

Evaluating the strategic impact of medical revalidation

Building an evaluation framework

Julian Archer, Sam Regan De Bere, Nicola Brennan & Nick Lynn

FINAL REPORT

December 2013



Table of Contents

List of Figures	iii
Executive Summary	iv
1. Background.....	1
1.1 What is revalidation?	1
1.2 Why evaluate revalidation?	2
1.3 What is evaluation?.....	3
1.4 Evaluating revalidation from a regulatory perspective	5
2. Methods.....	7
2.1 Literature Reviews.....	7
2.2 Interviews/Focus Groups with Stakeholders	8
2.3 Cultural Historical Activity Theory (CHAT)	9
3. Results & Discussion	12
3.1 Literature Reviews.....	12
3.2 Analysis of stakeholder interviews & focus groups	17
4. The Revalidation Evaluation Framework	35
4.1 Introduction.....	35
4.2 Evaluation Questions	38
4.3 Work Streams.....	40
4.4 Summary of Framework.....	62
5. Commissioning Considerations.....	64
5.1 The use of CHAT to help shape commissioning	64
5.2 Commissioning in practice	65
6. Conclusions.....	66
7. References.....	67

Appendices

Appendix 1: Interview Schedule.....	87
Appendix 2: Ethics approval letter	89
Appendix 3: Ethics approval letter – Focus Group Amendment	90
Appendix 4: Ethics application	91
Appendix 5: Information sheet for participants.....	98
Appendix 6: Consent form for participants.....	100
Appendix 7: Invitation letter to participate in research	101
Appendix 8: Completed Interviews and Focus Groups	103
Appendix 9: Evaluation methods that were investigated but not included in the framework	104
Appendix 10: Likely minimum sample sizes for qualitative interviews and audio-recorded appraisals if conducted annually for 3 years to achieve data saturation	109

List of Figures

Figure 1: Revalidation from a regulatory perspective	5
Figure 2: GMC design features of revalidation	6
Figure 3: Engestrom's diagrammatic illustration of activity theory adapted for revalidation	10
Figure 4: Flow diagram of revalidation evaluation framework	36
Figure 5: Pre-revalidation trend in negative outcomes.....	108

Executive Summary

1. Background

On 3rd December 2012, the regulation of medical doctors in the UK by the GMC underwent its most radical reform in 150 years with the introduction of revalidation. This initiative presents a unique and exciting opportunity to evaluate the design features of an innovative regulatory system; in order to assure and potentially further develop revalidation as a robust regulatory process.

Aims & objectives The aim of this research was to develop a framework for evaluation for consideration by the GMC encompassing suggestions/recommendations which could capture information about the key design features of revalidation over which the GMC has jurisdiction. The evaluation should illuminate the central processes of the policy as they play out in practice, and generate an evidence base for examining how specific processes could be developed or enhanced in the future. Therefore, specific objectives of the research were to:

1. Review the current literature
2. Explore with key stakeholders the focus, purpose and potential benefits of revalidation and its evaluation
3. Draw on the above to develop an evaluation framework
4. Outline various methods that could appropriately populate the evaluation framework.

2. Methods

A mixed method approach was used to develop the framework. This included:

1. A narrative review of the literature in two key areas:
 - a. medical revalidation, relicensure and recertification
 - b. the effect of appraisal on performance improvement
2. Interviews/focus groups with representatives of the GMC, and other key stakeholders including the departments of health in the four devolved nations, patient organisations, medical bodies and employers
3. Inclusion/exclusion of arising methodological designs into an evaluative framework underpinned by Cultural Historical Activity Theory (CHAT) as the conceptual framework.

3. Results

Literature review A relatively disparate research literature has focused on different aspects of the revalidation process, and little documentary evidence of revalidation's potential benefits moves beyond appraisers/appraisees' perceptions of the policy. The mainstream literature, carried out in primary care using questionnaires or qualitative interviews, anticipates increased employee motivation and performance. A significant body of literature explores the association between recertification in the USA and clinical outcomes. However

this American literature cannot be translated to revalidation: in the US participation is voluntary and is based on specialty, whereas revalidation is compulsory and generic for all doctors. Notwithstanding these differences, there have also been claims that there is limited evidence to establish a causal link between recertification and better patient outcomes.

Stakeholder interviews Thematic analysis of stakeholder interviews (n=24) and focus groups (n=2) revealed no clear consensus among the participants about what revalidation is intended for or what it will achieve. Interviewees suggested that evaluating revalidation could be “difficult” and that trial type methodologies would be problematic. Specifically it was highlighted that the global implementation of revalidation would negate the use of controlled studies and that revalidation judgement decisions are difficult to measure because they are complex and do not result in definite outcomes (pass/fail). Furthermore, performance concerns are raised about relatively few doctors, compared to the professional group as a whole. Comparative studies would therefore be difficult to power without very large sample sizes.

Stakeholders identified five key areas for evaluation:

1. Supporting information – Questions were identified around the quality of supporting information; whether doctors experienced difficulty in collecting it, its impact on practice and whether doctors were using supporting information for reflection. Some felt that examining the paperwork required for appraisal, i.e. the appraisal forms, would be an important way to explore some of these issues.
2. Appraisal – Stakeholders described appraisal as a “closed door”. They suggested that, while appraisal is core to revalidation, its purpose and function is poorly understood and varies from person to person. In order to understand the “mechanics” of appraisal and support its future development (such as through evolving guidance) stakeholders felt that appraisal in action should be “witnessed”.
3. RO judgement-making – Stakeholders identified the central importance of the RO judgement-making process in revalidation. They talked of the need to understand the RO’s “craft” as they undertake this fundamental role within revalidation.
4. PPI – stakeholders questioned the role of PPI in revalidation and the underlying tension and uncertainty amongst stakeholders as to how far PPI should feature in revalidation reinforced the importance of exploring this issue.
5. Sampling considerations – most stakeholders expressed a view on whether some groups of doctors might be disadvantaged by revalidation, and how “challenging” it might be to ensure inclusivity. Very few however offered concrete suggestions as to how this might be undertaken. Stakeholders felt that the evaluation needed to be longitudinal running for 3-5 years, concluding in time to inform the next revalidation cycle.

CHAT as a conceptual framework CHAT is proposed as a conceptual framework to map together the different dimensions of revalidation for data collection and analysis. CHAT affords an understanding of complex activities within systems, demonstrating how systems impact on activity (in this case revalidation activity), and vice versa. CHAT will help the GMC draw simple messages out of complex systems.

4. The Revalidation Evaluation Framework:

The evaluative framework was developed by triangulating the core findings from the literature, stakeholder interviews and underpinning revalidation standards, as outlined by the GMC. Within this framework a series of important evaluation questions have been established that could be explored by four overarching work streams. In summary:

Work Stream 1 - supporting information Work stream 1 aims to capture the nature, quantities and uses of supporting information through evaluative measures including a literature review, a survey and semi-structured interviews with appraisees and appraisers, audio-recording appraisal meetings and a documentary analysis of appraisal forms. These approaches will capture **how GMC guidance is being applied, what information doctors actually collect, what information they value, in what ways they reflect and change their practice in light of this information and how this is evolving over time.**

Work Stream 2 - appraisal Work stream 2 aims to explore the activity, quality and impact of appraisal through evaluative measures including a literature review, secondary data analysis of appraisal rates, a survey and semi-structured interviews with ROs, as well as a survey and semi-structured interviews with appraisees and appraisers. These approaches will capture **what appraisal activity is happening, when and how GMC guidance (e.g. GMP framework for appraisal and revalidation) is being used, and how revalidation is shaping professionalism through the adoption of the values and principles of *Good Medical Practice*.**

Work Stream 3 - RO judgement-making Work stream 3 attempts to understand the craft of ROs as they make their judgements through a literature review, a survey and semi-structured interviews with ROs, ELA's and other relevant personnel, and a case control study involving doctors who have been referred to the FTP procedures. These approaches will establish **how revalidation supports the early identification of concerning practice, how ROs make their decisions in relation to protocol, recommendation statements and processes.**

Work Stream 4 - PPI Work stream 4 explores the role of patient and public involvement in revalidation using a literature review, interviews with key stakeholders in revalidation, and a survey and in-depth semi-structured interviews with patients and the public. These approaches will explore **the opportunities for PPI to be reflected in revalidation processes and the level of involvement patients and the public want in revalidation.**

A mixed-methods approach is proposed for each work stream in order to address effectively all of the proposed evaluation questions. In addition this approach would add a further layer of robustness to the framework, as it would enable the triangulation of findings and different perspectives.

In order to explore issues of equity and inclusivity across practice settings, and to assure transferable findings, it is proposed that the work streams are implemented in six revalidation evaluation centres across the four nations (England, Northern Ireland, Wales and Scotland). The proposed sampling strategy enables the exploration of protected characteristics in the doctor population, including categories of gender, age, ethnicity etc.,

settings and roles, for example doctors as locums. A longitudinal design could be adopted within each work stream, collecting data over a three year period.

5. Commissioning Considerations

Strategic decisions will need to be made by the GMC, in terms of the parts of the proposed Revalidation Evaluation Framework that will be commissioned and the time frame. Funding is a limitation in any research design and this will have an impact on the implementation of the evaluation. While it is unlikely that all of the framework would be initiated as a longitudinal study, the GMC in collaboration with academic partner(s) could make strategic decisions about what to commission when and what should then follow.

By mapping results from initial research onto the framework informed by CHAT it will be possible to ascertain where the gaps in knowledge are over time. Revalidation is a complex intervention in the complex system of healthcare delivery. To this end our evaluative framework, for consideration by the GMC, has been underpinned by a conceptual framework that seeks to move thinking away from isolated research activities to a more robust multi-method approach.

It is important to try to capture information about as many of the revalidation processes as possible in order to be able to understand what is really happening as revalidation is implemented. While any proposed method will provide useful insights into aspects of each of the activity systems (components of revalidation) it is only with a systematic approach over time and through a range of methods that the GMC will be able to populate evidence across the complexity and begin to understand what is happening within and between revalidation activity systems as they play out and interact with each other over time.

