

COMMENTARY

Open Access

Judging research quality to support evidence-informed environmental policy

Simin Davoudi^{1*}, Gemma Harper², Judith Petts³ and Sarah Whatmore⁴

Keywords: Research quality, Social sciences, Systematic reviews, Evidence-informed

In August 2005, PLoS Medicine published an essay by John Ioannidis called: ‘Why most published research findings are false’ [1]. Since then, the paper has been viewed over a million times possibly because of its provocative title, but probably because of growing concerns about the reliability of scientific publications and diminishing confidence in the peer review process to deliver effective quality control. Most recently such concerns have been supported by a number of high profile cases of publication retraction, for example the withdrawal of some statements from articles published in the British Medical Journal regarding the adverse side effects of Statins, a cholesterol-reducing drug [2]. A series of articles in *the Lancet* has suggested that some \$200 billion (estimated to be 80% of the world’s spend on medical research) was wasted on ‘studies that were flawed in their design, redundant, never published or poorly reported’ [3]. Moreover, ‘when a prominent medical journal ran research past other experts in the field, it found that most of the reviewers failed to spot mistakes it had deliberately inserted into papers, even after being told they were being tested’ [4]. In the field of environmental studies too, concerns have been raised about ‘the limited effectiveness of peer-review as a quality-control mechanism’ [5].

The actual and perceived unreliability of scientific reports and papers is particularly problematic for governments’ scientific advisors who operate in the ‘messy’ world of policy making. Here even credible evidence is rarely the only influential factor [6] and policy (and evidence) contention, scientific uncertainty and even ignorance pose significant challenges. The Chief Scientific Advisor for the UK government’s Department for Environment, Food and Rural Affairs has called ‘for an auditing process

to help policy makers to navigate research bias’ [7] and suggested the need for establishing third-party certified auditors and international auditing standards that grade scientific studies or even journals. A case has also been made for adopting healthcare best practice quality assessment tools for environmental science [8]. Others have suggested the use of ‘formal consensus methods’ such as Delphi techniques to achieve better quality control [9]. Dr Ioannidis (mentioned above) has already institutionalised these ideas by launching in early 2014 a Meta-Research Innovation Centre at Stanford University. METRICS’ mission is ‘to undertake rigorous evaluation of research practices and find ways to optimize the reproducibility and efficiency of scientific investigations’ [10]. These efforts point to a growing call for the tightening of peer review, or even dispensing with it, in favour of post-publication evaluation in the form of appended comments.

We share the concerns raised by other commentators about the reliability of research and the evidence it produces, and support the efforts to promote quality assurance. However, our motivation for writing this article is to raise concerns about the perceived validity and value of social science evidence (compared with scientific outputs from the physical, natural, engineering, and medical sciences) in interdisciplinary research. Interdisciplinarity is particularly relevant in the field of environmental policy and management which often grapples with multiple questions that demand diverse research methods from both social and natural sciences. Defined broadly, the social sciences study societal processes and peoples’ lived experiences as these shape, and are shaped by, the world around them. Understanding what people, individually and in various forms of association with others, think and do poses unique research challenges. Studying society involves not only the objective system under scrutiny, but also the subjective system of scrutiny itself (known as double hermeneutic). In consequence, social

* Correspondence: simin.davoudi@ncl.ac.uk

¹School of Architecture, Planning and Landscape, Newcastle University, Claremont Tower, Newcastle upon Tyne NE1 7RU, UK

Full list of author information is available at the end of the article

science disciplines have developed a variety of quantitative and qualitative research techniques adapted to these challenges. They draw on data generated by a variety of methods including statistical analyses, survey questionnaires, in-depth interviews, participant observation, and group discussions. Some of these methods and the criteria appropriate for evaluating the reliability of the evidence that they generate may be unfamiliar to other scientific fields.

We are concerned that a desire to set universally applicable 'kite marks' and 'gold standards' may risk undermining: an appreciation of the complementarity of different methods (within and between the quantitative and qualitative); the importance of adopting an inclusive definition of evidence; the diversity of research designs and methods; and, the significance of 'fitness for purpose' in research design, conduct and reporting. There appears to be a tendency to consider qualitative methods as somehow inferior to experimental or quantitative methods *a priori* [11]. From our experience, sometimes decision makers question evidence that is based on analysis of narrative, discussion and commentary and use statistical representativeness and reproducibility as the primary criteria for assessing the quality of research. These tendencies are not random instances; they are indicative of power relations within and beyond science and embedded in the fabric of knowledge itself. Arguing against these tendencies is not new. It has a long pedigree in the field of environmental research and policy-making [12] whose interdisciplinary nature demands that diverse framings of the problem and multiple methods of investigation typically come together and challenge each other in producing new ways of knowing.

