HIGHER EDUCATION FUNDING COUNCIL FOR ENGLAND

“Bibliometrics and the REF”
&
“REF equalities analysis”

Mark Gittoes: m.gittoes@hefce.ac.uk

David Mawdsley: d.mawdsley@hefce.ac.uk

Hannah White: h.white@hefce.ac.uk

Research Funding

- The allocation of over £1.5bn of funding is driven by Universities' research performance
- Policy to focus on the highest quality research
- Previously allocated by periodic peer review exercises (the RAE)
- Push towards a (purely) metrics based allocation in the sciences (the REF).
- After an extensive pilot exercise we anticipate a system of peer review informed by metrics

Two parts

- Discussion on two technical aspects of HEFCE's preparatory work on the use of metrics in the REF:
  - Exploration of particular metric measures
  - Use of metrics in equality analysis
- Not intended to be a complete overview of the REF or HEFCE's approach to research assessment.

HIGHER EDUCATION FUNDING COUNCIL FOR ENGLAND

Bibliometrics and the REF

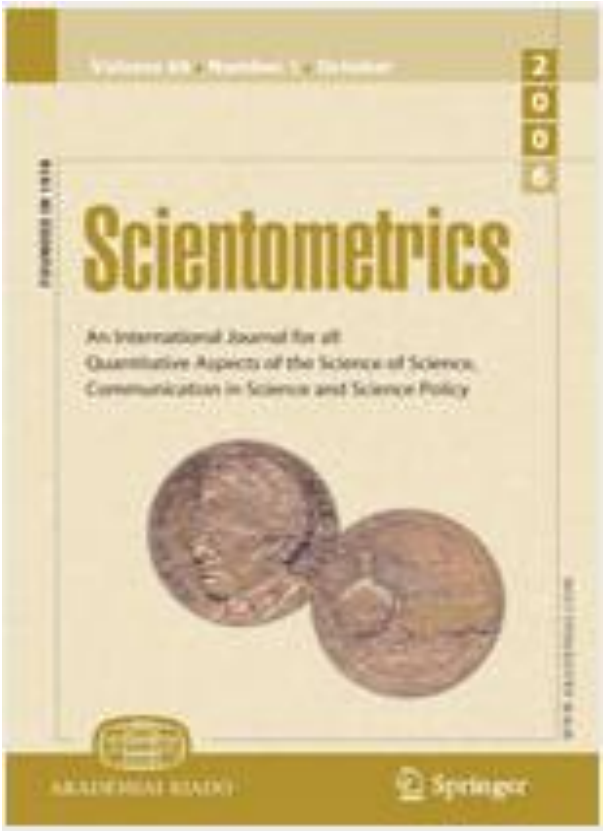
David Mawdsley

Contents

- Overview of bibliometrics
- How we adjust between subjects to present fairer picture
- How best to present information to groups of peer reviewers

How bibliometrics work…

- Take a journal…
- …and a paper inside it
- The bibliography references other papers



References

ADAMS, J. (1998), Benchmarking international research. *Nature*, 396 : 615–618.

ADAMS, J., BAILEY, T., JACKSON, L., SCOTT, P., SMALL, H., PENDLEBURY, D. (1998), *Benchmarking of the International Standing of Research in England – a Consultancy Study for the Higher Education Funding Council for England.* CPSE, University of Leeds. 108 pp. ISBN 1 901981 04 5.

GLÄNZEL, W., MOED, H. F. (2002), Journal impact measures in bibliometric research. *Scientometrics*, 53 : 171–193.

HIRST, G. (1978), Discipline Impact factor: a method for determining core journal lists. *Journal of the American Society for Information Science*, 29 : 171–172.

KOSTOFF, R. (2002), Citation analysis of research performer quality. *Scientometrics*, 53 : 49–71.