6. Conclusions

By working with the GMC and other key stakeholders, it has been possible to develop an innovative evaluative framework that will explore key questions of revalidation policy in action. The framework proposes a range of qualitative and quantitative methods to understand revalidation at four key points of inquiry across revalidation activity. These processes are: the uses of supporting information, appraisal, the judgement-making of ROs and the role of PPI in revalidation

The mixed-methods, multi-perspective approach proposed in the framework will support the ability to triangulate results, adding rigour to the transferability of findings into future policy and practice. The use of CHAT as a conceptual framework acknowledges the complexity of revalidation as activity in a complex system and the need for a combination of evaluative approaches, rather than singular or series of unrelated measures.

This framework will allow the GMC to understand the process and early regulatory impact of revalidation in practice and will establish opportunities by which the effectiveness and efficiency of revalidation design features might be improved in the future.

1. Background

1.1 What is revalidation?

Since 1858, the General Medical Council (GMC) has regulated the medical profession through the addition of doctors' names to a professional register.¹ This long-standing historical tradition assumed that doctors would maintain their fitness to practise throughout their careers.² In 2001 and 2004 this capacity to self-regulate was called into question when reports arising from the public inquiries into the Bristol Royal Infirmary and Harold Shipman were published.³ These reports demonstrated that trust alone was no longer a sufficient guarantee of a doctor's fitness to practise. Inside and outside of the profession the call was made for trust to be backed by objective assurance.¹

By this time the GMC were already developing plans for a new regulatory system that would require all doctors to be licensed to practise and to have their fitness to practise reviewed every five years: licences would only be renewed if doctors satisfied the new requirements.⁴ The proposed policy intervention that would guarantee objective assurance was entitled revalidation.⁵ In 2002, the GMC was empowered to introduce revalidation under the Medical Act (Amendment) Order.⁶ After a series of delays revalidation was finally implemented across the whole of the United Kingdom in December 2012. The GMC's aim is that all licensed doctors will be revalidated by the end of an initial revalidation cycle by March 2018.⁷

Revalidation, as it currently stands, rests on the continuing workplace evaluation of a doctor's fitness to practise. To maintain their licence doctors are required to participate in activities that support revalidation.⁸ In each workplace – a 'Responsible Officer' (RO) has been appointed under *The Medical Profession (Responsible Officers) Regulations 2010*⁹ and *The Medical Profession (Responsible Officers) Regulations (Northern Ireland) 2010*.¹⁰

The statutory duty of the RO is to ensure that systems for appraisal and the mechanisms for collecting and holding related information are in place. The RO regulations give ROs statutory responsibility for specified doctors to ensure that they are appraised annually. Every five years, the RO makes a recommendation to the GMC about the doctor's fitness to practise informed by a doctor's participation in five annual appraisals.¹¹ The GMC decision to license doctors is based on the RO's recommendation: this is the act of revalidation. In the event of concerns ROs are empowered to refer doctors to the GMC so that their fitness to practise can be investigated further.⁸

The GMC provides doctors with specific guidelines on the supporting information required¹² and content of appraisals for the purposes of revalidation.¹³ During the appraisal event, doctors are required to discuss their practice with their appraiser and use supporting information to demonstrate that they are continuing to meet the principles and values set

out in the GMC's core professional guidance: *Good Medical Practice*.¹⁴ *Good Medical Practice* requires that doctors provide supporting information across four domains:

1. Knowledge, skills, and performance
2. Safety and quality
3. Communication, partnership, and teamwork
4. Maintaining Trust.

Revalidation requires doctors to provide evidence of their on-going performance through reference to six types of supporting information¹²:

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.

1.2 Why evaluate revalidation?

While the process of revalidation in terms of regulatory policy is well established, there is little evidence about how revalidation is establishing itself and evolving in practice. Furthermore, little is known about how revalidation might need to be better supported and developed in the future in terms of mandated requirements, associated guidance and local/national infrastructure.

Any lack of clarity around purpose, guidance, support networks or future policy direction may result in unintended as well as intended consequences. As over time, these unintended consequences have the potential to impact upon or drive the process (the policy in action) rather than simply being the products of it.¹⁵ So it is paramount that complex interventions such as revalidation are evaluated to understand how they 'play out' in practice in relation to how they were initially proposed. This also provides opportunities to further develop the policy and associated guidance, and establish an increasingly nuanced understanding of the impact of the intervention in relation to the desired aims.

Lessons can potentially be learned from other medical regulatory systems around the world. A literature review conducted by the research team in the Collaboration for the Advancement of Medical Education Research & Assessment (CAMERA)¹⁶ to examine medical regulation across eight countries (UK, Ireland, Australia, Canada, USA, Germany, Hungary, Italy, New Zealand) found that while there were similarities between systems, there were also substantial differences. Notwithstanding the fact that revalidation shares a common goal to protect the public, by encouraging doctors to maintain their clinical knowledge and

skills, professional attitudes, and behaviours through on-going assessment,¹⁷ revalidation remains unique in that it is a national solution, mandatory for all doctors and overseen by a national single regulator.

The introduction of revalidation in the UK therefore brings with it a singular opportunity to evaluate an innovative regulatory system in action. While the early development of the medical revalidation process was piloted, these pilot schemes focused mainly on the interaction between employer organisations and their internal systems for performance review and other design decisions within revalidation implementation.^{18,19,20} The roll out of revalidation over the coming years provides an exciting opportunity to identify ways in which the effectiveness and efficiency of revalidation processes could potentially be developed further. By engaging in evaluative research it will be possible to develop an evidence-base of the relative merits and weaknesses of revalidation processes and begin to explore related consequences to healthcare delivery both intended or otherwise. This new knowledge can then be used to further develop policy and practice.

1.3 What is evaluation?

Evaluation is:

‘(the) systematic collection of information to explore effectiveness and characteristics of programmes/interventions, to improve outcomes and to examine worth and value’.^{21: 229}

Evaluation can be further defined as ‘a study designed and conducted to assist an audience to assess an object’s merit and worth’.^{22: 11} A variety of evaluation models are presented in the evaluation literature. ‘Evaluation models stipulate the question that a given type of evaluation seeks to answer; it also specifies how to set up the criteria for assessment’.^{23: 448}

In general, evaluation models fall into several categories:

- process models
- outcome models
- system models
- economic models
- actor models
- programme theory models.

Early in the research, the GMC expressed a desire that the evaluation framework should be informed by both process and outcome evaluation models: the two are briefly explained here as is the distinction between them.

Process evaluation explores how programme activities are delivered. It can help to determine the extent to which an intervention is implemented as planned and whether it reached the target participants. The quality of implementation is critical to maximising the

intended benefits and demonstrating strategy effectiveness.²⁴ Conducting process evaluation activities alongside programme implementation and using the results of these activities to perform continuous quality improvement is often overlooked.²⁵ A critical step in the success of any intervention is to establish that the various activities involved in implementing the intervention are carried out as planned. Process evaluation can contain a combination of qualitative and quantitative data collection strategies and can be carried out with or without outcome evaluation.²⁵

When designing a process evaluation the types of questions asked are different from those in outcome evaluations. Typically process evaluation questions include²⁵, but are not limited to:

- What intervention activities are taking place?
- Who is conducting the intervention activities?
- To what extent was the programme implemented as planned?
- Who is being reached through the intervention activities?
- What are barriers to programme delivery?
- What inputs or resources have been allocated or mobilised for programme implementation?
- What are possible programme strengths, weaknesses, and areas that need improvement?
- How is the programme received by the target group and programme staff?

A process evaluation may also include 'process outcomes'. Process outcomes do not rely on longitudinal variations in an outcome variable (i.e. the measurement of outcomes at different points in time) but depict the status or situation of participants after taking part in a programme.²⁶

Outcome evaluation 'tests a series of hypotheses concerning the intended changes by (1) making a comparison between conditions before and after participation in a programme (2) comparing individuals who participated in a programme with similar individuals who have not or (3) a combination of both. Outcomes can include short-term, intermediate as well as long term results'.^{26: 4}

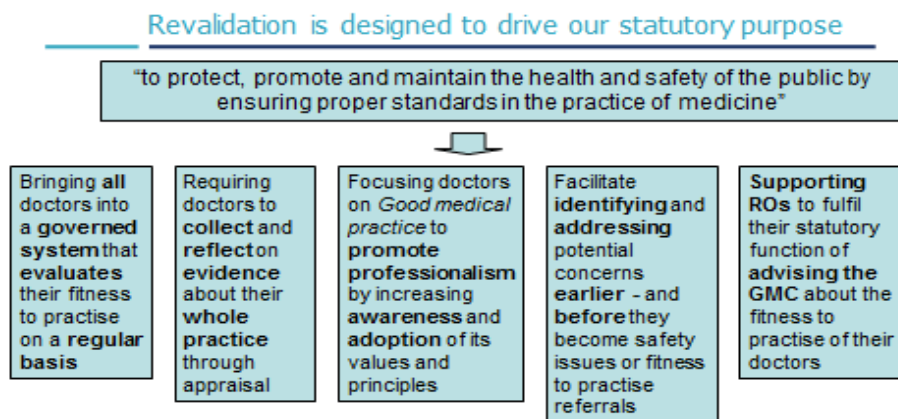
By using a process evaluation model to evaluate revalidation it will be possible to assess whether the various interconnected design features of revalidation are being carried out as proposed, and this can, in turn, inform decision-making about potential future changes to elements of the process, if required. An outcome evaluation model would assess the impact of revalidation.

1.4 Evaluating revalidation from a regulatory perspective

Revalidation has been designed to drive the GMC's statutory purpose "to protect, promote and maintain the health and safety of the public by ensuring proper standards in the practice of medicine" (Figure 1).²⁷ Specifically the mechanisms of revalidation are intended to:

1. Bring all doctors into a governed system that evaluates their fitness to practise on a regular basis
2. Require doctors to collect and reflect on evidence about their whole practice through appraisal
3. Focus doctors on *Good Medical Practice* to promote professionalism by increasing awareness and adoption of its values and principles
4. Facilitate identifying and addressing potential concerns earlier – and before they become safety issues or fitness-to-practise referrals
5. Support ROs to fulfil their statutory function of advising the GMC about the fitness to practise of their doctors.