In this context, evidence must be understood broadly to encompass the insights from the natural, physical *and* social sciences and provide space for 'a measured array of contrasting specialist views' [13]. While our focus here is on research, we believe that tacit and experiential knowledge by which 'much of the world's work of problem solving is accomplished' [14] should also be included in the definition of evidence. Similarly, quality should be defined inclusively and the mechanisms and criteria used to judge it should reflect the diversity of research methods and paradigms. This means that the criteria used to assess the quality of, for example, randomised controlled trials (RCT) may not be suitable for assessing qualitative methods.

The key point is that applying the same criteria universally to all types of research is imprudent. The approach to quality control should start by asking which method or mixture of methods are most appropriate for answering the research questions and the research project's intended uses. The validity and credibility of the method depends fundamentally on its fitness for purpose. For

example, while statistics can tell us the voting patterns of a given social group in a general election, they do not *explain* why the group and importantly the individuals within it, voted as they did. As with all sciences, what causes something to happen in a particular case may not necessarily be explained by the number of times we observe it happening. Finding out 'how' and 'why' people vote as they do necessitates an understanding of what voting for a particular outcome means to the individual voter. Such understanding requires a mixture of complementary quantitative and qualitative (or Q2) methods. Thus, appropriateness should be the first test of quality control. Once the appropriateness of the method is established, criteria relevant to that method can be drawn upon to assess its quality and distinguish between, for example, a high and low quality RCT, or a high and low quality focus group. This means that before asking whether this research is valid, we should be asking what 'this research is valid for' [15]. We would be amongst the first to acknowledge that there is low quality social science as well as low quality natural science but, no method can be judged better or worse than another in isolation from the research questions they aim to address. Accordingly, it is essential that we avoid the tendency to assess the quality of research methods by a universal set of criteria or worse even, to assess qualitative methods by the same criteria developed for and used in quantitative methods (such as the statistical validity of the participants sample size for a focus group).

There is a growing body of literature on criteria and checklists for assessing quality in social science research [16,17]. These have been applied to single research projects and syntheses of qualitative research, as well as systematic reviews similar to those conducted by Cochrane and Campbell Collaborations and the Collaboration for Environmental Evidence. Reports have suggested that there are over one hundred sets of proposals on quality in qualitative research [18]. However, there appears to be few attempts to develop method-specific approaches. Furthermore, in selecting papers for inclusion in the systematic reviews 'consensus about which aspects of design, execution, analysis and description are most crucial is yet to be reached' [19]. Furthermore, there is even a lack of consensus about whether such reviews are appropriate for studies using qualitative methods whose assessment involves an iterative process and does not follow the often linear approach used in experimental and quantitative research [20]. One area on which both social and natural scientists agree is the acknowledgment that assessing the quality of evidence is a subjective process and involves judgment. In the context of systematic reviews, structured approaches (such as checklists and tools) have long been proposed as a means of assessing the quality of research reports and reducing subjectivity. However, a comparison

of structured approaches and ‘unprompted judgement’ has shown that although the former ‘may sensitise reviewers to aspects of research practice’, they do ‘not appear more likely to produce a higher level of agreement between or within reviewers’ [21]. It is also important to note that there is a wide range of methods for synthesising qualitative research. Barnett-Page and Thomas [22], for example, have identified ten different methods spanning across the “realist – idealist” epistemological spectrum and each with their own criteria for quality assessment. It is therefore important that in undertaking systematic reviews of qualitative research, attention is paid to the suitability of the criteria for not only quality assessment but also the synthesis method itself.

To summarise, the main messages of this commentary are as follows:

- Evidence for environmental policy should be defined broadly and inclusively to incorporate the insights from all sciences.
- There is a diversity of social scientific research methods, each with its own specific contributions to environmental decision making.
- Mechanisms and criteria for judging research quality should take account of such diversity and be fit for purpose.
- To make the best of social sciences their contributions should be fully integrated at the beginning into environmental policy development and interdisciplinary research.

Abbreviations

RCT: Randomised controlled trials.

Authors' contributions

SD drafted and revised the manuscript. JP, GH and SW commented on and improved the draft manuscript. All authors read and approved the final manuscript.

Authors' information

SD is Professor of Environmental Policy and Planning at Newcastle University. GH is Chief Social Scientist at the Department for Environment, Food and Rural Affairs (Defra). JP is Professor of Environmental Risk Management at Southampton University and member of DEFRA Science Advisory Council. SW is Professor of Environment and Public Policy at Oxford University and member of DEFRA Science Advisory Council. All authors are members of the Defra and DECC (Department for Energy and Climate Change) Social Science Expert Panel.