MOED, H. F., BURGER, W. J. M., FRANKFORT, J. G., VAN RAAN, A. F. J. (1985), A comparative study of bibliometric part performance analysis and peer judgment. *Scientometrics*, 8 : 149–159.

PODLUBINY, I. (2005), Comparison of scientific impact expressed by the number of citations in different fields of science. *Scientometrics*, 64 : 95–99.

SCHUBERT, A., BRAUN T. (1993), Reference standards for citation based assessments. *Scientometrics*, 26 : 21–35.

SCHUBERT, A., BRAUN T. (1996), Cross-field normalization of scientometric indicators, *Scientometrics*, 36 : 311–324.

VAN RAAN, A. J. (2006), Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment. *Scientometrics*, 67 : 491–502.

ZITT, M., RAMANANA-RAHARY, S., BASSECOULARD, E. (2005), Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalization. *Scientometrics*, 63 : 373–401.

Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 67, No. 3 (2006) 491–502
DOI: 10.1556/Scient.67.2006.3.10

Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups

ANTHONY F. J. VAN RAAN

Center for Science and Technology Studies, Leiden University, Leiden (The Netherlands)

In this paper we present characteristics of the statistical correlation between the Hirsch (h-) index and several standard bibliometric indicators, as well as with the results of peer review judgment. We use the results of a large evaluation study of 147 university chemistry research groups in the Netherlands covering the work of about 700 senior researchers during the period 1991–2000. Thus, we deal with research groups rather than individual scientists, as we consider the research group as the most important work floor unit in research, particularly in the natural sciences. Furthermore, we restrict the citation period to a three-year window instead of 'life time counts' in order to focus on the impact of recent work and thus on current research performance. Results show that the h-index and our bibliometric 'crown indicator' both relate in a quite comparable way with peer judgments. But for smaller groups in fields with 'less heavy citation traffic' the crown indicator appears to be a more appropriate measure of research performance.

Introduction

In a recent paper, HIRSCH (2005) proposes an original, simple new indicator to characterize the cumulative impact of the research work of individual scientists: 'a scientist has index h if h of his/her N papers have at least h citations each, and the other (N-h) papers have no more than h citations each'.[1]

From the above definition follows that h is a measure of the absolute 'volume' of citations whereby h^2 provides an estimation of the total number of citations received by a researchers. Given the very skewed distribution of citations (C) over publications (P) described by a power law $P(C) = \alpha C^{-\beta}$ (VAN RAAN, 2006), particularly for the higher-C

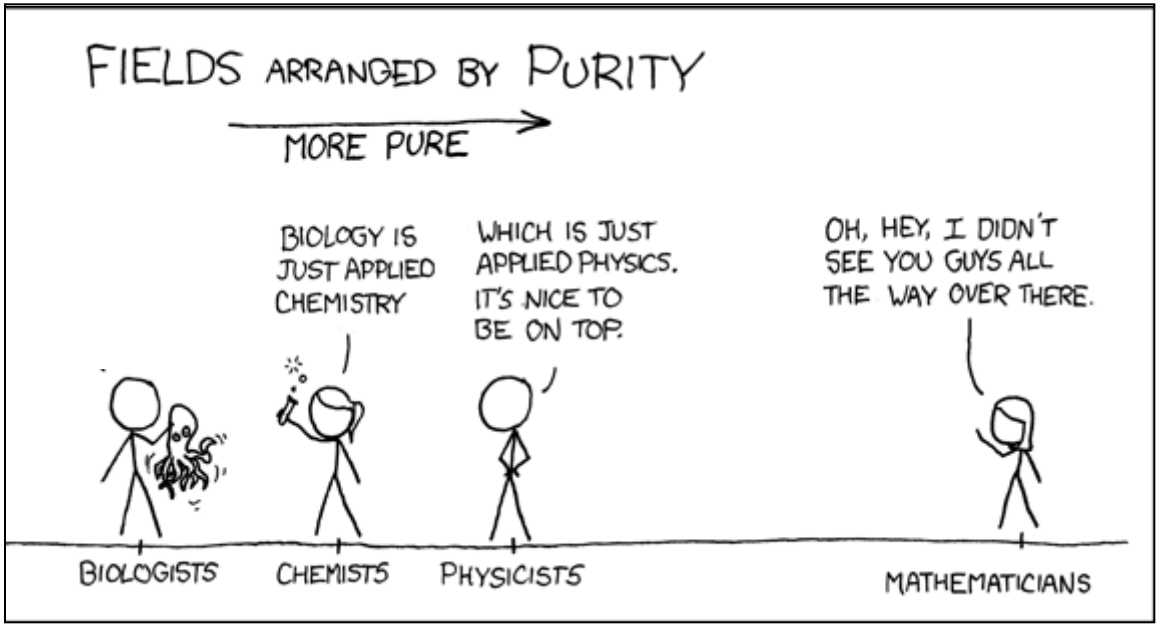
[1] For instance, if a scientist has 21 papers, 20 of which are cited 20 times, and the 21st is cited 21 times, there are 20 papers (including the one with 21 citations) having at least 20 citations, and the remaining paper has no more than 20 citations.