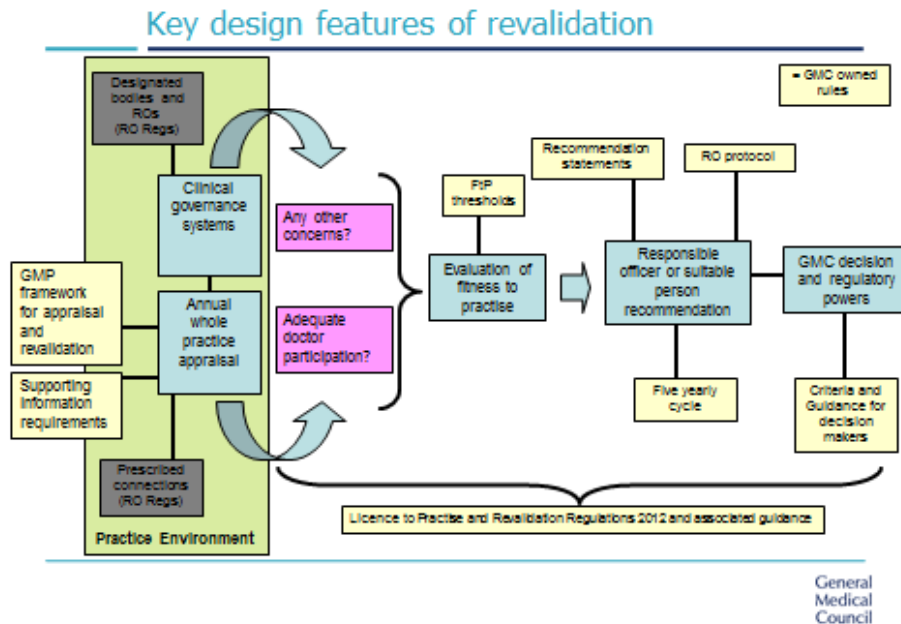
Figure 1: Revalidation from a regulatory perspective



The key design features of revalidation are outlined in Figure 2. While revalidation is a regulatory mechanism governed by the GMC its effective implementation requires the GMC to work in partnership with the NHS and its clinical governance systems. There are design features that the GMC have direct control over, i.e. the *GMP framework for appraisal and*

revalidation, the supporting information requirements and the RO protocol, but there are also areas which they do not, i.e. clinical governance systems and the actual oversight of appraisal. The design features owned by the GMC are highlighted in yellow.

Figure 2: GMC design features of revalidation



1.4.1 Aims & objectives of evaluation

The aim of this research was to develop a framework for evaluation for consideration by the GMC encompassing suggestions/recommendations which could capture information about the key design features of revalidation over which the GMC has jurisdiction. Such evaluation should illuminate the central processes of the policy as they play out in practice and generate an evidence base for examining how specific processes could be developed or enhanced in the future. It looks in particular to examine the value added to existing structures of supporting information, appraisal, recommendation processes and the role of PPI. Specific objectives of the research were to:

1. Review the current literature
2. Explore with key stakeholders the focus, purpose and potential benefits of revalidation and its evaluation
3. Draw on the above to develop an evaluation framework
4. Outline various methods that could appropriately populate the evaluation framework.

The methods utilised to address these objectives are delineated in the following section.

2. Methods

A mixed methods approach was used to inform the development of the framework. This included:

1. a narrative review of the literature in two key areas:
 - a. medical revalidation, relicensure and recertification
 - b. the effect of appraisal on performance improvement
2. interviews/focus groups with representatives of the GMC, and other key stakeholders including the departments of health in the four devolved nations, patient organisations, medical bodies and employers
3. inclusion/exclusion of arising methodological designs into an evaluative framework underpinned by cultural-historical activity theory (CHAT) as the conceptual framework.

2.1 Literature Reviews

2.1.1 International literature on revalidation, relicensure and recertification

A narrative approach to the literature review was considered most appropriate in order to investigate the impact of medical regulatory interventions as it describes and synthesises existing evidence.²⁸ The search strategy involved combining scoping searches of Google Scholar and Medline, with the CAMERA research team's expert knowledge of the field. This was then supplemented with 'ancestry' and 'forward citation' searches where appropriate. In ancestry searches, the reference list of a relevant article is scrutinised to identify additional relevant citations.²⁹ In a forward citation search, the researcher takes an identified citation or reference and tracks its after-life to identify which other subsequently published sources have cited it.

2.1.2 International literature on the effect of appraisal on doctors' performance

In the case of the international literature on the effects of appraisal on doctors' performance, our narrative review synthesised empirical literature on the appraisal process within the remit of medical regulation. The literature was reviewed between 2007 to the present and updated a previous review by Overeem et al.³⁰ Overeem's study reviewed the effectiveness and feasibility of methods of formative assessment used, in routine practice, to assess the performance of individual doctors: in eleven studies, the use of portfolios or appraisal were investigated.³¹⁻⁴¹ The search strategy involved combining scoping searches of Google Scholar, Medline and EMBASE for the period 2007-2013. These were searched using combinations of the following free-text search terms: 'appraisal', 'performance', 'medicine' and 'effectiveness'. The following MESH search terms were also used: 'employee performance appraisal/', 'empirical research/'. The results were combined with the CAMERA research team's expert knowledge of the field and then supplemented with ancestry searches and forward citation searches where appropriate.

2.2 Interviews/Focus Groups with Stakeholders

Scoping interviews and focus groups were undertaken with principal stakeholders in revalidation to consider what was important to them in any evaluative framework for revalidation. Engaging stakeholders is an important step in any evaluation.⁴²

Interviews continue to be a source of valuable information for social researchers; although the use of interviews as a 'technology for eliciting a range of psychologically and linguistically interesting responses'^{43: 218} has been reassessed. The revised view is that interviews are: 'arenas for interaction between two or more parties...analysing them in the same way that we might a telephone conversation between friends or the cross-examination in a court room.'^{43: 218}

Rather than negating the presence and input of the interviewer, the revised view actively acknowledges dialogue taking place. This dynamic and inherently dialogic element is also a feature of focus groups.

Focus group methodology is 'a method that facilitates the expression of criticism and the exploration of different types of solutions is invaluable if the aim of research is to improve services'^{44: 300}

The composition of the focus groups was guided by the literature review. Focus group research is an excellent method for eliciting respondents' opinions, feelings, beliefs, experiences and reactions.⁴⁵ The dynamics of a group context can elicit a wide range of responses that can then be discussed by the participants: thus can be useful in revealing where there is consensus or disagreement. This method is often frequently used to capture potentially diverse and complex perspectives.

A semi-structured approach was undertaken in the interviews and focus groups to allow participants to talk freely but with focus. The interview schedule has three sections (appendix 1). The first explored with participants was the purpose of revalidation and how they thought its implementation might change or shape healthcare. The second focused on developing measures that might be used to evaluate revalidation; in particular, the five domains outlined by the GMC and how revalidation may differ for various groups of doctors. Finally, the third section invited interviewees to prioritise their ideas and to consider an appropriate timeframe for evaluation. The focus group schedule was a shortened version of the interview schedule and focused on questions 3, 4, 6, 7 and 9 (also appendix 1).

The participants were identified as principal stakeholders i.e. people or organisations with a stake in the revalidation process including: appraisers, appraisees, responsible officers, patient groups, relevant GMC departments, the Care Quality Commission, the Royal Colleges, the British Medical Association, Ethnic Minority groups and the Departments of Health. Participants came from all four of the devolved nations. A purposive sampling strategy was adopted⁴⁶ and the research team developed a list of potential interviewees:

this was further developed in discussion with the GMC, who also provided the names of individuals in a number of organisations who might be approached.

Ethical approval for the research was granted by the Plymouth University Faculty of Health, Education and Society Research Ethics committee (appendix 2). A later amendment was made to allow for the focus groups (appendix 3). The application process involved completing an in-depth application form (appendix 4) outlining the purpose of the research and its ethical implications as well as the development of the study information sheet (appendix 5) and consent form (appendix 6) for participants.

A joint letter of invitation from the GMC and the CAMERA research group was developed and sent to potential participants (appendix 7). The letter included the approved information sheet and consent form.

Twenty-four individual interviews and two focus groups were conducted with stakeholders between June and September 2013 (appendix 8). The first focus group involved 8 participants from across different departments of the GMC. The second was with three participants from the GMC ethnic and diversity working groups. The interview and focus group data were analysed using a thematic analysis.⁴⁷ Detailed notes were taken of the emerging themes and triangulated within the team.

The data generated from the interviews and focus groups and the literature reviews (presented in the following section) were then triangulated and augmented with the expertise of the research team to develop the proposed evaluation framework.