Acknowledgement

We would like to thank members of the Department for Environment, Food and Rural Affairs (Defra) and the Department for Energy and Climate Change (DECC) Social Science Expert Panel, Defra Social Research Group, Professor Andy Stirling, and two anonymous reviewers for their insightful comments. The views expressed here are solely our responsibility.

Author details

¹School of Architecture, Planning and Landscape, Newcastle University, Claremont Tower, Newcastle upon Tyne NE1 7RU, UK. ²Department for Environment, Food and Rural Affairs, 5C, Noble House, 17 Smith Square, London SW1P 3JR, UK. ³Vice-Chancellor's Office, University of Southampton, Highfield, Southampton SO17 1BJ, UK. ⁴Keble College, Oxford OX1 3PG, UK.

Received: 12 November 2014 Accepted: 11 March 2015

Published online: 31 May 2015

References

- Ioannidis J. Why most published research findings are false. *PLoS Medicine*. 2005;2(8):696–701.
- British Medical Journal: Editorials, Adverse effects of statins, 2014, 348:g3563 published 15 May, available at <http://www.bmj.com/content/348/bmj.g3306> accessed 11 August 2014
- The Economist: Combating bad science, *Metaphysicians*, 2014, 15 March, p.78 available at <http://www.economist.com/news/science-and-technology/21598944-sloppy-researchers-beware-new-institute-has-you-its-sights-metaphysicians> accessed 11 August 2014
- The Economist: How science goes wrong? 2013, 19 October, Leader section, available at <http://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong> accessed 11 August 2014
- Bilotta G, Milner A, Boyd I. Quality assessment tools for evidence from environmental science. *Environ Evid*. 2014;3(14):1–14. p. 1.
- Davoudi S. Evidence-based planning: rhetoric and reality. *DisP: The Planning Review*. 2006;165(2):14–25.
- Boyd I. A standard for policy-relevant science. *Nature*. 2013;501:159–60. p.159.
- Bilotta G, Milner A, Boyd I. Quality assessment tools for evidence from environmental science. *Environ Evid*. 2014;3(14):1–14.
- Sutherland WJ. Review by quality not quantity for better policy. *Nature*. 2013;503:167–68. p.167.
- Available at <http://med.stanford.edu/metrics/> accessed 7/8/2014
- Veltri GA, Lim J, Miller R. More than meets the eye: the contribution of qualitative research to evidence-based policy making. *Innovation: The European Journal of Social Sciences Research*. 2014;27(1):1–4.
- Burgess J, Goldsmith B, Harrison C. Pale shadows for policy: reflections on the Greenwich open space project. *Stud Qual Meth*. 1990;2:141–67.
- Stirling A. Keep it complex. *Nature*. 2010;468:1029–31. p. 1030.
- Lindblom CE, Cohen D. *Unusable knowledge: Social sciences and social problem solving*. New Haven CT: Yale University; 1979. p. 91.
- Garside R. Should we appraise the quality of qualitative research reports for systematic reviews, and if so, how? *Innovat Eur J Soc Sci Res*. 2014;27(1):67–9. p.76.
- HM Treasury. *Quality in qualitative evaluation: a framework for assessing research evidence (supplementary Magenta Book guidance)*. London: HM Treasury; 2012.
- National Centre for Social Research. *Quality in qualitative education: a framework for assessing research evidence*. London: National Centre for Social Research / UK Cabinet Office; 2003.
- NHSCRD (NHS Centre for Reviews and Dissemination): *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews NHS Centre for Reviews and Dissemination, 2001, York University, York YO10 5DD (CRD Report 4: second edition)*, p. 221. Available at: <http://www.york.ac.uk/inst/crd/report4.htm> accessed 7/8/2014
- Garside R. Should we appraise the quality of qualitative research reports for systematic reviews, and if so, how? *Innovat Eur J Soc Sci Res*. 2014;27(1):67–9. p.68.
- Freeman M, de Marrais K, Pressie J, Roulston K, St Pierre EA. Standards of evidence in qualitative research: An introduction to discourses. *Educational Researcher*. 2007;36:25–32.
- Dixon-Woods M, Sutton A, Shaw R, Miller T, Smith J, Young B, et al. Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. *Journal of Health Service Research Policy*. 2007;12(1):42–7. p. 46.
- Barnett-Page E, Thomas J. Methods for the synthesis of qualitative of research: a critical review. *BMC Med Res Meth*. 2009;9:59.