Received December 1, 2005

Address for correspondence:
ANTHONY F. J. VAN RAAN
Center for Science and Technology Studies, Leiden University
Wassenaarseweg 52, P. O. Box 9555, 2300 RB Leiden, The Netherlands
E-mail: vanraan@cwts.leidenuniv.nl

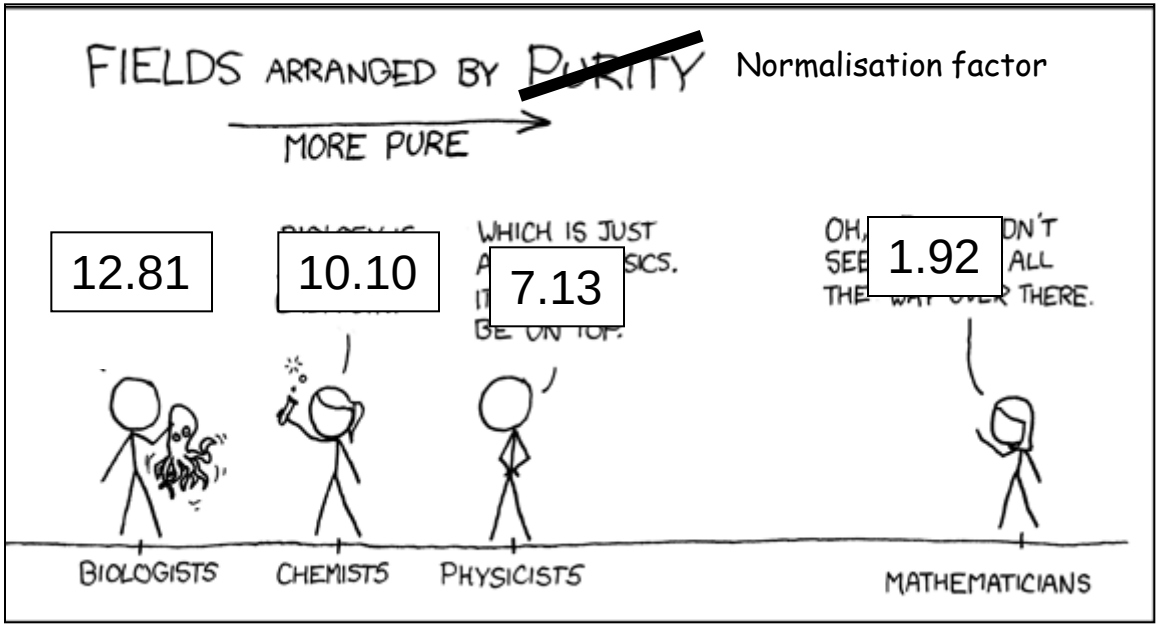
0138–9130/US $20.00

- What's a citation worth?
  - Old documents tend to be cited more
  - Some types of document are cited more
  - Subject dependence



http://xkcd.com/435/

- What's a citation worth?
  - Old documents tend to be cited more
  - Some types of document are cited more
  - Subject dependence

