BMJ Open

Understanding how appraisal of doctors produces its effects: a realist review protocol

Nicola Brennan,[1] Marie Bryce,[1] Mark Pearson,[2] Geoff Wong,[3] Chris Cooper,[2] Julian Archer[1]

To cite: Brennan N, Bryce M, Pearson M, *et al*. Understanding how appraisal of doctors produces its effects: a realist review protocol. *BMJ Open* 2014;**4**: e005466. doi:10.1136/bmjopen-2014-005466

▸ Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2014-005466).

Received 11 April 2014
Revised 29 May 2014
Accepted 2 June 2014

[1]Collaboration for the Advancement of Medical Education Research and Assessment, CAMERA, Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth University, Plymouth, UK
[2]Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, University of Exeter, Exeter, UK
[3]Centre for Primary Care and Public Health, Queen Mary University of London, London, UK

Correspondence to
Dr Julian Archer;
julian.archer@plymouth.ac.uk

ABSTRACT

Introduction: UK doctors are now required to participate in revalidation to maintain their licence to practise. Appraisal is a fundamental component of revalidation. However, objective evidence of appraisal changing doctors' behaviour and directly resulting in improved patient care is limited. In particular, it is not clear *how* the process of appraisal is supposed to change doctors' behaviour and improve clinical performance. The aim of this research is to understand *how and why* appraisal of doctors is supposed to produce its effect.

Methods and analysis: Realist review is a theory-driven interpretive approach to evidence synthesis. It applies realist logic of inquiry to produce an explanatory analysis of an intervention that is, what works, for whom, in what circumstances, in what respects. Using a realist review approach, an initial programme theory of appraisal will be developed by consulting with key stakeholders in doctors' appraisal in expert panels (ethical approval is not required), and by searching the literature to identify relevant existing theories. The search strategy will have a number of phases including a combination of: (1) electronic database searching, for example, EMBASE, MEDLINE, the Cochrane Library, ASSIA, (2) 'cited by' articles search, (3) citation searching, (4) contacting authors and (5) grey literature searching. The search for evidence will be iteratively extended and refocused as the review progresses. Studies will be included based on their ability to provide data that enable testing of the programme theory. Data extraction will be conducted, for example, by note taking and annotation at different review stages as is consistent with the realist approach. The evidence will be synthesised using realist logic to interrogate the final programme theory of the impact of appraisal on doctors' performance. The synthesis results will be written up according to RAMESES guidelines and disseminated through peer-reviewed publication and presentations.

Trial registration number: The protocol is registered with PROSPERO 2014:CRD42014007092.

BACKGROUND

What do we know about appraisal of doctors?

Appraisal can be defined as the process by which an appraiser examines and evaluates an appraisee's work behaviour by comparing it with preset standards. The results of the comparison are then documented and used to provide feedback to the appraisee on their performance, to show where improvements are needed and why.[1] It is widely assumed that appraisal of doctors leads to an improvement in clinical performance.[2] However, it remains unclear whether its perceived benefits are actually realised. According to one academic, "it's one of those wonderfully simple, obvious questions the literature does not address."[3]

A systematic review conducted by Overeem *et al*[4] in 2007 investigated the effect of appraisal (and other assessments) on doctors' performance. They found that the majority of doctors undertaking appraisal were satisfied with the evaluation they received in their appraisal and reported performance improvements. A further scoping review of the literature aiming to update the review by Overeem *et al*[5] found that a number of other studies have subsequently been published on the effect of appraisal on performance. Again these studies were based on doctors' perceptions or beliefs that appraisal was having a positive effect on their performance rather than an objective measure that appraisal results in performance improvement.

While not a study of performance improvement, a paper by West in 2002 found an association between appraisal and improved patient care. Human resources (HR) directors from 61 acute hospitals in England completed questionnaires or interviews exploring HR practices and procedures.[6] The interviews asked about the extensiveness and sophistication of appraisal and training for employees and the percentage of staff working in teams. Data were also collected on patient mortality. The results demonstrated strong associations between HR practices and patient mortality in general. The

extent and sophistication of appraisal in the hospitals were significantly associated with measures of patient mortality. There was also a correlation with the sophistication of training for staff and the percentages of staff working in teams. Despite the fact that this study shows an association between appraisal and patient outcomes it does not prove causality that is appraisal results in improved patient care.

