workshop attendees did not reach consensus on this issue. It was thought that requiring all staff or a designated proportion to submit may reduce game-playing. It was also felt that assessing individuals would allow funds to be allocated more fairly through banded scores (e.g. 15% of staff at 5*; 25% of staff at 5; 40% of staff at 4; remainder below 4). Funding could still be allocated formulaically as at present with banded scores of this kind, and this would protect small centres of excellence within larger groups of less excellent people.
- *Assessing all publications.* It was thought that some measure of productivity (number of papers) was needed, alongside quality measures. However, it was acknowledged that these measures would have to be normalised to account for differing scientific areas.
- *Reducing the number of Units of Assessment.* The general consensus was that this would be unworkable. There is already concern that some panels were constituted adequately to judge the quality of areas within their UoA and aggregating UoA would make it progressively more difficult to judge the quality of research across a wide range of disciplines.
- *Administrative burden.* It was thought that the current system was very labour intensive for both those undertaking the assessments and those making submissions, although there was a feeling that technology could be improved to simplify the process. It was also noted that the software that is used to create submissions could be more user friendly.
- *Clinical research.* It was noted that the RAE does not recognise that clinical research has a different set of problem to basic research and these differences need to be taken into account by the research assessment process. It was also thought that the RAE had led to a divergence in university based clinical research from NHS based work, the quality of university based clinical research being higher.

Assessment Methods

A. Expert review

6. Expert review was thought to be the most useful method of assessing the quality of research. When used alongside algorithmic methods it was felt that experts play an important and necessary role ensuring the correct interpretation of metrics for specific fields. It was noted that the constitution of the panels was very important to ensure fair assessment of all disciplines within a UoA.
7. A number of suggestions were made for how the current process could evolve, although it was acknowledged that each had its own drawbacks. These methods included:
 - Expert panel undertaking site visits as opposed to simply reviewing paper submissions;
 - A representative of a research group being present during the review of paper submissions to clarify any issues that may arise;
 - A written consultation to allow those being assessed to address feedback and panel decisions; and
 - Closer examination of those groups on the borderlines of ratings.
8. It was agreed that panels should assess work retrospectively, rather than be used to review prospective strategic bids for funding.

B. Algorithms

9. Algorithms were generally thought to be a useful tool in the assessment of research, but one that should not be used in isolation. It was thought that they could be used in a variety of new ways, particularly in association with expert review to accomplish a lengthening in the assessment cycle. This could be achieved for example by having expert panel review every 10 years alongside a 5 yearly algorithm based assessment to determine if research groups are maintaining the same levels of quality.
10. Other suggestions included a system where groups only entered in to the review process if they thought they would move to a higher rating, with algorithm techniques used to determine if research groups are maintaining the same levels of quality.
11. Publications were seen as a key metric for use in algorithmic methods, as they build some degree of peer review into the system. It was believed that bibliometric methods may play a key role in reducing the workload of expert panels and help assist in the assessment of interdisciplinary work for which

panels may not have sufficient expertise. However, it was thought that bibliometric methods need to be used with careful interpretation and expert validation for a number of reasons, including:

- *Subject specific differences.* The research process differs between subject areas and measures need to be developed that are subject specific and fit the needs and opportunities of particular assessment areas (this would include permitting longer reporting periods within areas where books are the norm, as well as using different measures in different areas).
- *Impact factors can be misleading.* Some subjects (e.g. surgery) are difficult to get published in the highest impact journals. It was also noted that not all papers in the same journal are of equal status and quality. Lastly it was noted that pieces of 'wrong' research are often highly cited.
- *Teams.* Some research requires large teams (e.g. bioinformatics) and it was questioned how we ensure that research resulting in papers with a large number of authors are rewarded appropriately.
- *Outputs – type and quantity.* The research process has different time-scales in different disciplines and may also result in differing end products (e.g. books). It was also believed that measures should try to include both quality and volume of output but these must be normalised for scientific area. It was questioned whether an analysis of review articles or those publications that are highlighted in journal editorials could be used to unpick the differences in quality between current 5/5* rated research groups.

12. Other metrics were considered by the group but the general consensus was that none alone were suitable for assessment of excellence for the following reasons:

- *Research student numbers.* This was thought to be an inadequate metric as the method of allocating studentships was undergoing considerable change within the Research Councils. It was also perceived that universities varied in their ability to raise their own scholarships.
- *External research income.* It was felt that only peer reviewed research income should be included, but it was also questioned whether the link between research quality and research income was really that strong.
- *Reputation.* It was felt that reputation had no role to play in the assessment of research excellence as assessment of individuals out of their departmental/group context was difficult and highly subjective.

13. It was thought by the group that the Review Team should undertake some modeling exercises to explore how algorithm methods might be better used in the future.

C. Self-assessment

14. This method was seen to be too reminiscent of the Quality Assurance Agency (QAA) teaching assessment and was not discussed at any length.

D. Historical ratings

15. The group felt this was not a unique method and that the discussions of expert review and algorithms had covered options for historical review.

Cross-cutting themes

What is excellence in research?

16. It was acknowledged that publications and expert review are key indicators of excellence, it was thought that excellence should be measured by a diversity of outputs (e.g. including activities such as patent applications, biomedical resources).

17. There was discussion around the concepts of international and national excellence. The concept of national excellence was questioned. It was thought by the group that work, which had national standing and relevance was probably in fact 'internationally excellent' though not necessarily of relevance internationally.

What should/could an assessment of the research base be used for?

18. It was thought by the group that given resource constraints that the primary use of an assessment of quality of the research base should be to provide a fair and transparent mechanism for the allocation of funding.

How often should research be assessed? Should it be on a rolling basis?

19. It was thought that the period between assessments should be lengthened for all participants. A rolling programme of assessments was thought to be unworkable. It would create a long-term burden on assessment teams and would unfairly give some groups a little time and others a long period to develop submissions.

Should research assessment determine the proportion of the available funding directed towards each subject?

20. It was believed that the assessment process should ensure that there is adequate funding for each subject area especially those that have greater research costs such as the biomedical sciences. It was also suggested that information emerging from the transparency review could be used to provide more accurate details of the costs of research in the various subject fields.

Should each institution be assessed in the same way?

21. As was noted earlier there may be ways of developing more flexible assessment methods allowing institutions to lengthen the period between assessment procedures.

Should each subject or group of cognate subjects be assessed in the same way?

22. As noted earlier measures need to be developed that are subject specific and fit the needs and opportunities of particular assessment areas, this may include using different measures in different subject areas.

How much discretion should institutions have in putting together their submissions?

23. As described the group felt that the discretion afforded in the current system has allowed sophisticated game-playing to occur. It was suggested that the next research assessment process should require all members of staff to be submitted. This would eliminate guess work about the relative trade-offs between quality ratings and volume factors, and ensures that ratings really represent the achievements of all the people within the unit of assessment.

How can a research assessment process be designed to support equality of treatment for all groups of staff in Higher Education?

24. While it was thought that the primary purpose of a research assessment process should be to provide a fair and transparent mechanism for the allocation of funding, it was highlighted by a number of participants that a future system could be used to drive good practice in this area.