

# Measuring research impact: not everything that can be counted counts, and not everything that counts can be counted

**Jane Grimson<sup>1</sup>**

School of Computer Science and Statistics, Trinity College Dublin,  
The University of Dublin, Ireland

## Introduction

Virtually every sector of society and the economy across the world is subject to ongoing monitoring and assessment through an increasingly complex and bewildering array of measurements, metrics and key performance indicators. All of this is being carried out in the name of openness, transparency and accountability. These performance indicators are being used internationally to benchmark countries against each other, nationally to inform government policy and locally to improve performance within an organization, from increasing shareholder value in a company, to reducing waiting times for elective surgery in a hospital. Even the Millennium Development Goals have given rise to a whole new industry in M&E (Monitoring and Evaluation) of the effectiveness of international aid.

The world of research is no different, and international rankings, citations and impact factors have an increasingly important influence on deciding where investment should be made, whether someone should be promoted and even whether or not they are eligible to apply for a grant.

Although measurements of research quality (bibliometrics) have been around for over 100 years, it is really only in recent years that they have begun to play such an important role in academic life. However, virtually everyone agrees that even the most sophisticated measurement is, at best, a poor proxy for evaluating the richness and the diversity of research and of its social and economic impact. New tools, enabled by web technologies, which are emerging together with the older more traditional citation and impact counts, may offer more comprehensive approaches which are less open to unintended consequences, gaming and perverse incentives.

This chapter does not attempt to conduct an exhaustive study of the issues surrounding the use of bibliometrics as a tool (or set of tools) for the measurement of research quality; rather, it draws parallels with the field of healthcare quality and investigates whether there are any lessons to be learnt which could improve our approach to assessing research quality.

<sup>1</sup>Email: jane.grimson@tcd.ie

## Transparency, accountability and trust

Where public expenditure is concerned, the demands for transparency and accountability have increased dramatically in recent years [1] and have been given a further impetus in the current economic climate of cut-backs and austerity. No sector is immune. For the research sector of higher education, the impact has been dramatic with a growing number of articles with catchy titles such as *Users, narcissism and control - tracking the impact of scholarly publications in the 21st Century* [2], *Bibliometrics as weapons of mass citation* [3] and *The follies of citation indices and academic ranking lists* [4]. Although much that has been published has been well researched and evidence-based, and makes recommendations for improved methods of assessing research quality, others simply restrict themselves to a detailed analysis of the inappropriateness of the metrics proposed without putting forward any alternatives, with the remainder representing what could be loosely described as a polemic against such measurement in the name of academic autonomy and independence. That is not to say that academic autonomy and independence are not important; indeed, there are far too many lessons to be learnt from history where their loss has had devastating results for human rights and society as a whole. However, the challenge, as Michael Power puts it, is “balancing the aspiration for autonomy with external pressures for accountability” ([5], p. 2). Power asks whether auditing mechanisms of control (and bibliometrics are essentially a method of auditing the quality of research and ultimately, therefore, of control) are themselves out of control ([5], p. 30) and argues that the burden and cost of the procedures and processes surrounding auditing for both those being audited and the auditors need to be reduced, whereas on the other hand, there needs to be a “rehabilitation of trust” ([5], p. 40).

The issue of trust and the need for a return to it was a major theme of Onora O’Neill’s Reith Lectures in 2002 [6]. O’Neill concluded that, “some of the regimes of accountability and transparency developed over the last 15 years may damage rather than reinforce trustworthiness”. Similarly, Power concluded that the “technologies of audit may paradoxically achieve the opposite of their intended effect” ([5], p. 27). Commenting on the situation in universities, O’Neill described the forms of accountability to which universities are subjected as “a clumsy attempt to achieve accountability for the greatly increased public revenues supporting teaching and research while maintaining respect for academic freedom and university autonomy” [7].

## Measuring healthcare quality

Given this widespread enthusiasm for indicators and performance measurement in general, it is interesting to examine what, if anything, can be learnt from other sectors which might be usefully applied in attempts to measure research quality. Healthcare quality has been a focus of interest and debate within the health sector from the time of Hippocrates, if not before [8]. Today, most researchers and practitioners agree that healthcare quality has six key dimensions, safe, effective, patient-centred, timely, efficient and equitable [9], and all need to be considered

in any attempt at assessment. This led to the development of a variety of metrics or KPIs (Key Performance Indicators). There is a growing body of evidence concerning what makes a 'good' KPI, as shown in Table 1 [10].