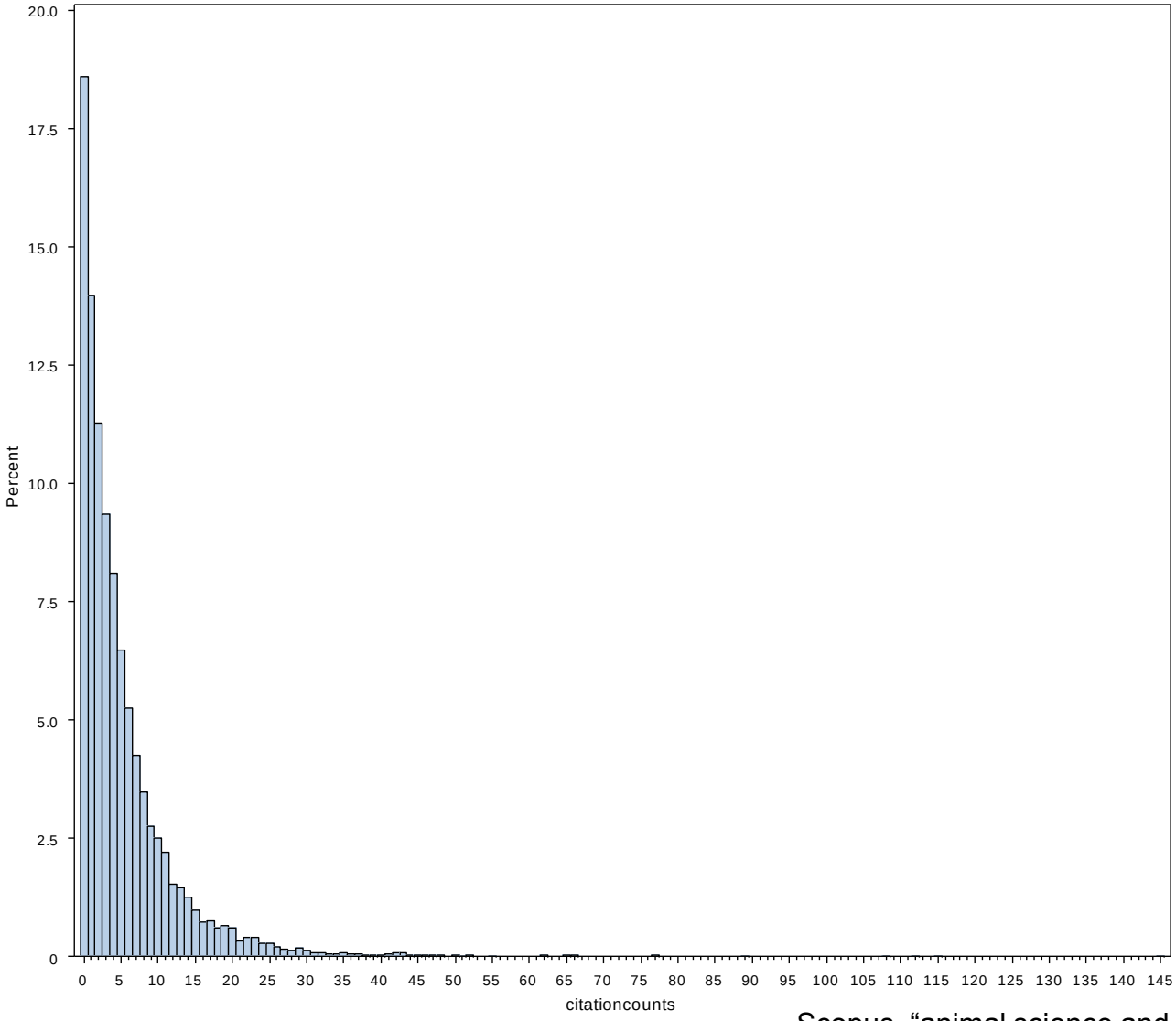


Factors are for articles in English published in 2004.
Subject categories are biochemistry & molecular biology, chemistry (multidisciplinary), Physics (multidisciplinary), and mathematics