2.3 Cultural Historical Activity Theory (CHAT)

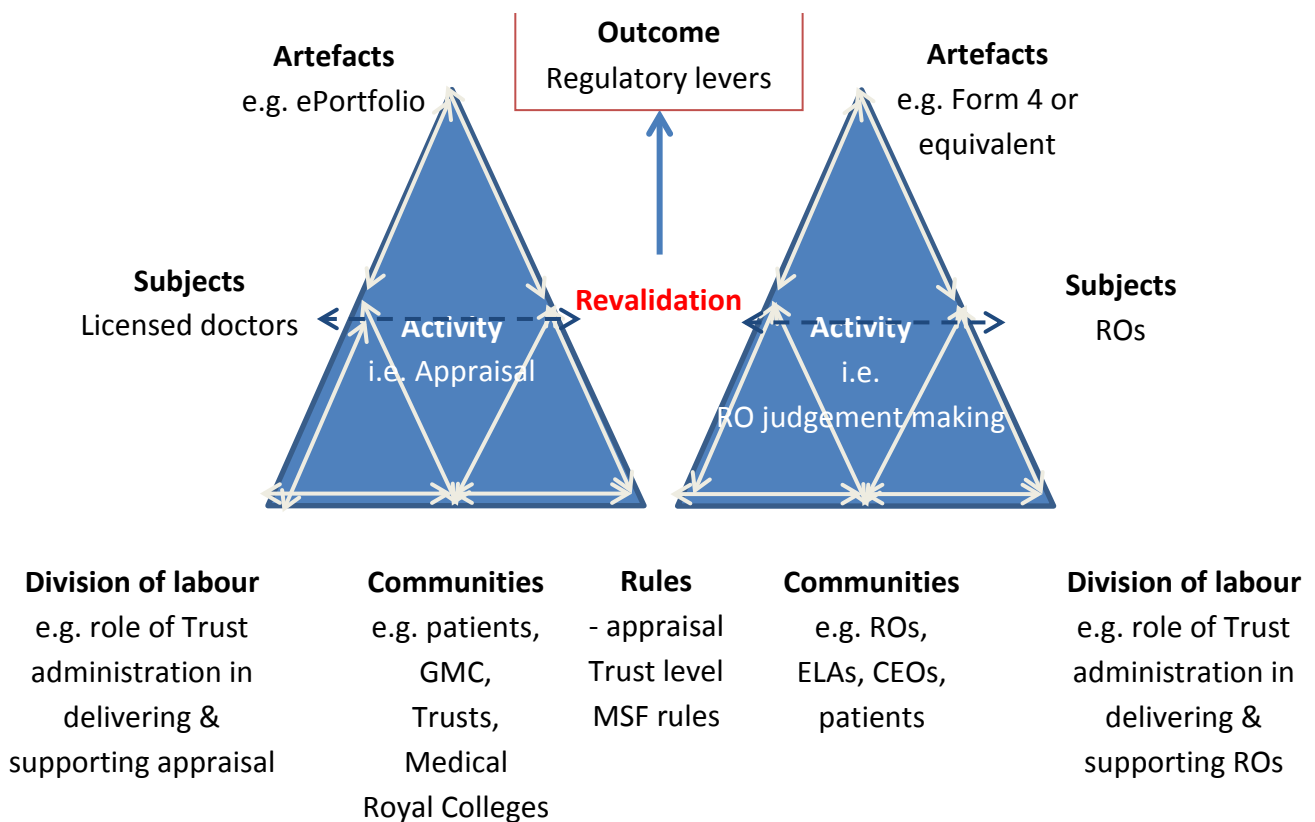
2.3.1 Framing revalidation as activity in the wider system of healthcare

There are a number of potential frameworks available within the policy and wider educational literature that might help researchers decide what and how revalidation could be evaluated. In policy research, simple frameworks have been developed that seek to understand 'actors' (individuals and organisations) as a triangle of context, content and process within policy implementation.⁴⁸

In the development of this framework, we have proposed that any approach to an evaluation should have the potential to bring together the many strands of evaluative research ideally needed to fully understand revalidation. Drawing on policy impact⁴⁸ and social discourse theories,⁴⁹ Cultural-Historical Activity Theory (CHAT) is one such theoretical framework. CHAT⁴⁹ is presented as 'an integrated road map for educational research and practice'⁵⁰ and has been frequently adopted by workplace theorists⁵¹ including in medicine.⁵² CHAT theorises that people continually shape, and are shaped by, their social and cultural contexts.

CHAT is relevant to examining revalidation because it provides an accessible and flexible framework with which to identify and examine interactions, as well as components of, a system as it is implemented. Engstrom's⁵³ summary of an activity system forming the basic unit of analysis in activity theory is shown in Figure 3. We have adapted it here to demonstrate its applicability to revalidation evaluation, by exploring just two of four possible work streams we present later; the interrelated systems of appraisal and RO judgement-making, within the 'practice' of revalidation.

Figure 3: Engstrom's diagrammatic illustration of activity theory adapted for revalidation



In this model, *Artefacts* are mediational means or instruments (e.g. an adopted ePortfolio), employed by the *subjects* (doctors and ROs) to achieve the object (revalidation). The *object* is the central issue to which *activity* is directed (revalidation), which leads to an *outcome* (the desired series of changes e.g. earlier identification of doctors in difficulty), as a consequence of *activity*, appraisal and or ROs judgement making.

The *community* comprises multiple individuals and/or sub-groups who share the same general object of delivering revalidation. Our communities include patients and the public, practising doctors, appraisers, ROs and wider stakeholders such as CEOs and Royal Colleges etc. The *division of labour* or *roles* refers to both the horizontal division of tasks between the members of the community and to the vertical division of power and status. Engstrom intended that *rules* refer to the explicit and implicit regulations, norms and conventions that

constrain actions and interactions within the activity system.⁵⁴ These rules are both explicit in medical regulation and convention, and implicit in professional and institutional culture.

Understanding activity such as appraisal or RO judgement-making within a system is as much about understanding the interactions between all of the components within and between systems (as well as the components themselves). This is illustrated by the multiple arrows on the activity system (Figure 3). So, while the main activity is seen across the middle of the triangles with subjects (doctors or ROs) undertaking activity (appraisal or judgement making) to achieve an object (revalidation) and ultimately potentially a series of regulatory outcomes there are important interactions between all the components.

It is important to try to capture information about as many of these practices as possible, in order to be able to understand what is really happening as revalidation is implemented. While any proposed method will provide useful insights into aspects of each of the activity systems (components of revalidation), it is only with a systematic approach over time - and through a range of methods - that the GMC will be able to populate evidence across the complexity and begin to understand what is happening within and between revalidation activity systems as they play out and interact with each other over time. CHAT will help the GMC draw simple messages out of complex systems.

Activity theory is also helpful in evaluative models in terms of its optimism towards positive change, in that subjects, objects and artefacts may be modified through action. It encourages people with fixed characteristics (attitudes, demographics and abilities) to achieve new things. Part of this for example might be about rethinking the tools at their disposal; techniques (like an appraisal meeting) which could be seen differently, rather than as a tool to achieve a positive outcome rather than simply as a set of rules to be followed. This provides an opportunity for the evaluation to interactively shape revalidation and not just simply observe it.

3. Results & Discussion

In this section we set out in detail the results that arose from the literature reviews and stakeholder interviews.

3.1 Literature Reviews

3.1.1 International literature on revalidation, relicensure and recertification

Despite its importance in policy terms, the research literature on medical revalidation, recertification or relicensure is limited. This observation is corroborated by a recent literature review conducted by the CAMERA research team for the Medical Council of Ireland on the maintenance of professional competence arrangements in Australia, Canada, Germany, Hungary, Italy, New Zealand, the UK and the USA.¹⁶ That review revealed a dearth of rigorous empirical research that specifically evaluates medical regulatory interventions including their potential impact on professional practice, quality improvement and patient safety.

In the UK, a diverse research literature exists on different aspects of the revalidation process. Some specific examples include Mugweni's⁵⁵ work on appraisal (discussed in following section 3.1.2), Wright et al.'s research on multisource feedback,⁵⁶ Murphy's examination of insightful practice,⁵⁷ and Ivers et al.'s work on quality improvement activities.⁵⁸

Wright et al.⁵⁶ investigated the psychometric properties of two questionnaires developed to provide multisource feedback to doctors from patients and colleagues. The study involved a cross-sectional design in 11 UK healthcare organisations from 2008-2011. Using the General Medical Council Patient Questionnaire (PQ) and Colleague Questionnaire (CQ), patients (n=30,333) and colleagues (n=17,012) rated the professional performance of 1,065 practising doctors. The factors influencing patient and colleague responses on questionnaire items were explored using regression modelling. The study found that while the questionnaires demonstrated satisfactory internal consistency, test-retest reliability, and convergent validity, patient and colleague ratings were heavily biased toward favourable impressions of doctor performance. The authors concluded that multisource feedback obtained using these questionnaires should not be used in isolation to inform decisions about a doctor's fitness to practise medicine.

Murphy et al.'s⁵⁷ study tested a method in which 60 general practitioners (GPs) and 12 GP appraisers, in the Tayside Region in Scotland, were assessed on their reflection and response to a set of external feedback. Facilitated by a website called Tayside In-Practice Portfolio, GPs collected specified data on patient and colleague opinion, open book self-evaluated knowledge test and complaints. External practice-level data including clinical quality and prescribing safety was also generated and provided to the GPs. GPs were then

asked to reflect on this specified suite of feedback in a portfolio to be submitted for appraisal. In order to ensure the content validity of the feedback and whether it covered the required GMC attributes of Good Medical Practice, GPs' carried out a mapping exercise before and after the study to assess the feedback tool's capacity to test the GMC attributes. GPs then underwent appraisal and GPs' professionalism was examined by their appraiser's assessment of their level of insightful practice. This was defined as: engagement with, insight into and appropriate action on feedback data. Using Generalizability theory, the reliability of GPs' assessment of their insightful practice and subsequent recommendations on GPs' revalidation by face-to-face and anonymous assessors were investigated. The study found that face-to-face assessment was unreliable. Anonymous global assessment by three appraisers of insightful practice was highly reliable ($G=0.85$), as were revalidation decisions using four anonymous assessors ($G=0.83$).

Ivers et al.⁵⁸ conducted a systematic review to evaluate the effect of audit and feedback on the behaviour and performance of healthcare professionals. The review found that audit and feedback generally results in small but potentially important improvements in professional practice. The effectiveness of audit and feedback is dependent on baseline performance and how the feedback is actually provided.

However the majority of existing research relevant to medical revalidation is from the US, where the broad equivalent of medical revalidation is Maintenance of Certification (MOC).⁵⁹ ⁶⁰ Sponsored by the American Board of Medical Specialties (ABMS) and its 24 member boards, MOC requires that most medical specialists recertify on a periodic basis – mostly every 10 years. ⁶⁰ However, unlike revalidation, MOC is not compulsory and the retention of a doctor's licence to practise is not dependent on successful participation.

MOC schemes differ for each specialty board, however all consist of four core components⁶¹:

1. Evidence of professional standing: A doctor must hold a valid State Licence.
2. Evidence of lifelong learning and self-assessment: This continuing medical education component specifies that doctors must participate in the educational and self-assessment programmes that fulfil specialty-specific standards agreed by their member board.
3. Evidence of cognitive expertise: Through formal examination, doctors must demonstrate that they have the fundamental, practice-related and practice environment-related knowledge to provide quality care in their specialty.
4. Evidence of performance in practice: Doctors are evaluated in their clinical practice according to specialty-specific standards for patient care. Doctors are asked to demonstrate that they can assess the quality of care they provide compared to their peers and national benchmarks and then apply the best evidence or consensus recommendations to improve that care using follow-up assessments.