Indicators should measure what they are supposed to measure (validity); they should give the same answer if measured by different people (reliability); they should be able to measure small changes (sensitivity); they should measure actual changes (specificity); and they should be underpinned by research (evidence-based) [10].

Today, there are a bewildering array of KPIs which are used in healthcare, many of which are reported internationally for comparison purposes to the World Health Organization and the Organisation for Economic Co-operation and Development. Not surprisingly, as with research indicators and metrics, much has been written about the benefits and risks associated with their use. On the positive side, unless healthcare providers evaluate or measure what they are doing and how, how can they be confident that they are delivering a safe and high-quality service? And everyone agrees that there must be both quantitative and qualitative measurements to give a balanced picture across all six dimensions of quality. On the negative side, however, there is extensive evidence of 'gaming', concerns over data quality, and an over-reliance on targets and indicators leading to unsafe services [11]. One example of 'gaming' which has been identified in a number of countries concerns waiting times in Emergency Departments. In England, for example, a maximum waiting time target of 4 hours for 95% of patients was introduced in 2004, but when hospitals realized that they were going to miss the target, they refused to accept the patients from the ambulances as the clock did not start until the patient was registered in the Department [12]. In order to counteract this, the National Health Service had to start measuring ambulance turnaround times!

It is also the case that some indicators which are widely used internationally have subsequently turned out to be flawed, and yet they continue to be used. An example is the HSMR (Hospital Standardised Mortality Ratio) which is intended to measure, appropriately risk-adjusted, whether a particular hospital has an unusually high mortality rate. It is used as a proxy for the safety of services provided by that hospital. On the face of it, this seems a reasonable measure, and patients have a right to know whether they are at risk. However, following

**Table 1**  
**Criteria for a good KPI**

KPI attribute	Meaning
Validity	Does the KPI measure what it is supposed to measure?
Reliability	Does the KPI provide a consistent measure, i.e. does it give the same answer if measured by different people?
Sensitivity	Can the KPI measure small changes?
Specificity	Can the KPI measure actual changes?
Evidence-based	Is the KPI supported by scientific evidence or the consensus of experts?

systematic reviews of the evidence, a number of researchers have shown that the HSMR is an unreliable measure of the quality of care [13,14]. Thomas and Hofer [14] concluded that the calculation of HSMRs is subject to both systematic and random errors. Pitches et al. [13] pointed out that, as approximately 98% of all patients survive their hospital stay, the HSMR is based on only 2% of the population, and it would not be unreasonable to assume that most of the deaths which did occur were unavoidable and not the result of poor care. Through a series of simulations, they showed that the HSMR is a reliable measure of quality of care only if a minimum of 15% of deaths were unavoidable; any less than that will not be picked up by the HSMR. They concluded that even though it is widely used in a number of countries, the HSMR is simply a bad indicator [15] with poor sensitivity and specificity [16].

Thus in spite of many decades of the use of measurement and metrics in healthcare, there are still concerns as to whether they do differentiate between good and poor quality care. They can certainly measure activity levels, waiting times and so on, but actual health outcomes are much more difficult to measure, even in situations where the outcome is apparently as definitive and measurable as death. Furthermore, where funding decisions and league tables are based on these indicators, “the evidence reveals how the system rewards providers not for preventing adverse events or ill health, but for treating patients, even if the illness is caused by the service itself” [17].

## Implications for measuring research quality

So what, if anything, can be learnt from the healthcare domain which might ensure that the measurement and metrics used to assess research quality are more robust? Is it possible to come up with a set of indicators that really do differentiate good-quality research from poor-quality research, that will not be subject to gaming and that will not distort practice? And our first problem is to define research quality, and to date, there does not seem to be a consensus [18]. Is research quality the same as research excellence? Where does impact fit in? The Impact of Social Sciences Project at the LSE (London School of Economics) divided research impacts into ‘academic impacts’, which are instances when research influences actors in academia or universities and ‘external impacts’, or instances when research influences other actors outside of academia including business, government and civil society [19]. Academic impacts are traditionally measured through bibliometrics, whereas external impacts are measured in a whole variety of different ways ranging from, for example, coverage in the media, to policy influences, patents and spin-off companies. The key question which all of those involved in the evaluation of research personally, institutionally, nationally and internationally seeks to answer is: “What is the scientific and social impact of an individual’s research?” ([2], p. 5). The equivalent question in healthcare is: “What is the impact (or outcome) on the individual the health of an individual and/or population at large?” Outcome is regarded as an important aspect of healthcare quality, and therefore it seems reasonable to include impact as part of research quality, rather than regard it as something separate. In healthcare, a broad view

of outcome is taken, so a similarly broad view of impact should be taken which encompasses knowledge creation in the sense of increasing our understanding as well as knowledge which is potentially useful socially or economically. In both cases also, it is important to consider both the short- and long-term impacts. The long-term impacts of research can be virtually impossible to track using conventional bibliometrics. For example, Hamilton discovered quaternions in 1843, but it was not until the late 20th Century that they found widespread practical application in computer graphics [20].