Computing a normalised citation score (NCS)

- Number of citations
- Divided by the average number of citations for its *peer group*:
  - All documents of the same type (article, review, etc)
  - Published in the same year
  - In the same subject category
- A normalised citation count of 1 = world average

Distributions of citations



Scopus, “animal science and zoology”, articles, published 2003

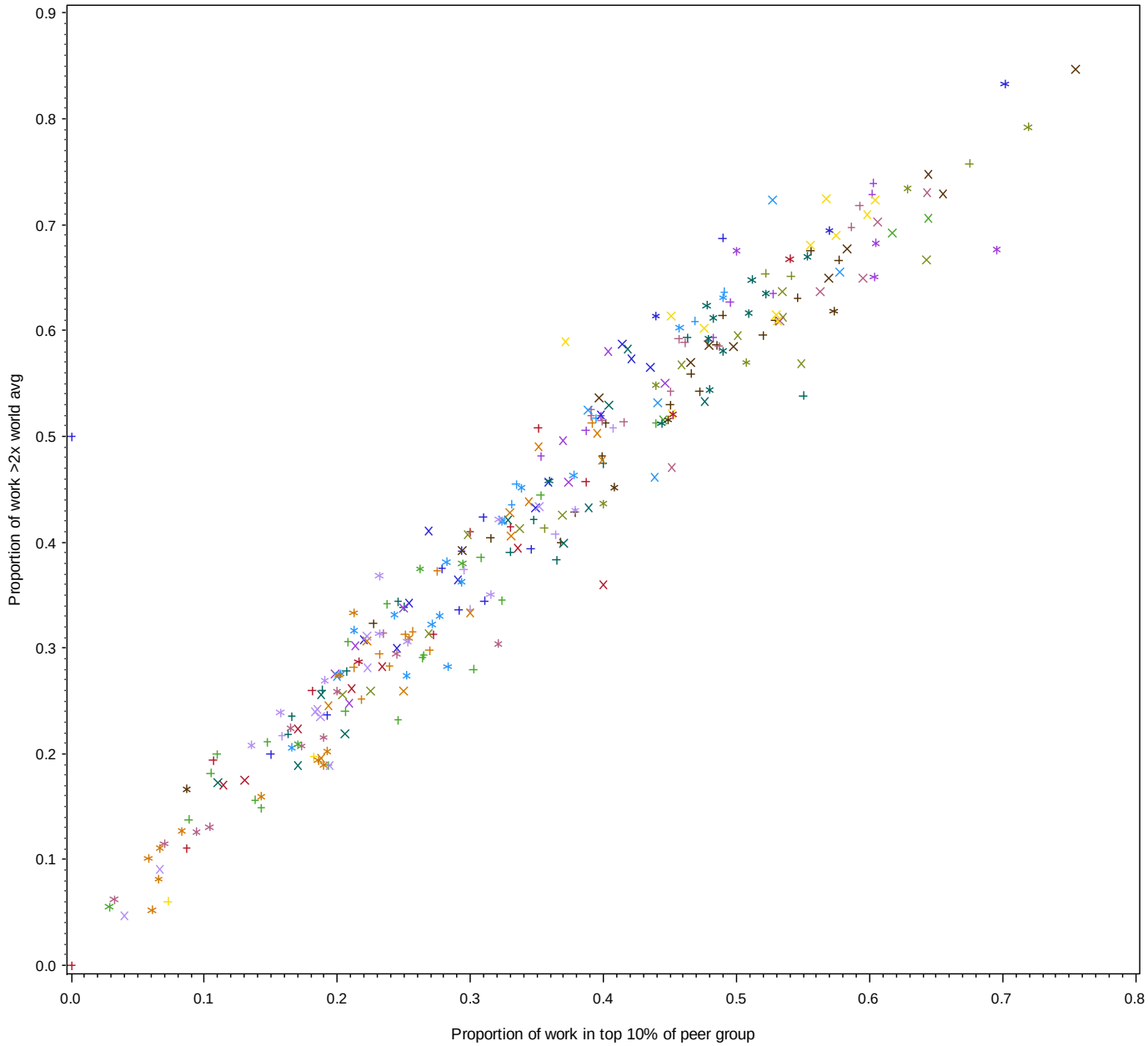
How do we benchmark against this?