The research on MOC falls into three main categories.⁶² The first category explores whether there is a link between MOC and improved clinical performance and outcomes. Many studies have found that physician engagement in MOC activities is associated with enhancement in clinical competence, improvement in care processes and the gathering of valuable patient feedback.⁶³⁻⁷³ However a systematic review exploring the relationship between the certification of physicians and clinical outcomes, published in 2002, was inconclusive.⁷⁴ The review identified 13 papers⁷⁵⁻⁸⁷ containing 33 separate relevant outcomes published between 1966 and 1999. Of the 33 findings, 16 demonstrated a significant positive association between certification status and positive clinical outcomes, three revealed worse outcomes for certified physicians and 14 showed no association. But while the majority of studies showed a link between MOC and improved clinical performance and outcomes, these studies were almost exclusively observational and cross-sectional in nature and therefore fail to prove causation.^{88, 89}

The second category of research on MOC investigates support for MOC programmes. A growing body of literature suggests that the public⁹⁰⁻⁹⁴, as well as doctors, support MOC.^{68, 71, 95-103} However there are dissenting voices in the literature. The experts against MOC agree that the concept of recertification is sound but disagree about the process.¹⁰⁴ MOC draws strong criticism from physicians who assert that MOC is too expensive and the process too time-consuming.⁶⁰ There is concern about the requirement that a secure examination (one of MOC's four components) be completed without access to outside sources of information. This condition contradicts what medical students and residents are currently taught: that they should take advantage of the best sources of information rather than rely entirely on their memory.⁶⁰ MOC is also criticised for what is called the “grandfathering issue” and there are calls for all physicians to have to recertify. In most specialities existing physicians were “grandfathered” into Board MOC programmes and received time-unlimited credentials and thus do not have to re-certify every ten years. Among physicians who are currently board certified by the American Board of Internal Medicine, approximately 23% of those less than 70 years of age are either not required to or have chosen not to recertify.⁶² This demonstrates an inequality in the system. There are calls for all physicians be held to the same standard and demonstrate on a regular basis that their medical knowledge is up to date.¹⁰⁵

The third category of research on MOC explores its educational principles. The body of work demonstrates how the latest principles in adult learning such as self-directed practice improvement modules, simulations and interactive workshops are being incorporated into MOC activities.^{63-65, 68, 97, 98, 106-110}

While it would be desirable to draw on the MOC literature to inform an evaluative framework for revalidation, its value/relevance/transferability is limited. Firstly the MOC literature fails to provide any studies that prove causality e.g. that with the introduction of MOC, doctor’s performance or patient care has improved as a direct result. This is because

the MOC research is confounded by the lack of randomisation. MOC doctors have better outcomes but they are a self-selecting group as they have chosen to undergo the MOC process. Therefore causality cannot be demonstrated.

Other important differences between the two systems include that while revalidation has been informed by the Medical Royal Colleges, it is implemented by the UK's medical regulator, the GMC. Lastly, decisions on a doctor's fitness-to-practise are initially made at a local level and are woven into the appraisal process and NHS clinical governance systems.

In conclusion, the aim of this literature review was to identify research on medical revalidation, relicensure or recertification that could inform the design of the evaluation framework. Key findings relevant to the development of the framework were identified.

1. A diverse research literature exists on different aspects of the revalidation process however there is no comprehensive up-to-date evidence synthesis on each aspect of the revalidation process. This includes the six pieces of supporting information (1. CPD, 2. quality improvement activities, 3. significant events, 4. feedback from colleagues, 5. feedback from patients and 6. review of complaints and compliments) and appraisal itself. Synthesising the existing literature on these areas, particularly around their potential effect on doctors' performance, would provide an evidence base that would help inform decisions about the development of revalidation and would help prioritise the secondary research agenda in these areas.
2. The review as reported in this report has identified a large body of literature exploring the association between MOC and clinical outcomes; however this research does not provide any directly useful insights into causation or the causal mechanisms at work in regulatory processes. Rigorous in-depth qualitative methods could address this issue, and would generate the rich data necessary to understand the process of regulation in practice.

3.1.2 International literature on the effect of appraisal on performance improvement

Building on a wider scoping review, the purpose of this more concise narrative review was to update the Overeem et al.³⁰ review published in 2007. Overeem et al.'s review investigated the effect of appraisal (and other assessments) on doctors' performance. They found that the majority of doctors undertaking appraisal were satisfied with their evaluation and reported performance improvements. Since that review there have been a number of other studies published.

In 2008, Colthart et al.,¹¹¹ published a cross-sectional study of GPs that investigated the outcomes of GP appraisal and whether it resulted in changes in medical practice, education and learning, career development, attitudes to health and probity, how GPs organise their

work and their perceptions of the overall value of the process. The study found that appraisal can have a substantial effect on all aspects of a GP's professional life. In particular those GPs who placed a greater value on the process reported a sustained benefit in the management of their professional development; although some GPs perceived little or no benefit.¹¹¹

A small study by Wakeling et al.¹¹² also in 2008 used qualitative interviews to explore the views of 18 GPs on appraisal in Scotland. They found that the majority of interviewees obtained some benefit from appraisal. This mostly related to the development of learning needs and the support experienced through a confidential discussion with a colleague. Direct impact on clinical practice was described as limited. There were mixed views as to whether appraisal should become more summative, about the appropriate level of challenge and the standard of evidence required. The majority of GPs were complimentary about the skills of their appraiser and felt that appraisal by an appraiser from a different specialty would not be as useful.

In 2009, Zolle¹¹³ used a questionnaire study of GPs to explore the educational impact and quality of the appraisal at a primary care trust in Wessex. The study found that appraisees reported a growth in confidence and feeling valued in their jobs. They also reported an improvement in overall patient care which contributed for some in a postponement in retirement.¹¹³

A study by Finlay(2009) also investigated the attitudes of GPs towards the appraisal process and its impact on their learning, their practice and their CPD.¹¹⁴ The key findings, of their postal questionnaire in West Kent, were that 47.5% (n=131) of doctors said that participating in the appraisal process had enhanced their learning, 40.2% (n=111) thought the appraisal process had resulted in an improvement in their practice and 55.8% (n=154) stated that the appraisal process had encouraged their CPD. The thematic analysis of qualitative data generated in the open questions in the questionnaire highlighted that it was crucial that participants viewed appraisers as respected peers. Independence in appointing appraisers was considered to be important. However respondents felt the appraisal process was time-consuming and that 'protected time' was necessary to complete the paperwork and CPD engagement. The authors concluded that in order to maintain the confidence of patients and doctors, it was important to ensure the quality, consistency and the nature of appraisal.

In 2011, Mugweni conducted a literature review on the benefits that GPs thought appraisal had to offer.⁵⁵ GPs recognised that appraisal provided the opportunity to reflect on their personal development, and promoted their educational activity. Eight studies were included^{38, 111, 112, 114-119} offering both quantitative (n=4) and qualitative analysis (n=5) of the perceived benefits. This review found that there was a strong perception that appraisal promotes changes in clinical practice and offers other benefits such as mentorship and

motivational support for the doctor. The review concludes that GPs, and patients, would continue to benefit from the appraisal process after the revalidation is introduced.

In conclusion, there have only been a limited number of small studies exploring the impact of appraisal on performance and patient outcomes. Relatively little empirical evidence of revalidation's potential benefits exists, beyond opinions of its value from appraisers and appraisees. Most of the literature expects increases in employee motivation and thus performance improvement: including changes to levels of insight, reflective activity, and motivation which should all be considered as legitimate outcomes of the appraisal process.¹²⁰ Most of the research has been carried out in primary care settings and has been mainly limited to cross-sectional studies using questionnaires or qualitative interviews. The use of observational techniques to witness and capture what actually happens in the appraisal meeting would address many of the substantial knowledge gaps that exist.

3.2 Analysis of stakeholder interviews & focus groups

The analysis of the stakeholder interviews and focus groups highlighted a number of key themes. These are presented under the following headings in this section of the report:

1. What is revalidation for and how might it be evaluated?
2. Supporting information
3. Appraisal
4. RO judgement making
5. Influence of practice settings and roles
6. Patient and public involvement
7. Professionalism
8. Unintended consequences
9. Timeline for evaluation
10. Measures – qualitative or quantitative
11. Summary of thematic analysis of interviews

3.2.1 What is revalidation for and how could it be evaluated?

There was no obvious consensus among the participants about what revalidation is intended for or what it will achieve. This was as true of those who support the idea of revalidation as it was of those who are more sceptical of its aims. Some felt the aim of revalidation was to drive up standards in the practice of medicine:

“To raise clinical standards: It’s not a particularly complex view of it.”

“[A] mechanism that the individual doctors cannot avoid and must cooperate with ... it is a means to ensure proper standards of practice.”

Others thought it would provide a mechanism to identify or deal with poor performers:

“I think you have got a minority of doctors who are not very good and no one has really known what to do about them. I think what revalidation will do is it will give Trusts the mechanism and the infrastructure to deal with poorly performing doctors.”

Many believed it was about providing reassurance to the public and that doctors are up-to-date and fit to practise:

“It’s supposed to be helping the public feel confident that doctors are doing the right thing, that they’re safe to practise and patients are properly looked after ... I think it has a lot more to do with doctors being fit to practise and helping to develop doctors to become more reflective ... and picking up issues before they become concerns.”

A number of participants felt the arrival of revalidation was no guarantee that professional behaviours or standards of practice would change. Some felt the procedural elements of the revalidation process had the potential to be circumvented or undermined and to become a ‘tick-box’ exercise or a form of paperwork compliance^{121, 122}:

“Just because you collect the data you have to tick boxes doesn’t necessarily tell you that much I don’t think, I think it ticks the boxes it keeps everyone happy.”

“I suppose it could change healthcare ... if it’s working properly and it’s more than simply a tick box exercise...it should shape healthcare because it should shape practitioners.”

When strong proponents of revalidation were pressed to suggest measures to evaluate revalidation, they were less assured in their responses. What all those interviewed did agree upon is that evaluating revalidation is “difficult”. In particular there was significant agreement that trial methodologies would be very difficult. Potential outcome measures such as patient mortality present many challenges, in that trends are already established and it would be hard to prove any future change in the trend was caused by revalidation:

“We need to look at how [revalidation] can bring improvement in patient care, and actually I don’t think that’ll be measurable in the current structure because I don’t think that revalidation as it’s designed will actually quickly produce identifiable benefits in patient care. It might bring some but I don’t think it’ll bring big measurable benefits quickly.”