Just as in healthcare, there is ample evidence of the potential for gaming of indicators from citation cartels to ghost writers and guest authors (see below). In research, bibliometric indicators, such as citation counts, journal impact factors and publication rates, rely for their validity as measures of research quality on the fact that published research has been subject to peer review [18]. Given the widespread criticisms of peer review, this assumption is not necessary always valid. It would be beyond the scope of this chapter to critique the peer-review system. However, there are three issues which are particularly relevant to the quality debate, namely the inherent conservatism, the problems associated with the identification of plagiarism and fraud, and issues around authorship. There are many examples over the years of the resistance by scientists to scientific discoveries, including those of major figures such as Lister and Maxwell [21], and the House of Commons investigation into peer review in scientific publications identified a “perceived bias towards conservative judgements” or a lack of risk-taking ([22], p. 18). The ultimate “post-publication punishment: retraction”, although still quite rare, is increasing [23]. Van Noorden [23] points out that in the 2000s, there were only 30 retraction notices per annum, whereas in 2011 the number of retractions had grown to 400 “even though the total number of papers published has risen by only 44% over the past decade”. Perhaps even more alarmingly, retracted papers continue to be cited in the literature with the authors seemingly unaware of the retraction. Fortunately, new approaches are being developed such as CrossMark (<http://www.crossref.org/crossmark/>), which will draw the reader’s attention when downloading a pdf of an article that it has been subsequently retracted or corrected.

Some fraudulent research can have devastating consequences. In 1998, a doctor, Andrew Wakefield, published what subsequently turned out to be a fraudulent paper in *The Lancet* which provided evidence of a link between the combined MMR (measles, mumps and rubella) vaccine and autism ([24], retracted). The findings were widely reported in the media with the result that there was a rapid decline in vaccination uptake which was followed by significantly increased incidence of measles and mumps, resulting in deaths and severe and permanent injuries [25]. The paper was partially retracted in 2003 and fully retracted in 2010. The negative effects are still being felt today.

Authorship has been described as “the main currency in the world of science ... Authorships enable scientists to accumulate citations ... are key to getting grants and winning promotions” [26]. In research generally, but in the biomedical sciences in particular, “authorship provides recognition, but also establishes accountability and responsibility” [27]. Yet, in spite of the central role authorship plays in bibliometrics, it is in fact a poor indicator of the contribution

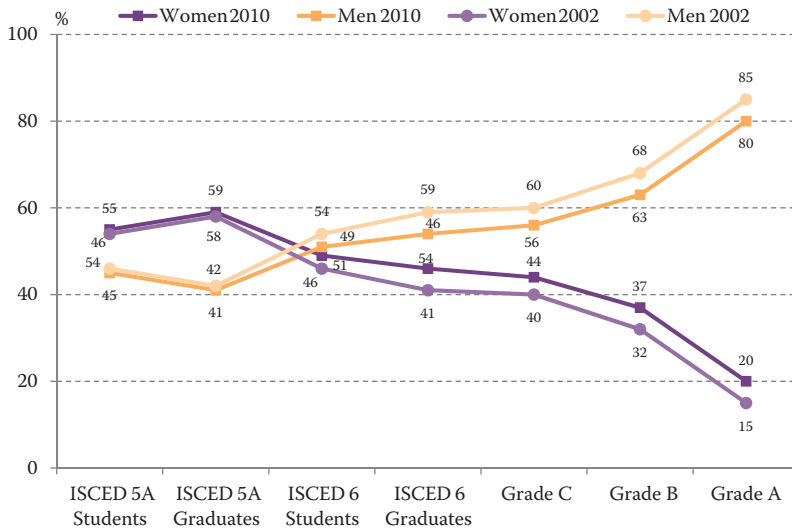
the author has made to the research. Guest authorship, where someone's name is added to a paper even though they have contributed nothing to the intellectual content, research or writing, is not unusual. Equally, ghost authorship, where someone's name is deliberately omitted from the list of authors even though they have made a substantial contribution, has been identified [28]. Both practices are unacceptable. One of the most egregious and widely publicized examples concerned the clinical trial results for the drug rofecoxib (marketed as Vioxx). Rofecoxib is a non-steroidal anti-inflammatory drug used to treat acute and chronic pain. It was very widely prescribed (an estimated 80 million users worldwide), but Merck voluntarily withdrew it from the market in 2004 after disclosures that the company had withheld information for over 5 years about increased risks of heart attack and stroke associated with long-term use of the drug [29]. It was one of the most widely used drugs ever to be withdrawn from the market. The original papers were ghost-written by sponsor employees, but first authorship was attributed to academics who did not always disclose that they had received financial support from Merck [27].