- Typically the mean is used
  - Not ideal
- “Department” level indicators:
  - Mean NCS
  - Proportion of work above a threshold
- Percentiles

Percentiles

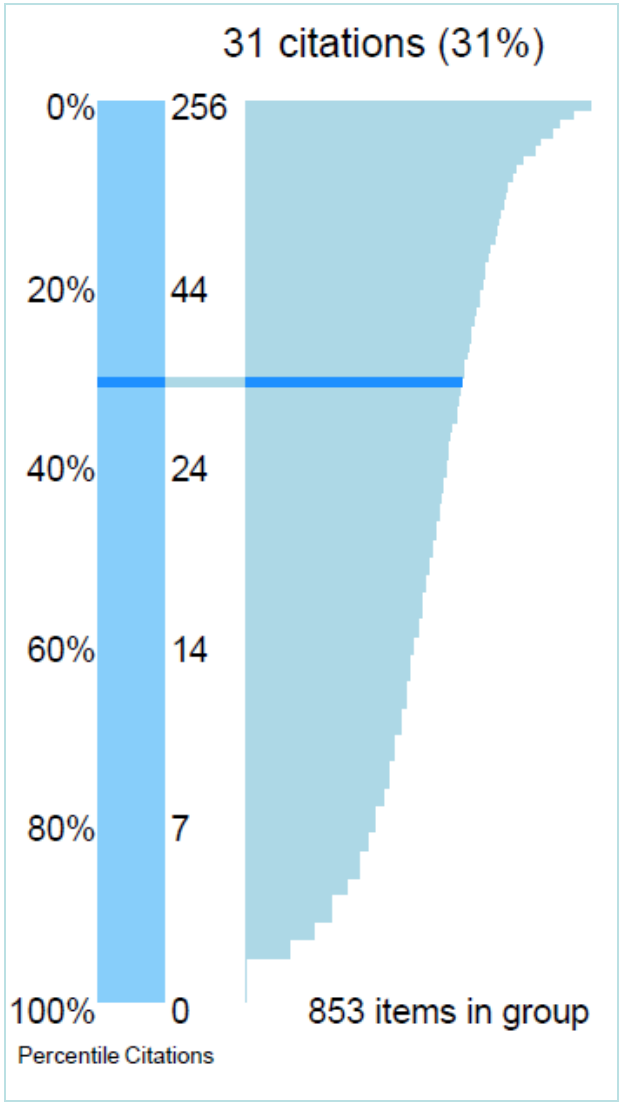
- For each *peer group*, we rank all the papers by the number of citations
- For each paper we report which percentile it is in.

- For each department we then calculate the proportion of items that are in the top 10% of their *peer groups*.


Proportion of work >2x world avg
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Proportion of work in top 10% of peer group
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

To inform peer review…

- As we're not using formulaically the *peer group* is less important
- Compare the output's score against two *peer groups*
  - Everything else within the UOA
  - Everything else within the journal
- Communicate visually



HIGHER EDUCATION FUNDING COUNCIL FOR ENGLAND

REF equalities analysis

Hannah White

Motivation

Bibliometrics move away from the peer review method used in previous RAEs. This equalities analysis looked at addressing the question:

- Were the equalities differences observed in the RAE still present when bibliometric measures were used?