Referring to clinical care improvement:

“You’re never going to prove it. You may have two parallel lines on a graph showing improvement but to actually say it’s due to revalidation I think the people who are pushing that irrevocably, if that’s the case, are doing it for political reasons not because clinically you can prove it.”

Another participant, when asked about how he might evaluate revalidation, expressed these reservations:

“It depends what you mean by evaluate ... I’ve got a really deep scepticism that you will be able to find a measurable impact on fitness to practise through the achievement of the proxy markers. By which I mean how would we measure improvement of fitness to practise? The only way I can think of is by a reduced number of doctors going through fitness to practise procedures, in other words more people being identified and remediated in advance and yet so few doctors go through fitness to practise procedures that I’m not sure that any change in those numbers is going to be significant.”

Another participant noted, in a similar vein:

“What we have to say is well, okay, what is it we think revalidation is going to change amongst these doctors and how big a change can be legitimately estimated it might be? And then go from there and that’s difficult because I don’t think the things you’re trying to change are necessarily easily measurable through the dataset that will be collected from revalidation.”

Others were less sceptical and took a pragmatic view. While some were specific about where to look and the methods that might be used to evaluate revalidation; others could suggest where to look but were vague on precise methodologies:

“I think we need to have a measurement tool that measures whether performance of an individual doctor is reflected in their behaviours when working within a team. So, for example, the surgeon that is difficult to work with, will only do things in their own way and that is not conducive to a safe environment in which to be able to do that procedure, so we need to see whether colleague feedback is telling us something.”

“I think the only way to get quite specific information is through most of the qualitative work, ...ask people about how they feel about revalidation, ... what work they did because of revalidation and what behaviours they feel they changed.”

I think numbers is not relevant. What is relevant is competencies and I think that to go back and re-evaluate people in the workplace, as we do our trainees, is probably appropriate for senior doctors as well.”

“I think you could design some interesting studies that use quite a bit of the routine data of the appraisal and revalidation process in order to measure something.”

These four observations reveal the diversity of opinion among the participants as to where, in the revalidation process, we might go to unearth data that would be useful in any evaluation. These observations also remind us of the amount of information that revalidation and clinical governance generates. However what was also evident from the

interviews, and this was a specific question on the interview schedule, was the lack of existing datasets that could be used to effectively evaluate revalidation.

As we move on to consider what the participants had to say about ‘supporting information’, questions about the quantity and quality of the information generated and used in the revalidation process are common.

3.2.2 Supporting information

The research participants felt it was important to evaluate the supporting information that doctors are required to collect. In particular, there was some agreement around evaluating the quality of supporting information that doctors are required to bring to appraisal; whether doctors experienced difficulty in collecting each type of supporting information; which type of supporting information impacts most on practice; and whether doctors were ‘reflecting’ upon their supporting information:

“We need to know ‘Is any of this of any value and what is the quality of this evidence?’, ‘What’s the quality of the supporting information that doctors are bringing?’ ... ‘How much evidence do you need to bring?’, ‘How much work do you need to do to show that you’re up-to-date and fit to practise?’”

In making this observation the participants demonstrate they are alert to how the revalidation process might alter as it translates from policy to practice. They recognise too that numbers can be manipulated and decontextualized, and that ‘supporting information’ is equally malleable:

“There’s something about the quality of evidence which I think is really interesting because, obviously, you’ve only got to tick the box to say you’ve included it and there are very minimal requirements now for anyone to demonstrate any advanced reflection on their data.”

To evaluate effectively, it is important that the paperwork required for appraisal – the appraisal forms - are examined. This was felt by many interviewees to be an important way to explore how policy becomes practise:

“What I’d do is I’d go through an appraisal form as it is at the moment ... to see whether ... doctors [are] adhering to what is required for appraisals. So you’d look at each section and you’d take a sample of doctors and you’d look at whether they are completing [it]. For example, if they’re providing the supporting information are they providing clinical audits or discussion groups summaries? And so you’d go through each of the different sections and see quantitatively whether doctors are actually complying.”

Clearly, the context in which these documents are created is significant and any analysis must be done in conjunction with other information arising from the appraisal event.

Together, an analysis of appraisal *and* the supporting documentation makes for a strong evaluative procedure. Indeed, the participants were agreed that because appraisal was central to revalidation, it was central to any evaluation of revalidation.

3.2.3 Appraisal

In accepting that it was important ‘to evaluate whether revalidation is bringing all doctors into a governed system that evaluates their fitness to practise’, participants were unanimous in recommending that appraisal rates across the four nations be examined. Many also felt that some method of “*witnessing*” appraisal was required: “*we have to get into the appraisal room*”, to “*get behind the closed door of the appraisal.*”

Another observed that much of the revalidation process is currently beyond the gaze of researchers, observers, and regulators; and that this warrants the need for inquiry. One participant summed this up by posing the simple but challenging question:

“What goes on in the revalidation encounter?”

Another pointed out that:

“It’s a very closed discussion.”

While someone else remarked upon the potential for variation:

“Appraisals can be anything from a tick-box exercise to actually something really beneficial and meaningful. So I think it really does rely on not just the fact that it’s taken place but the experience of the person that’s being appraised, the experience of the person doing the appraisal, what it covers and again that’s an interesting idea to film the appraisal because that gives you a lot of information all in one go although I would imagine it would be quite nerve-racking ... these things are only as good as the people doing them and the level of engagement.”

The lack of knowledge around how appraisers and appraisees use and interpret supporting information during the appraisal, and how decisions or outcomes are negotiated was seen as problematic:

“When the appraiser receives the information in the six elements as I call it, is that adequate in terms of assisting them in coming to a decision whether to say ‘yeah okay’ or ‘no sorry there’s a problem.’ ... Is the quality and the quantity of information being given to the appraiser fit for purpose to help them make their decision. [At] the other end of the scale ... how easy is it for doctors to collect that supporting information?”

The value of digitally recording appraisals to capture the details of the interaction, either through the use of video or audio recording was heavily supported by participants. The likely resistance to the recording of appraisal was also recognised:

“It is challenging ... but it is also evidencing that someone is up to speed and fit to practice. If we want to do that then we have to get into the appraisal room and we have to find a way of looking at the quality of the appraisers and the types of questions they’re asking and the follow-up questions that they’re asking in order to make that appraisal an effective and meaningful appraisal. I can’t see how we do it any other way ... but I don’t think we should let that put us off.”

“If you were to video a number of appraisal meetings ... over a number of years, I mean if this project went on for three years ... you got three appraisals [with the] same doctor, I think that’s going to be absolutely fascinating because that will show really whether there have been significant changes in the way that doctors practise.”

The need to unpack and unpick what goes on in the appraisal room and to record the interaction was a common refrain. Although the participants did not necessarily state it in this way, many implied or hinted that the ‘witnessing’ of appraisal would introduce some quality assurance - thereby safeguarding the integrity of the process and, by default, those involved.

From a research and evaluation standpoint, the recording of appraisals would be significant as it effectively turns the appraisal event into a digitally-fixed, ‘for-the-record’ document. As with supporting documentation this then makes it accessible for analysis and evaluation.

Many of the participants felt that access to appraisal might discourage or uncover subversive and collusive practices such as the *“nod and wink appraisal”* or ‘tick-boxing’:

“What evidence people are bringing?,[is the evidence real?], does it mean anything to our patient? Is there a real outcome measure or is it just a ticking box and talking shop?”

“My last appraisal had no links with clinical governance information ... it’s possible to get a good ORSA assessment without that necessarily reflecting the reality on the ground.”

Similarly, a number of participants were concerned that appraisers who were not of the same clinical specialism as the appraisee might, through a lack of content knowledge, be less well placed to reach an informed decision:

“I don’t see how a non-anaesthetist has the knowledge and skills to appraise an anaesthetist ... it is too easy for the appraisee to pull the wool over an appraiser’s eyes if the appraiser doesn’t come from that speciality.”

There also appeared to be a general concern over the consistency of appraisers and others were concerned that the appraisal process was not standardised enough:

“Some Trusts [are] operating on paper-based systems ... I visited a client recently who sort of showed me the cabinet in the back room where all the previous appraisals and their appraisal history is a massive cabinet in a back room ... it’s stuffed to the gills.”

The contribution of the participants indicates awareness, on their part, that the appraisal process can be undermined at various stages. As ethnographic studies in other institutions and organisations demonstrate, informal occupational practices *often* develop to shortcut or circumvent formal rules and procedures^{123, 124}. Very often, as one of the participants noted earlier, this arises through pragmatism – to make things work - rather through any criminal or Machiavellian intent – although that too, is a factor that may well be insufficiently recognised.

In turning to RO judgement-making, many of the concerns that are evident in the participants’ thinking and reasoning on appraisal are carried over: notably, the similar desire to ‘witness’ the judgement-making process.

3.2.4 RO Judgement-making

The need to evaluate the judgements that ROs make was a theme that most participants felt was important. The participant cohort contained a number of ROs and it was useful to examine what they, as ‘insiders’, had to say – especially in relation to judgement-making. Here an RO describes the nature of the role:

“I’m speaking for myself, certainly each decision at the moment is something I actually have to sit down and concentrate on.”

Another non RO observed:

“I’m not sure exactly how the RO’s decision is made.”

Participant stakeholders had very little idea about how ROs actually reach the judgements they make:

“I imagine it is a mechanism that takes you through the process to form a judgement on the fitness to practise.”

Implicit in this reply is the assumption that the revalidation process steers ROs through a dialectical process but it is not necessarily clear how it is done. One participant was concerned about his own revalidation experience:

“The Medical Director has no clue who I am and I’ve never done an appraisal there so they were at a loss whom to ask when I sent them the form ... [afterwards] I was going through it and I didn’t know the name of the person who signed it but he signed it off saying I was a person of good standing. So it’s a completely disjointed process.”

A common view was that ROs, especially those in large Trusts or organisations would be almost entirely reliant upon the decisions made by appraisers:

“I see the ROs as essentially having a rubber stamp job and it’s the appraisers that have got the difficult job. The appraiser is the one that does the work and the RO just basically rubber stamps it.”

“I don’t think ROs typically would look at supporting information.”