## Gender and research quality

In healthcare, ensuring equality of access to treatment is an important dimension of quality; this raises the equivalent question in research, namely whether all researchers have an equal opportunity to contribute to high-quality research and 'maximize their bibliometric score'. Not all disciplines, geographical regions and languages are covered in the same way [30,31]. Consideration has also been given as to whether or not bibliometrics are gender-neutral [32,33]. Certainly by virtue of the fact that the bibliometrics favour the sciences and women are more likely to work in the humanities, they are at a disadvantage relative to their male colleagues in science. However, this bias is due to differences between disciplines and not gender. Bibliometrics rely for their validity on peer review and "the issue of gender and peer review is highly contested" leading to often contradictory conclusions [30]. The first evidence that the peer-review system might not be gender-neutral was a study carried out by two Swedish researchers, Wenneras and Wold [34]. They examined the award of postdoctoral fellowships by the Swedish Research Council and showed that women had to publish 2.6 times as much as men to receive the same score. Although there has been subsequent criticism of their methodology [35], it did lead to most research funders tightening up their review procedures [32]. Even if we accept that the peer-review system is largely gender-neutral, there is irrefutable evidence of unequal participation of men and women in academia which, in turn, damages research quality. "Universities and research institutes are regarded as liberal, meritocratic institutions united in a commitment to academic excellence" [32]. Yet, the "political arithmetic of gender in the academy" tells a different story.

The latest figures published by the European Commission and reproduced in Figure 1 show that although women constitute the majority at undergraduate level (55%), their numbers decrease rapidly, with women comprising just 11% of full professors. Even in fields which have always had larger female partici-

Figure 1



### Proportion of men and women in a typical academic career, students and academic staff, EU-27, 2002–2010

*ISCE, International Standard Classification of Education. Figure reproduced from [36].*

pation, the numbers show that the glass ceiling in academia is still in place, with 28.4% of full professors in the humanities and 19.4% in the social sciences [36]. The numbers in the traditionally male disciplines of engineering and technology are much lower at just 7.9%. Importantly, the report concludes that “there is no evidence of spontaneous reduction of gender inequality over time”. A study by Swedish researchers of the differences in career trajectories of men and women with particular reference to achieving full professorships confirms this, concluding: “Measured as chances of advancing to the position of full professors, the career prospects for female university researchers are as bleak today as they were twenty years ago, despite policy programs aimed at promoting female professors” [37]. If one assumes that the ability to conduct good-quality research is equally distributed between the genders, this ‘leaky pipeline’ where women are lost disproportionately along the career ladder has major implications for research quality. A study by Valian [38] argued that “men and women are socialized into operating with implicit assumptions about gender differences” which result in men being over-rated and women under-rated, by both women and men. More generally, Merton concluded that the contribution of established researchers are consistently given more credit than is due, referred to as the ‘Matthew effect’ (to those who have more shall be given) [39]. From a gender perspective, this is not so much discrimination against women researchers, but rather the effect of a male bonus.

Research is fundamentally concerned with the quest for new knowledge and understanding or, as the English chemist William Henry (1775–1836) put

it, research is “a blind date with knowledge”. Research is thus a highly creative endeavour, and “since modern science is primarily carried out in groups, success depends not just on the creative individuals but also on creative groups” [33]. There is extensive evidence from the business world that gender-balanced teams are more innovative and creative [40] and that organizations with women in leadership positions perform better than those run just by men [41]. “Gender balance yields creativity. Focusing on scientific quality ... entails focusing on gender equity” [33].