History

Equalities analyses were carried out for staff selection to RAE2001 and RAE2008. Findings included:

- Men were more likely to be selected than women and modelling showed this to be significant over the age range 30-50.
- Staff with non-UK nationality were significantly more likely to be selected than staff with UK nationality.

Comparison

RAE

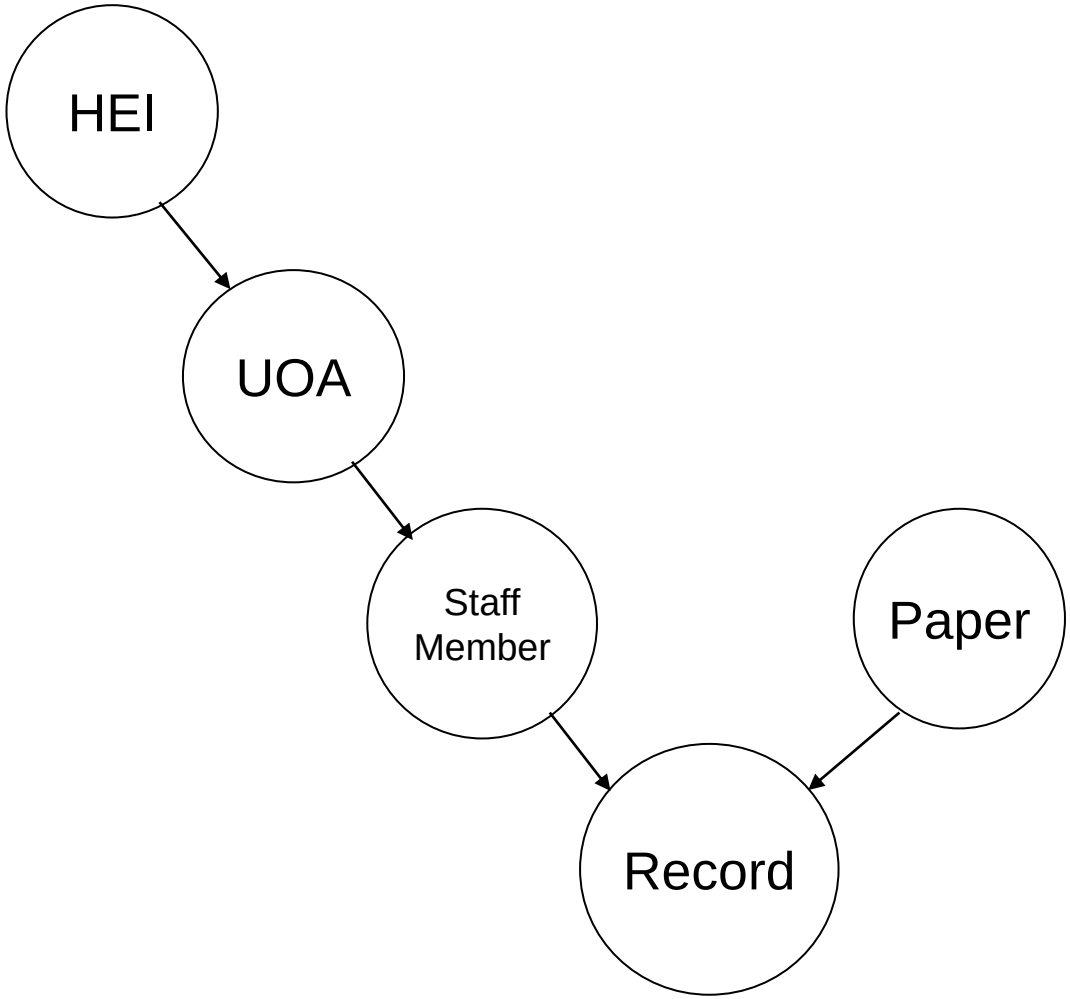
- Data analysed after the assessment
- Compared the selection rates of groups of staff