Why is appraisal important?

Medical revalidation was introduced in the UK in December 2012 by the General Medical Council (GMC).[7] This represented a major shift in the regulation of the profession from a system of self-regulation.[8] Revalidation requires all doctors' licences to practise to be reviewed every 5 years[9] by a responsible officer (RO) who is normally the most senior doctor locally. The RO makes their judgement based on an output summary from each annual appraisal and, where relevant, supplementary clinical governance data such as morbidity rates, serious complaints and other markers of performance.

Appraisal is therefore at the heart of revalidation, which, as Murie described it, is "perceived to be the 'cement', which binds the system for the revalidation of doctors."[10] During the appraisal meeting, doctors are required to discuss their practice with a trained appraiser, usually another doctor. They are asked to provide *supporting information*[11] to demonstrate that they are continuing to meet the principles and values set out in the GMC's core professional guidance: *Good Medical Practice*.[12]

The GMC currently lists six types of supporting information:

1. Continuing professional development (CPD)
2. Quality improvement activity
3. Significant events analysis
4. Feedback from colleagues
5. Feedback from patients
6. Complaints and compliments.[11]

Importantly, appraisees are expected to reflect on their supporting information, and discuss their intentions to develop or change their practice as a result.[11] In this way, the GMC has tried to establish appraisal as an active process potentially supporting performance change and not simply an audit trail.

Identifying the need for further research on appraisal of doctors

The implementation of revalidation in the UK with appraisal being a key component has resulted in an increased importance of the appraisal process. However, there is currently little objective evidence of appraisal changing doctors' behaviour and directly resulting in improved patient care.[13] This may be because the impact of appraisal on performance is quite difficult, and possibly unfeasible, to measure accurately. In order to measure appraisal outcomes, longitudinal studies would need to be conducted comparing specific outcomes before and after the introduction of appraisal. As appraisal systems have been implemented for many years, the opportunities to carry out such studies are limited.

Studies are also limited in explaining how and why the underlying mechanisms of appraisal are supposed to produce their intended effect. While we use the term 'appraisal' as if it is a 'universal' process, it is operationalised in different ways in different settings, has different objectives and means different things to different people that is, we are not all talking about the same 'thing'. Similarly, the fact that there are very disparate views on the goal of revalidation will have a direct impact on the appraisal process. A discourse analysis of the revalidation policy[8] found that revalidation is perceived by some as a way to identify 'bad apples', requiring a summative approach and minimum standards. Others take a professional stance and view revalidation as a process by which all doctors improve; requiring evolving standards and a developmental model. These two discourses are not simply divergent; indeed most documents and participants used them interchangeably, but they are in some regards at odds. This dichotomy could have an impact on the mechanisms at play in the appraisal process.

This brief overview of the literature on appraisal highlights a clear gap in the evidence base, which is the need to understand how and why appraisals of doctors are actually supposed to produce their effect. By identifying the causal mechanisms at work in the appraisal process it may be possible to design experimental research studies to effectively investigate the effect of appraisal on doctors' performance. A better understanding of 'how appraisal works' will also inform decision-making about how to tailor and implement appraisal processes at a local level.

METHODS

Research questions

The aim of the current research is to understand how and why appraisal of doctors produces its effect. In particular, what are the mechanisms by which appraisal is believed to result in its intended outcomes? Our research questions are as follows:

1. What are the mechanisms by which appraisal of doctors is believed to result in its intended outcomes?
2. What are the important contexts which determine whether the different mechanisms produce their intended outcome?
3. In what circumstances is appraisal likely to be effective?

Realist review

The research questions will be addressed using a realist review approach. Realist review is a theory-driven, interpretive approach to the synthesis of evidence. It seeks to interrogate the theories that underpin the intervention being studied, in this case appraisal, to produce an

explanatory analysis of it, that is, what works, for whom, in what circumstances, in what respects.[14] A realist synthesis takes a 'generative' approach to causation, that is, "to infer a causal outcome (O) between two events (X and Y), one needs to understand the underlying mechanism (M) that connects them and the context (C) in which the relationship occurs."[15]

Realist synthesis is typically used to understand complex interventions. Complex interventions "often have multiple components (which interact in non-linear ways) and outcomes (some intended and some not) and long pathways to the desired outcome(s)."[16] One of the central processes of a realist review is the development of a programme theory. "The term 'programme theory' refers to an abstracted description and/or diagram that lays out what a programme (or family of programmes or intervention) comprises and how it is expected to work."[16] A realist approach is particularly useful for the current research because appraisal is a complex intervention that is context sensitive. At present there is little or no understanding of how and why appraisal of doctors leads to particular outcomes and of the contexts under which such outcomes might occur. The realist review will not provide a summative judgement on whether appraisal is 'good' or 'bad' but instead will explain how, why, in what contexts, for whom and to what extent appraisal 'works'.