There is another aspect of gender and research which has received increasing attention in recent years, namely the role of gender in the research itself, the type of research questions being asked and how the research is conducted [42]. Research that is gender blind can also simply be bad research [33]. Men and women are equally represented among society and therefore one would expect, for example, that trials of drugs would consider their effects on both men and women. Heart disease manifests differently in men and women yet the majority of studies have been performed on men with the result that women are often mis- or under-diagnosed ([42], p. 22). Equally, research on osteoporosis-related hip fractures was based on women with the result that men (who account for one-third of the total) are rarely evaluated or appropriately treated [43]. Climate-change research is generally scientifically and technically driven and therefore largely a male-dominated area. Yet, according to the Food and Agricultural Organization of the United Nations, women account for 60–80% of rural farmers and produce nearly 90% of the food on the continent [44]. They are therefore the ones most immediately affected by climate change and need to be actively involved in the development of appropriate solutions and approaches in setting the research agenda. At a recent conference on Hunger, Food and Climate Justice, Etrida Luhanga, a rural farmer from Malawi, pointed out that often the people who were sitting around the table deciding how best to adapt or mitigate the effects of climate change know nothing about farming and are only there “because they can read and write”. She argued strongly for the inclusion of rural farmers and women in particular who are experiencing the effects of climate change on a daily basis in the development of effective solutions [45].

## Conclusions

This paper has examined traditional bibliometrics as a method of measuring research quality through the lens of the use of KPIs to measure healthcare quality. It is widely recognized in both fields that the use of such metrics can be associated with gaming, perverse incentives and distortions of practice, thus undermining their validity. There are many ways in which the current set of bibliometrics could be improved which would help to reduce the inbuilt biases and distortions and the potential for gaming. Although it is possible to eliminate self-citation, citation cartels will inevitably be more difficult to identify since those working in a given field of research will inevitably cite the work of others in the same field. Further investigation would be required in order to be sure that the citation was actually of direct relevance to the particular paper. Clear and verifiable rules around



authorship would be highly desirable. But as Biagioli [46] concludes: “rather than pursue the chimera of the one conceptually ‘right’ definition, one may take a more pragmatic position by acknowledging that authorships (scientific or not) has always been a matter of compromises and negotiations, and that no new conditions have emerged to change that”.

Web 2.0 technologies including social networks have generated a myriad of methods for tracking and commenting on research in what Van Noorden [47] described as “a Cambrian explosion of metrics”. Wouters and Costas [2] identified and analysed 16 different tools and concluded that there was a need for a concerted research effort into understanding “the dynamics, properties, and potential use of new web based metrics” and how they relate to established indicators in the hope that this “may contribute to the development of more useful tools for the scientific and scholarly community” ([2], p. 45). In particular, they might provide evidence of impact, of “how new research findings are being read, cited, used and transformed in practical results and applications” ([2], p. 5). Bollen et al. [48] conducted a detailed study of new impact measures and concluded that “scientific impact is a multi-dimensional construct that cannot be adequately measured by one single factor”.

In healthcare, there is widespread recognition that indicators are just that, simply a proxy indication of quality, and that in order to truly understand whether the care being provided is safe and of good quality, it is necessary to consider many other, generally qualitative, issues. The risk of data-driven, as opposed to evidence-driven, indicators is widely recognized in healthcare [49], and similar problems have been identified with bibliometrics. Just because it is possible to measure an increasingly large number of aspects of research, does not necessarily mean that the results provide a better or more accurate measure of research quality ([47], p. 1). It seems that, to some extent, it is traditional bibliometrics which define what constitutes research quality rather than providing objective measures of research quality. Furthermore, they do not fully measure up to the requirements of validity, reliability, sensitivity, specificity, and underpinned by evidence, which should characterize a good indicator [10]. There needs to be the same recognition, particularly when assessing the performance of the individual researcher, that bibliometrics measure, however imperfectly, only some of the dimensions of research quality and that there is no substitute for a detailed study of the researcher’s output.

But the biggest concern in both research and healthcare is the potential to distort practice or ‘what gets measured gets done’. And there is plenty of evidence of this from the healthcare domain. The recent public enquiry into events at Mid-Staffordshire NHS Foundation Trust concluded that the failures which resulted in the deaths of 400–1200 patients over a 4-year period from 2005 were, in part, a consequence of “...allowing a focus on reaching national access targets, achieving financial balance and seeking foundation trust status ... at the cost of delivering acceptable standards of care” [50]. In research, there are many ways in which bibliometrics can distort practice, from discouraging risk-taking to focusing on research that is likely to be published rather than on research that will provide evidence to influence policy for the benefit of society. This is particularly evident in the context of international development research which seeks to address major

global challenges [31]. The challenge is that in attempting to measure the quality of healthcare and of research, the ultimate and definitive assessment can only be made in the long term with the benefit of hindsight. Medical treatment is given on the basis of the best available knowledge at that time, which may subsequently prove to be wrong, but this does not necessarily reflect negatively on the quality of the original research. In research, Tim Berners-Lee posted links to the computer code for the World Wide Web which in 1991 he had developed on the alt.hypertext discussion group so others could download it and play with it. No one at that time predicted the dramatic impact this invention was to have on all aspects of society across the world a decade later. It is only in the long term that we can really understand whether or not a health intervention has been beneficial or whether a particular piece of research is truly groundbreaking.