REF

- Data analysed to inform development of REF
- Compared the citation scores of groups of staff previously selected for the RAE

REF model structure

Cross-classification multi-level model



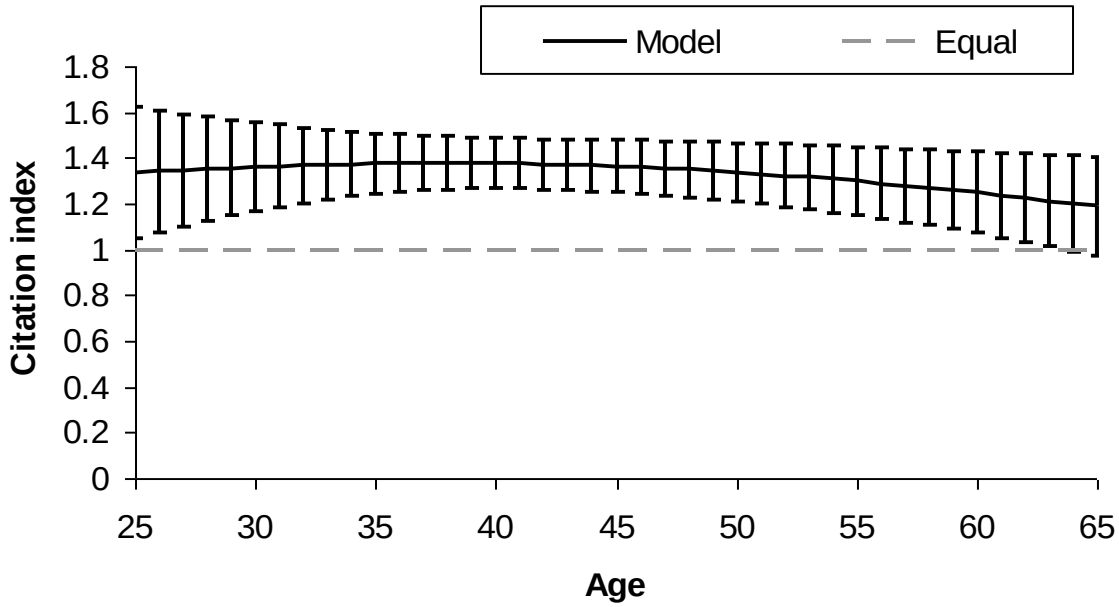
Model variables

The attributes simultaneously allowed for in the statistical models were:

- age; gender; ethnicity; nationality; disability; ECR status;
- mode of employment (part-time or full-time); subject area; clinical status; contract status (permanent, fixed-term or atypical); employment function (research and/or teaching); senior position holder; grade; member of ECR department;
- paper; institution; unit of assessment; person;

Findings

- Age and sex



- Nationality – There were no significant differences found between UK and non-UK nationals in the proportion achieving the threshold.

Conclusions

Age and sex differences possibly:

- Men are producing better quality papers which are then more highly cited than women.
- There is some systematic bias towards men within the research assessment or production process.

Nationality differences:

- The difference in selection rate is likely to be based on the quality of papers produced by the two groups of staff rather than selection bias.

Summary (1)

- Pitfalls and caveats with all bibliometric measures
- Key challenge is how these interact with REF expert panels
- Aiming for consistency of information
- But not necessarily one size fits all

Summary (2)

- Consultation document:
  - HEFCE publication 38 of 2009;
  - Assessing output quality;
  - Central group of equality specialists and REF panel members
- Spring 2010:
  - Announce consultation outcomes;
  - Invite nominations for panels;
  - Development of REF data collection systems.