Those ROs working in small organisations were felt to have an advantage in that they are likely to have first-hand knowledge of the doctors awaiting revalidation:

“The appraisal process only formally documents what I largely know intuitively or actually through day to day contact.”

Some ROs did offer general information about how they set about judging a doctor’s fitness to practise and the craft skill involved:

“Which rules are you following? ... Which documentation are you looking at?... Have you got the right support to make the decision? I guess there’s potentially something about following it up after the decisions as well in terms of whether the decision, I don’t know whether anybody’s going to turn round and audit decisions whether I’ll get a phone call one day and somebody will say ... ‘okay get out all your evidence as to why you made this decision for this person on this day.’ Clearly there will be an audit trail for those that subsequently fail.”

It was clear however, that several participants felt the quality of ROs was variable. The unspoken implication was that this would manifest itself in the judgement-making:

“The role of the RO is actually something I’m trying to understand actually to be honest ... there are different levels of ROs.”

To evaluate the craft of ‘judgement-making’ will, as with appraisal, require some form of ‘witnessing’. In addition, to be thorough in capturing the thinking and reasoning processes involved – to witness the craft of the RO – any digital witnessing must be accompanied by access to the completed ‘recommendation statements’ or ‘form 4’s’. The experience of conversation analysts and discursive psychologists is that in the process of ‘writing up’ decisions, arguments and logical reasoning undergo a varying degree of transformation.¹²⁵. Unravelling and understanding data of this sort is the prerogative of qualitative research – a fact that a large number of the participants recognised:

“I think that qualitative research would also give you real insights into some of these questions, the behaviours, reflexivity and so on they can all be explored qualitatively.”

“I think the only way to get specific information is through most of the qualitative work.”

“You need to analyse and compare, in my view, RO’s decisions ... and if the decision is also to refer to remediation then we [need] to know why they’ve been referred to remediation, was it related to their clinical practice?”

Whilst the participants highlighted the importance of analysing the qualitative aspects of the RO’s craft, they also volunteered ways in which evaluators might maximise their resources by suggesting how one might sample:

“I think looking at the edges will be very worthwhile.”

“Is there a higher proportion of deferrals amongst people of particular characteristics ... that would indicate groups that ROs are struggling to deal with so they’re always deferring them for a little bit.”

“What will happen in reality is that the majority of appraisers are going to say that the majority of the doctors are okay so those ones are rubber stamped. So maybe you could look at ones where the initial response from the appraiser was ‘we’ve got a problem here’ and then to look to see how many of those, by the time they’ve been through the hands of the RO, the RO has passed. Because you don’t know whether the appraisers are going to err on the side of leniency or err on the side of being too tough and similarly you don’t know what ROs are going to do. Another way you can look at it is, how many, when the appraiser says it’s okay, does the RO say ‘hang on a minute, you say it’s okay I want another look at that.’”

What participants appear to be advocating, both explicitly and implicitly, is that to concentrate attention on ‘outliers’ might be a useful *initial starting point* for any qualitative analysis of RO judgement-making. ‘Deviant case’ analysis, as it might loosely be termed, looks to explain ‘variation’. Many of the participants suggested auditing Fitness-to-Practise (FtP) cases in an investigative and qualitative way:

3.2.5 Influence of different practice settings & roles

The issue of outliers was also pertinent to the discussions on whether revalidation would adversely affect some groups of doctors or those who work in diverse practice settings. While most of the participants had a view, opinions were divided:

“So there’s this kind of errant population of doctors out there who are probably the ones that represent the greatest risk to the public because they skip between employers or they’re in and out of the country or ... they jump from locum agency to locum agency [and] still don’t sit within a governed system ... it’s not going to capture them terribly well ... even in the locum sector which is obviously where we can speak the most confidently about we’ve got doctors that relate to us or just work for us periodically and they move between agencies, it is desperately hard for those doctors to continue practising without an appraisal.”

“Locums I think are a special group, especially those that aren’t in agencies so people who do intermittent work, directly employed, probably for Trusts largely that know them but ... don’t have a single overriding employer [I] think ... there are all kinds of difficulties in terms of the practicalities for them.”

The ‘invisibility’ of locum doctors within the revalidation process by virtue of their occupational setting may, ironically, result in a greater visibility within fitness-to-practise referral lists as ROs insist upon information that some locum doctors are not able to provide:

“They [ROs] don’t want to put their neck on the line. I guess we’re all being a bit more careful now, I know we are. I guess it’s affecting everyone.”

People who take leave for a variety of reasons i.e. maternity, sabbatical were also highlighted as groups that may have potential difficulties:

“you mentioned maternity, can we just add, because maternity leave seems to be singled out. There are issues like sabbatical, there are issues like the Registrar of the Royal College, [who] could be spending 80% of their time being a Registrar and 20% of their time doing their clinical work. I think we shouldn’t bundle maternity as a separate issue, this is about putting that sort of group together to say how are they going to be managed”

One participant suggested that an unintended consequence of revalidation may be to ostracise particular groups of doctors; although he was not specific about how that might occur:

“Far and away the very biggest concern is that we may have inadvertently introduced a process that favours or disfavors particular groups, based not any form of medical objective criteria, but on something else, and there continues to be concern that black and minority ethnic doctors are heavily over-represented in GMC fitness-to-practise procedures.”

Others were equally sure that culture and ethnic origin would not or should not be a factor in revalidation; however, there was a realisation that implementing revalidation might be difficult in certain sectors of the profession:

“People make a great song and dance about minority groups ... it doesn’t matter where we come from, we’re all doctors and we have to practise to the same standards. It’s very challenging, the most challenging of all [are] the locum doctors ... often young people are flitting around doing odd nights here and there and getting supporting information for those people and making meaningful decisions when often they’re not known by anybody, that’s very challenging.”

“There does certainly have to be consideration about differences within ... how the system is implemented, and rather than a blanket one-fits-all approach ... I would

guess that you would need to cover, in terms of the evaluation study, a cross section of doctors and not just in terms of their personal characteristics but also the settings in which they work.”

Notwithstanding that most of the participants offered a view on whether some groups of doctors might be disadvantaged by revalidation, and how ‘challenging’ it was to ensure no discrimination occurred; very few had any concrete suggestions as to how one might evaluate the revalidation in relation to these “*itinerant groups*” or “*orphan doctors*”. A similar diversity of opinion and lack of ideas was evident in terms of Patient and Public Involvement (PPI).

3.2.6 Patient & public involvement (PPI)

Almost all the participants recognised that the nature of healthcare and the asymmetry of the doctor-patient relationship have changed. As one put it:

“I think it’s one clear direction of travel that we’re seeing, is that patients want to be involved in their own healthcare, so ‘nothing about me without me.’”

Not everyone felt this was a good thing. Not surprisingly perhaps, the views expressed were factionalised and tended to reflect the stakeholder’s perspective: those with a strong stake in patient groups were keen to see greater involvement; while doctors’ opinions varied. One experienced doctor was forthright in his view of patient and public involvement:

“Well, I think the vulnerability of the whole process at the moment is the involvement of patients and patient feedback and again I accept there are token attempts to do patient feedback in terms of [requiring] twenty-five patients with feedback, compliments and comments etc. etc. But I think there is a pretty imprecise element, certainly for patients that I deal with in the role that I have.”

Another interviewee observed that patient public involvement was currently quite limited:

“Patient feedback is a very minority element of the overall revalidation process ... it’s a bit tokenistic at the moment so what contribution does it make in reality to making judgements about a doctor’s fitness to practise?”

He then proposes ways in which PPI might be increased:

“RO decisions do need to be evaluated and it’s important that in evaluating them that part of that evaluation exercise contains a patient or patient representative because we do, as patients, do come across on many occasions with different, very different viewpoints. So ... any sort of framework you develop has to have in it, in my opinion ... a framework where a patient or patient representative can sit down alongside clinicians, the ROs or whoever it might be doing the analysis and say this is my view of it. I mean one of the criticisms several organisations have had is that at

the moment the system is doctors evaluating doctors, doctors appraising doctors and I think from a patient's point of view we need a system which includes a non-medical patient/public viewpoint."

For one doctor, patient and public input in the form of patient feedback was something that needed to be part of the "evolution of the process", but "at this moment in time it seems a bureaucratic encumbrance that doesn't add much value."

One of the few practical suggestions for evaluating revalidation from the PPI perspective involves obtaining data from the National Joint Registry:

"You can use the patient satisfaction data because they use the Oxford, Oxford New Score I think it's called ... anyway they use a satisfaction tool and you can check how many revisions, can check if there's problems with any particular joint replacements."

Another doctor felt patient input would be beneficial but was not helped by current arrangements:

"I think if patients are allowed to contribute to the doctor's development ... through feedback and I think one of the things that is not so good about the whole appraisal revalidation process at the moment is this limited sort of five year review."

A different participant was "gloomy" about whether revalidation would encourage patient input:

"It would be very nice if the input of patient groups ... was greater in informing what revalidation does. The particular group that I come from we're very conscious that in one of the main clinics in the country maybe in the world that deals with gender identity problems there are some very, quite profound problems amongst the staff in terms of their handling of the patients who go to us for advice and support so I'm one [who] would like to see ways of unearthing that kind of inappropriate medical behaviour."

Latent tension amongst the various factions involved in revalidation was evident in this participant's suggestion about how to increase PPI:

"One question that I raised at the meetings ... at a meeting earlier was that: 'Why can't you put on the GMC website, on the online register, if a doctor's been revalidated?' And the answer was 'well that's a bit difficult' mainly because if you go onto the website because if you go onto the revalidation website ... you click onto public and patient information and there is a piece there, your doctor is registered with us so forth, you want to see these files to on to the online register, you click on the online register and it says nothing about the doctor going through revalidation ... So I've suggested that perhaps we should be doing that now ... and we'll get to a position where if they're not revalidated they won't go on the register ... there's a lot

of talk about revalidation being for patients but I think ... more work needs to be done in demonstrating to patients, the public, it is for them."

The underlying tension and uncertainty amongst revalidation's assorted stakeholders as to how far PPI should feature in revalidation appeared to inhibit the participants' suggestions on how PPI could be included in the evaluation.