In addition to the problem of hindsight, there are other aspects of research quality which are not currently measured, either because they have not been considered as a dimension of research quality or because they are inherently difficult to measure. The OA (open access) movement is one such example. OA is concerned with making scholarly peer-reviewed publications freely available on the Internet [51,52]. The increase in the adoption of OA publication mandates represents an important development in making research results more widely available while at the same time benefiting the researchers through increased citation rates. Making a publication openly available should be reflected in the assessment of research quality. But arguably it is the open-data movement which may have greater implications for the assessment of research quality. Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions [53,54]. Providing OA to research results and data potentially speeds up the process of knowledge discovery, and therefore there is an argument to be made for including it as part of a suite of research quality indicators. The benefits of OA to research data was clearly demonstrated by the now infamous paper by Reinhart and Rogoff [55] which had a major impact on global economic policy and which was subsequently shown by Herndon, Ash and Pollin [56] to have been based on an incorrect Excel spreadsheet. The errors in the spreadsheet were such as to invalidate the conclusions of the paper, but not in time to prevent the results of the original paper contributing to widespread adoption of programmes of austerity in many countries.

Finally, existing approaches to the measurement of research quality are largely gender blind; yet, there is ample evidence to show that gender plays an important role in research quality, in terms of setting the research agenda, but also in ensuring that the 'system' supports the best researchers to do the best research. Good male researchers have nothing to fear from the inclusion of gender as a dimension of research quality!

So in summary, conventional bibliometrics give, at best, only a partial indication of research quality. As Wouters and Costas [2] have recommended, there is a need for much more research into the validity and impact of them and of the new web-based metrics. But, above all, there is a need to agree, first, on what exactly constitutes high-quality research, and secondly, whether and how it might be effectively measured.

## References

1. Engwall, L. and Scott, P. (2013) Setting the scene. In *Wenner-Gren International Series Volume 86: Trust in Universities* (Engwall, L. and Scott, P., eds), pp. 1–13, Portland Press Ltd, London
2. Wouters, P. and Costas, R. (2012) Users, narcissism and control – tracking the impact of scholarly publications in the 21st Century. SURF Foundation, Utrecht, The Netherlands. <http://research-acumen.eu/wp-content/uploads/Users-narcissism-and-control.pdf> (Accessed 21 February 2013)
3. Molinié, A. and Bodenhausen, G. (2010) Bibliometrics as weapons of mass citation. *Chimia* **64**, 78–89
4. Ernst, R. (2010) The follies of citation indices and academic ranking lists: a brief commentary to ‘bibliometrics as weapons of mass citation’. *Chimia* **64**, 90
5. Power, M. (1996) *The Audit Explosion*. DEMOS, London. <http://www.demos.co.uk/files/theauditexplosion.pdf> (Accessed 21 February 2013)
6. O’Neill, O. (2002) A question of trust. BBC Radio 4 Reith Lectures 2002. <http://www.bbc.co.uk/radio4/reith2002/> (Accessed 21 February 2013)
7. O’Neill, O. (2012) Integrity and quality in universities: accountability, excellence and success. *British Academy Review* **20**, 41–43
8. Merry, M.D. and Crago, M.G. (2001) The past, present and future of healthcare quality. *Physician Executive* **5**, 30–35
9. Institute of Medicine (2001) *Crossing The Quality Chasm: A New Health System For The 21st Century*. National Academy Press, Washington DC
10. Health Information and Quality Authority (2013) *Guidance on Developing Key Performance Indicators (KPIs) and Minimum Data Sets*. <http://www.hiqa.ie/publications/guidance-developing-key-performance-indicators-kpis-and-minimum-data-sets-monitor-health> (Accessed 3 May 2013)
11. Bowen, S. and Kreindler, S.A. (2008) Indicator madness: a cautionary reflection on the use of indicators in healthcare. *Healthcare Policy* **4**, 41–48
12. Campbell, D. (2008) Scandal of patients left for hours outside A&E. In *The Observer*. Sunday 17 February 2008, <http://www.guardian.co.uk/society/2008/feb/17/health.nhs1> (Accessed 4 May 2013)
13. Pitches, D.W., Mohammed, M.A. and Lilford, R.J. (2007) What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Services Research* **7**, 91
14. Thomas, J.W. and Hofer, T.P. (1998) Research evidence in the validity of risk-adjusted mortality rate as a measure of hospital quality of care. *Medical Care Research and Review* **55**, 371–404
15. Taylor, P. (2013) Rigging the death rate. *London Review of Books* **35**, 12–15
16. Scott, I., Brand, C., Phelps, G., Barker, A. and Cameron, P. (2011) Using hospital standardised mortality ratios to assess quality of care: proceed with extreme caution. *Medical Journal of Australia* **194**, 645–648
17. Øvretveit, J. (2009) *Does Improving Quality Save Money?* The Health Foundation, London. <http://www.health.org.uk/publications/does-improving-quality-save-money/> (Accessed 12 May 2013)
18. Méndez, E. (2012) *Evaluating Research Excellence: Main Debates*. International Development Research Centre, Ottawa. <http://www.idrc.ca/EN/Documents/Brief-Final-English.pdf> (Accessed 4 May 2013)
19. LSE Public Policy Group (2011) *Maximizing the Impacts of Your Research: A Handbook for Social Scientists*. London School of Economics, London. [http://eprints.lse.ac.uk/35758/1/Handbook\\_PDF\\_for\\_the\\_LSE\\_impact\\_blog\\_April\\_2011.pdf](http://eprints.lse.ac.uk/35758/1/Handbook_PDF_for_the_LSE_impact_blog_April_2011.pdf) (Accessed 4 May 2013)
20. Shoemake, K. (1985) Animating rotations with quaternion curves. *Computer Graphics* **19**, 245–254
21. Barber, B. (1961) Resistance by scientists to scientific discovery. *Science* **134**, 596–602
22. House of Commons Science and Technology Committee (2011) Peer review in scientific publications. <http://www.publications.parliament.uk/pa/cm201012/cmselect/cmsctech/856/856.pdf> (Accessed 5 May 2013)
23. Van Noorden, R. (2011) The trouble with retractions. *Nature* **478**, 26–28
24. Wakefield, A., Murch, S., Anthony, A. et al. (1998) Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* **351**, 637–641 (Retrieved 5 September 2007); retraction **375**, 445 (2010)