Study design

This review protocol follows Pawson's five practical stages in conducting realist reviews[17] and is summarised in figure 1.

Step 1: locate existing theories

As a first step we will identify existing theories in order to develop an initial overall programme theory of appraisal and thus explain how appraisal is supposed to result in its intended outcome. We are looking for theories that help us to explain specific aspects of appraisal (eg, how feedback on performance is meant to change behaviour) but also to explain how this section of the appraisal process fits in with the other sections. We will achieve this in two ways. First, by consulting with key stakeholders in appraisal, for example, doctors/appraisees, appraisers, academic experts on appraisal, the GMC, responsible officers, human resources personnel to better understand appraisal. This process will involve a series of expert panel meetings using facilitated discussions centred on evolving programme theory. Formal ethical approval will not be required but informed participation will be sought. Second, by searching the literature to identify existing theories on how and why appraisal is meant to work. These theories will form the basis of our initial programme theory. This initial programme theory will then be tested against data from studies included in the review.

As well as helping identify the existing theories the stakeholder group will also contribute to other stages of the review. They will act as a 'reality check' to see whether the theories in the literature about 'why appraisal improves doctors' performance' make sense from their experience of clinical practice. They will also sense check the emerging findings as the review progresses. The group will meet regularly throughout the study and will also communicate via email.

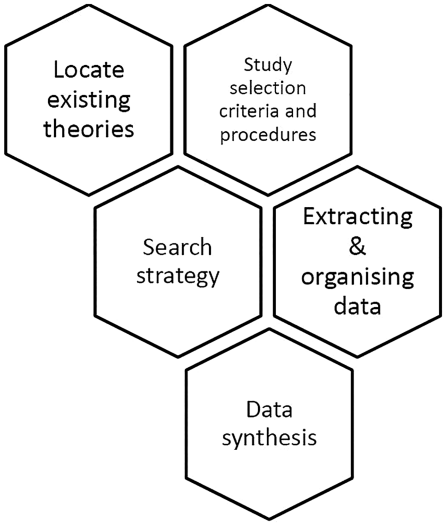

Figure 1 Study design using Pawson's five practical stages[17] (The diagram is tessellated to demonstrate that this is not a linear process. Realist reviews are iterative and the process often requires movement between stages.).

Step 2: search strategy

The search strategy will involve two phases. First, we will search for data that explain how appraisal is meant to work to produce its desired outcomes. The second phase will involve seeking additional relevant data to enable testing and refinement of our programme theory. We anticipate our search strategy to include a combination of the following search methods:

1. Electronic database searching (using keywords based on the theories identified): EMBASE, MEDLINE, ERIC, the Cochrane Library, PsycINFO, HMIC, Social Policy and Practice, CINAHL, British Nursing Index Conference Proceedings Citation Index, Web of Science, ASSIA and any other relevant databases identified by the Information Specialist or the team;
2. 'Cited by' articles search;
3. Citations contained in the reference lists of included papers;
4. Contacting authors;
5. Grey literature searching.

The search for evidence in a realist review is iterative and will be progressively extended and refocused (based on the identified sources) as the review evolves.

Step 3: study selection criteria and procedures

Documents will be selected based on relevance (ie, can provide data that inform programme theory development and refinement). These are likely to include

editorials, opinion pieces, commentaries, process evaluations, qualitative research, programme manuals and systematic reviews. A random sample of 10% of articles will be selected, assessed and discussed by two review authors using a preliminary set of inclusion/exclusion criteria. The remaining 90% will be completed by one reviewer. However, a number of these may require discussion between the two reviewers as they could be pivotal papers which need discussion to integrate into the review. In realist reviews, the study itself is rarely used as the unit of analysis; instead realist reviews consider small sections of the primary study to test a very specific hypothesis about the relationships between context, mechanism and outcomes.[18] We will thus select and review studies based on what new knowledge they bring to our thinking about the programme theory of the impact of appraisal on doctors' performance.