25. Pepys, M.B. (2007) Science and serendipity. *Clinical Medicine* 7, 562–578
26. Frische, S. (2012) It is time for full disclosure of author contributions. *Nature* 489, 475
27. Ross, J.S., Hill, K.P., Egilman, D.S. and Krumholz, H.M. (2008) Guest authorship and ghostwriting in publications related to rofecoxib. *Journal of the American Medical Association* 299, 1800–1812
28. Stemwedel, J.D. (2011) Scientific authorship: guests, courtesy, contributions, and harms. In *Scientific American*. 4 November 2011, <http://blogs.scientificamerican.com/doing-good-science/2011/11/04/scientific-authorship-guests-courtesy-contributions-and-harms/> (Accessed 5 May 2013)
29. Knox, R. (2004) Merck pulls arthritis drug Vioxx from market. <http://www.npr.org/templates/story/story.php?storyId=4054991>
30. Coryn, C.L. (2005) The use and abuse of citations as indicators of research quality. *Journal of MultiDisciplinary Evaluation* 4, 115–121
31. Tijssen, R.J., Mouton, J., van Leeuwen, T.N. and Boshoff, N. (2006) How relevant are local scholarly journals in global science? A case study of South Africa. *Research Evaluation* 15, 163–174
32. Rees, T. (2011) The gendered construction of scientific excellence. *Interdisciplinary Science Reviews* 36, 133–145
33. Rice, C. (2011) Scientific (E)quality. *Interdisciplinary Science Reviews* 36, 114–124
34. Wenneras, C. and Wold, A. (1997) Nepotism and sexism in peer review. *Nature* 347, 341–343
35. Mutz, R., Bornmann, L. and Daniel, H.-D. (2012) Does gender matter in grant peer review? An empirical investigation using the example of the Austrian Science Fund. *Zeitschrift für Psychologie* 220, 121–129
36. European Commission (2012) She figures 2012: gender in research and innovation. Directorate-General for Research and Innovation. [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/she-figures-2012\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/she-figures-2012_en.pdf) (Accessed 4 May 2013)
37. Danell, R. and Hjerm, M. (2012) Career Prospects for Female University Researchers: Event History Analysis of Career Trajectories at Swedish Universities. Proceedings of 17th International Conference on Science and Technology Indicators (Archambault, E., Gingras, Y. and Larivière, V., eds), 228–235, Science-Metrix and OST, Montréal
38. Valian, V. (1999) *Why So Slow? The Advancement of Women*. MIT Press, Cambridge, MA
39. Merton, R.K. (1968) The Matthew effect in science. *Science* 159, 56–63
40. London Business School (2007) Innovative potential: men and women in teams. [http://www.london.edu/assets/documents/facultyandresearch/Innovative\\_Potential\\_NOV\\_2007.pdf](http://www.london.edu/assets/documents/facultyandresearch/Innovative_Potential_NOV_2007.pdf) (Accessed 5 May 2013)
41. McKinsey & Company (2010) Women Matter 3: women leaders, a competitive edge in and after the crisis. [http://www.mckinsey.de/downloads/publikation/women\\_matter/women\\_matter\\_3\\_brochure.pdf](http://www.mckinsey.de/downloads/publikation/women_matter/women_matter_3_brochure.pdf) (Accessed 12 May 2013)
42. European Commission (2012) Structural change in research institutions: enhancing excellence, gender equality and efficiency in research and innovation. Directorate-General for Research and Innovation. [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/structural-changes-final-report\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/structural-changes-final-report_en.pdf) (Accessed 5 May 2013)
43. Schiebinger, L., Klinge, I., Sánchez de Madariaga, I., Schraudner, M. and Stefanick, M. (eds.) (2011–2013) Gendered Innovations in Science, Health & Medicine, Engineering and Environment. Osteoporosis research in men: rethinking standards and reference models. <http://genderedinnovations.stanford.edu/case-studies/osteoporosis.html>
44. Food and Agricultural Organisation (2013) Women, agriculture and rural development: a synthesis report of the Africa region. <http://www.fao.org/docrep/X0250E/X0250E00.htm> (Accessed 5 May 2013)
45. Luhanga, E. (2013) Thanks from Grassroots. Conference on Hunger, Nutrition and Climate Justice, Dublin Castle, 15–16 April 2013. [http://www.youtube.com/watch?v=KaWIsodrFQ&feature=youtu.be\\_gdata\\_player](http://www.youtube.com/watch?v=KaWIsodrFQ&feature=youtu.be_gdata_player) (Accessed 4 May 2013)
46. Biagioli, M. (1998) The instability of authorship: credit and responsibility in contemporary biomedicine. *The FASEB Journal* 12, 3–16
47. Van Noorden, R. (2010) A profusion of measures. *Nature* 465, 864–866
48. Bollen, J., Van de Sompel, H., Hagberg, A. and Chute, R. (2009) A principal component analysis of 39 scientific impact measures. *PLoS ONE* 4, e6022
49. Bowen, S., Erickson, T., Martens, P. and Crockett, S. (2009) More than “using research”: the real challenges in promoting evidence-informed decision-making. *Healthcare Policy* 4, 87–102
50. Francis, R. (2013) Report of the Mid-Staffordshire NHS Foundation Trust Public Inquiry. <http://www.midstaffspublicinquiry.com/report> (Accessed 5 May 2013)

51. European Research Advisory Board (2006) Scientific Publication: Policy on Open Access. Final report. [http://ec.europa.eu/research/eurab/pdf/eurab\\_scipub\\_report\\_recomm\\_dec06\\_en.pdf](http://ec.europa.eu/research/eurab/pdf/eurab_scipub_report_recomm_dec06_en.pdf) (Accessed 6 May 2013)
52. Houghton, J. and Sheehan, P. (2006) The economic impact of enhanced access to research findings. CSES Working Paper No. 23, July 2006. Centre for Strategic Economic Studies, Victoria University, Melbourne. <http://www.cfses.com/documents/wp23.pdf> (Accessed 11 August 2013)
53. Auer, S.R., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. and Ives Z. (2007) DBpedia: A Nucleus for a Web of Open Data. The Semantic Web. Proceedings of 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, 11–15 November 2007 (Aberer, K., Choi, K.-S., Noy, N. et al. eds.), 722–735, Springer, Lecture Notes in Computer Science 4825
54. Open Data Institute. What is open data? <http://theodi.org/guides/what-open-data>
55. Reinhart, C.M. and Rogoff, K.S. (2010) Growth in a time of debt. *American Economic Review* 100, 573–578
56. Herndon, T., Ash, M., and Pollin, R. (2013) Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. Political Economy Research Institute, University of Massachusetts Amherst, Working Paper Series No. 322. [http://www.peri.umass.edu/fileadmin/pdf/working\\_papers/working\\_papers\\_301-350/WP322.pdf](http://www.peri.umass.edu/fileadmin/pdf/working_papers/working_papers_301-350/WP322.pdf) (Accessed 11 August 2013)