Step 4: extracting and organising data

The realist review method synthesises information by note-taking and annotation rather than standardised data extraction as used in a traditional systematic review. Documents are examined for theories on how an intervention is supposed to work which are then highlighted, noted and given an approximate label. The reviewer may make use of data extraction forms to assist the sifting, sorting and annotation of primary source materials but they do not take the form of a single, standard list of questions as used in a traditional systematic review. Thus in this review, data extraction will be carried out in different (and appropriate) ways at different review stages.

Quality assessment will use the concept of rigour—whether the methods used to generate the relevant data are credible and trustworthy. Rigour will be assessed using a hybrid appraisal tool based on previous critical appraisal work, which enables sources to be classified as conceptually rich (thick) or thin (weaker) in description.[19 20] This tool has been found to be practical and useful in theory-driven reviews as it allows the reviewer to focus on the stronger sources of programme theories without excluding weaker sources that may make an important contribution.[21]

Step 5: data synthesis

We will synthesise the data using a realistic logic of analysis to interrogate the final theory which will be to determine what it is about appraisal that works and for whom, in what circumstances, in what respects and why. Specifically, we will seek data from included sources to test and refine each section of our initial programme theory. For the outcome of each section of our initial programme theory we will seek data to help us to infer what the causal mechanism(s) might be and the context (s) when the mechanism might be triggered.

Synthesis of the data from diverse sources of evidence included in a realist review is conducted through a process of reasoning that is structured around the following activities:

A. Juxtaposition of sources of evidence—for example, where evidence about performance improvement in one article allows insights into evidence about outcomes in another article;
B. Reconciling of sources of evidence—where results differ in comparable circumstances, these will be examined further to find possible reasons for the different results;
C. Adjudication of sources of evidence—based on methodological strengths or weaknesses;
D. Consolidation of sources of evidence—where outcomes differ in particular contexts, an explanation will be constructed on how and why these outcomes occur differently;
E. Situating sources of evidence—when outcomes are different in particular contexts, a possible explanation will be developed as to why they differ.[22 23]

The final realist programme theory will be summarised through narrative synthesis, using text, summary tables, a logic model and where appropriate graphics to summarise individual papers/reports and draw insights across papers/reports. The results of the synthesis will be written up according to the 'Realist and Meta-Review Evidence Synthesis: Evolving Standards' (RAMESES) standard for reporting realist reviews.[16]

DISCUSSION

Importance of the research

The findings of this research will provide a crucial insight into how appraisal is supposed to produce its intended effects and ultimately change clinical practice. This information will be important for doctors, the employment sectors and policymakers to improve the appraisal process, as well as other stakeholders in medical appraisal and revalidation in the UK including appraisers, ROs, the GMC, the National Health Service and the Departments of Health. In addition to medicine, the results could be utilised by other healthcare professions, with the nursing[24] and pharmacy[25] professions in the UK in the process of designing their own revalidation processes of which appraisal may be a part.

The GMC is the first regulator in the world to implement a compulsory revalidation process[26] but other medical regulators are considering the implementation of revalidation style systems and look to the UK model for guidance. For example, the Australian medical regulator (Medical Board of Australia) is currently proposing the introduction of revalidation incorporating appraisal as a core element.[27] Further still the results will also be of benefit to the wider employment sector outside of healthcare with most of the top international listed companies using appraisal for their employees.[28]

Finally, the findings of this research will provide useful information for academics, researchers and policymakers. By identifying the causal mechanisms at work in the appraisal process it may be possible to design experimental research studies to effectively investigate the

impact of appraisal on doctors' performance. A better understanding of 'how appraisal works' will also inform decision-making about how to tailor and implement appraisal processes at a local level.

Contributors JA conceptualised the study. NB led the design and drafting of the review protocol which was critically reviewed by JA, MB, MP, CC and GW. CC scoped and designed the search strategy. Methodological advice was given by MP and GW. NB wrote the first draft of this paper. JA, MB, MP, CC and GW critically reviewed it and provided comments to improve the manuscript. All authors have read and approved the final manuscript.

Funding This article presents independent research funded by the UK Department of Health and the National Institute for Health Research (NIHR) under its Fellowship Scheme (NIHR-CDF-2011-04-004). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES

1. Smither JW, London M. *Performance management: putting research into action*. John Wiley & Sons, 2009.
2. Wilson F. Dilemmas of appraisal. *Eur Manag J* 2002;20:620–9.
3. Zemke R. Do performance appraisals change performance? *Training* 1991;28:34–9.
4. Overeem K, Faber MJ, Arah OA, *et al*. Doctor performance assessment in daily practise: does it help doctors or not? A systematic review. *Med Educ* 2007;41:1039–49.
5. Archer J, Regan De Bere S, Brennan N, *et al*. *Evaluating the strategic impact of revalidation: building an evaluation framework*. Plymouth: CAMERA, Plymouth University Peninsula Schools of Medicine and Dentistry, 2013.
6. West MA, Borrill C, Dawson J, *et al*. The link between the management of employees and patient mortality in acute hospitals. *Int J Hum Resource Manag* 2002;13:1299–310.
7. General Medical Council. *Starting revalidation*. General Medical Council, 2013.
8. Archer J, Regan De Bere S, Nunn S, *et al*. *Revalidation: in policy: Peninsula Medical School*. Plymouth, 2012.
9. General Medical Council. *Consultation document: revalidating doctors. Ensuring standards, securing the future*. London: General Medical Council, 2000.
10. Murie J. Knowing me, knowing you: personality and peer appraisal. *Br J Gen Pract* 2010;60:382.
11. General Medical Council. *Ready for revalidation: supporting information for appraisal and revalidation*. London: General Medical Council, 2012.
12. General Medical Council. *Good medical practice*. London: General Medical Council, 2013.
13. Finlay K, McLaren S. Does appraisal enhance learning, improve practice and encourage continuing professional development? A survey of general practitioners' experiences of appraisal. *Qual Prim Care* 2009;17:387–95.
14. Pawson R, Greenhalgh T, Harvey G, *et al*. Realist review—a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;10:21–34.
15. Kastner M, Estey E, Perrier L, *et al*. Understanding the relationship between the perceived characteristics of clinical practice guidelines and their uptake: protocol for a realist review. *Implement Sci* 2011;6:69.
16. Wong G, Greenhalgh T, Westhorp G, *et al*. RAMESES publication standards: realist syntheses. *BMC Med* 2013;11:21.
17. Pawson R, Greenhalgh T, Harvey G, *et al*. Realist review–a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;10(suppl 1):21–34.
18. Kastner M, Estey E, Perrier L, *et al*. Understanding the relationship between the perceived characteristics of clinical practice guidelines and their uptake: protocol for a realist review. *Implement Sci* 2011;6:69.
19. Ritzer G. *Meta-theorizing in sociology: Lexington Books*, New York, 1991.
20. Roen K, Arai L, Roberts H, *et al*. Extending systematic reviews to include evidence on implementation: methodological work on a review of community-based initiatives to prevent injuries. *Soc Sci Med* 2006;63:1060–71.
21. Pearson M, Hunt H, Cooper C, *et al*. Intermediate care: a realist review and conceptula framework. Final Report. NIHR Service Delivery and Organisation Programme, 2012.
22. Pearson M, Chilton R, Woods H, *et al*. Implementing health promotion in schools: protocol for a realist systematic review of research and experience in the United Kingdom (UK). *Syst Rev* 2012;1:48.
23. Pawson R. *Evidence-based policy: a realist perspective*. Sage, 2006.
24. Vidall C. Revalidation for chemo nurses: are you ready? *Br J Nurs* 2014;23:S3.
25. Schafheutle EI, Hassell K, Noyce PR. Ensuring continuing fitness to practice in the pharmacy workforce: understanding the challenges of revalidation. *Res Social Adm Pharm* 2013;9:199–214.
26. GMC. Revalidation gets its own appraisal. In: Billings J, ed, 2014 http://gmcuk.wordpress.com/2014/03/05/revalidation-gets-its-own-appraisal/
27. Breen KJ. For debate. *Med J Australia* 2014;200:153–56.
28. Ghorpade J. Managing five paradoxes of 360-degree feedback. *Acad Manag Exec* 2000;14:140–50.