3.2.7 Professionalism

A similar uncertainty over the issue of 'professionalism' – what it is, and how one might measure such a contested concept - also hindered the participants' responses on that topic:

"The thing about Good Medical Practice as a sort of document is [it] is mostly common sense and so I'm not sure whether revalidation will raise the profile of it or not? ... Most of it is kind of really obvious stuff."

"When you talk about professionalism it's one of the things that I think people are still struggling with. I think people can say when someone is being unprofessional, not that it's a lightly done thing."

"Okay let's put it another way round ... Good Medical Practice is virtually unknown amongst patients. It's never distributed nobody ever sees I mean who's ever heard of it if you're not working in that sort of confines of the system? Whilst I think revalidation deals with specifically with issues raised during the appraisal I'm not sure that it deals with a more general requirement to act in a way which is consistent with the aspirations of Good Medical Practice. And if you look at Good Medical Practice and say 'How can I require a doctor to comply with anything here?' I don't think you can. So if you read through it's full of very positive stuff and if you then look through the booklet and say 'hang on ... I don't think my doctor's doing X so what do I do about it' and there's nothing you can do really because you're not going to complain to the GMC that the doctor hasn't done paragraph seventy three."

Those who did volunteer ideas recognised that some form of qualitative analysis would be necessary:

"You need to get underneath what happens behind the closed door of the appraisal. This is very definitely about whether the discussion properly covers ... the aspects of Good Medical Practice evaluation. I mean there will be documentation there will be appraisal documentation of some sort to establish whether those around Good Medical Practice happened or not. The only problem is this is highly confidential, I mean would you like another company ferreting around in your appraisal documentation?"

"I think also by qualitative means one will get an idea of how people feel about the process and whether they feel the reflection, whether they feel the aggregate of all

the supporting information has been positive for them, how they feel they've been changed by the process. So I think for me it's mainly a qualitative research issue."

"I guess one of the richest sources of how to evaluate doctor's professionalism may be patient feedback."

For many of those interviewed, the difficulty in evaluating whether or not *Good Medical Practice* was shaping practice was the uncertainty that, even if it was, would that equate to professionalism?:

"I think Good Medical Practice is a set of standards which sets out what you want to be if you're going to be a good doctor it's not necessarily saying there's professionalism. Professionalism I think is essentially about self-governance as an individual or collective community ... I'm trying to figure out how revalidation connects to professionalism."

The point was also made, on more than one occasion, that many doctors might not read *Good Medical Practice*:

"There would be in some cases a sea of blank faces when we asked: Who's read Good Medical Practice or who's read it recently – people would look at us a little bit blankly ... I can't say whether it's everyone's bedtime reading really."

Knowledge tests and surveys were proposed; but all recognised that doctors could revise *Good Medical Practice* - and that this would still not evidence 'professionalism' or that they routinely use it to shape their practice. Direct observation would be required for that. But as has already been mentioned, the perceived difficulty with direct observation or 'witnessing' revalidation activity, and indeed potentially with the implementation of revalidation itself, is that important conversations, such as in appraisal, might "*choke off*" what were once perceived to be "*safe*" conversations. Several participants saw this as an unintended consequence.

3.2.8 Unintended consequences

Very few of the interviewees volunteered much in the way of unintended consequences. Whilst fear of regulatory consequences was held to be a potential inhibitor to open discussion for some, another concern was that the obligations revalidation places upon employers and ROs will have human resource and financial implications. One doctor recounted how, in dealing with an issue of underperformance, his employer:

"Ended up involving a whole lot of people when we had to put the processes in place to try and support this person they weren't easy, so it was a question of coaching or mentoring and then the Trust didn't have any funds ... we ended up feeling that we do not support our senior doctors so well so that we were good at identifying there

was a problem but once we identified the problem putting the solution wasn't very easy."

The same participant also wondered if the evaluation of revalidation were to involve the use of crude output measures, whether this might encourage some practitioners to become risk averse with 'knock-on' consequences:

"I mean when I was speaking to the senior surgeon the feeling was 'well if this vascular surgeon has identified that he's got a higher mortality rate', well many surgeons have to operate on say abdominal aortic aneurisms and are not assured of an ITU bed, well the simple thing would be 'well I shall not operate unless I am assured of an ITU bed' well this means that waiting lists will go up."

Several others thought that the requirements of revalidation would disadvantage those doctors who work in places or situations where there is little governance, to the point where a shortage of licensed doctors in these situations might occur:

"You've got this errant and strange population of doctors who are overseas, who are on cruise ships, who are retired or approaching retirement, who are full-time Locums who have never really been or haven't been in a governed system for ages and then can easily not be part of one because they are self-employed."

The point being made here is that the revalidation process *"favours or disfavors particular groups based not on any form of medical objective criteria but on something else."* The difficulty for all the participants is that the scale and aim of revalidation is such that after less than a year, it remains difficult to determine how it will evolve.

Another participant spoke of a developing *"appraisal industry"*, whereby appraisees learn how to negotiate the process to the point where they are not *"actually tested or stretched in any way."* He suggested that appraisal and revalidation had become *"rather self-contained ... without too much reference to the outside world."*

Others felt that an unintended consequence of revalidation might be to:

"Choke off the kind of open conversation that there was before because of the threat that it might lead to a deferral or might lead to a fitness-to-practise, so I think we've got to be careful there that we are clear about how an individual doctor can talk about areas that they're concerned about without being concerned that they will be struck off the register."

"I think it [appraisal] was a fairly safe, we were moving hard to make it a very safe process, very clear about if issues were shared that needed to be taken on they would be ... now I'm worried that people won't feel safe to share that."

3.2.9 Timeline of evaluation

Overall participants felt that the evaluation needs to run for a minimum of 3-5 years until the end of the current revalidation cycle. This gives the process time to become established in practice:

“Absolutely minimum three years, realistically five. The reason I say that is the majority of doctors are going to go through this 2013 to 2016 schedule.”

“Should probably run over two or three years I would have thought.”

“I think it has to be for a minimum of five to ten years,[as] the cycle for revalidation is five [years]. But I think a proper evaluation should take place building on evidence that we get each year, five to ten years in the life cycle of an expensive doctor is probably not far out.”

A more cynical view was that:

I don't think it'll be valid until it's been going for at least three years and probably five years and I think this is an important point as the GMC has definitely tried to make it easy for itself ... the easier doctors to revalidate are being revalidated at the beginning ... the NHS GPs, now they're quite easy to revalidate because they've had appraisal systems going for a while and they're very very managed by the primary care trust ... it's the more difficult ones that are being revalidated later so if you did a full assessment now or in the next twenty four months you're going to get a skewed result because everything would look better than it really was.”

All the participants were convinced of the importance of getting any evaluation right; that it should be inclusive; and that it should be very carefully devised. The gravity of the enterprise well captured by the interviewee who observed:

“You've got a great opportunity for an actual experiment and you ought to think about how you use that opportunity.”

3.2.10 Measures – qualitative or quantitative

An enduring thread throughout was that a large number of the participants were convinced that qualitative methodologies were essential if the interactional and interpretative aspects of the revalidation process were to be properly evaluated:

“Is the quality and the quantity of the information ... being given to the appraiser fit for purpose to help them make their decision?”

“You look at the standard of portfolios that are being put in.”

A small number of participants were clear in suggesting 'quality metrics' as a way to evaluate revalidation as in studying referral or deferral rates for outliers or significant increases or decreases. Others felt this kind of measure was "crude". Whilst metrics have undoubted value, in isolation they provide no information about the prior decision-making involved. As one participant observed:

"Um, one person's borderline is another person's good enough"

Overall, the feeling amongst these interviewees was that some quantitative methods could still be useful and that a combination of qualitative and quantitative methodologies would provide the best opportunity for a thorough evaluation:

"To have the qualitative information is obviously extremely beneficial in it gives a rich a richness to hopefully identify the benefits of revalidation but just the ... quantitative elements establishes the baseline for obviously what is a new system."

"I've been involved quite heavily in survey development in medicine ... I'm an advocate to some of it but I'm one of its biggest critics."

"If you can get hold of any of those resources to see whether doctors are actually documenting their reflection: I think there's a difference between reflection and documenting their reflection. Where doctors do reflect ... they don't actually document that reflection which you're asking for."

The quality of any evaluation therefore appears to rest on the mix of methods used, and the skill of those evaluating in choosing what areas of the revalidation process define it and will give the greatest knowledge return.

3.2.11 Summary of thematic analysis of stakeholder interviews & focus groups

In summary, the stakeholder interviews identified five areas to focus on in the evaluation of revalidation.

The first was around supporting information: particularly the quality of supporting information, whether doctors can collect each type, which type of information is impacting on practice, and the levels of reflection on supporting information. Some felt that examining the paperwork required for appraisal, i.e. the appraisal forms, would be an important way to explore some of these issues.

The second area to focus on was appraisal. Stakeholders were very keen to get inside the 'closed' door of appraisal and find out what happens in an appraisal by recording the meeting. They also felt it was important to examine appraisal rates.

Thirdly, stakeholders felt it important to evaluate RO judgment-making, and particularly *how* ROs actually make their decisions. Again, the documentary evidence ROs use to make

their decisions e.g. form 4s, could be examined. Another evaluation measure proposed was looking at deviant cases or 'outliers' i.e. doctors referred to FtP for example.

Fourthly, the underlying tension and uncertainty amongst revalidation's assorted stakeholders as to how far PPI should feature in revalidation highlighted the importance of exploring this issue.

Finally, there were some sampling considerations identified by stakeholders. The majority expressed a view on whether some groups of doctors might be disadvantaged by revalidation, and how "challenging" it might be to ensure inclusivity. Very few however offered concrete suggestions as to how this might be undertaken. Stakeholders felt that the evaluation needed to be longitudinal running for 3-5 years, concluding in time to inform the next revalidation cycle.

The findings from the literature reviews and thematic analysis of interviews and focus groups were used to develop and inform the evaluation framework that follows.

4. The Revalidation Evaluation Framework

4.1 Introduction

In this section of the report the proposed Revalidation Evaluation Framework is described in detail. The development of the framework was informed, as described above, by:

- a. the CAMERA team working closely with the GMC to conceptualise revalidation from a regulatory perspective
- b. literature reviews of:
 - (i) the impact of medical regulatory interventions on doctors' performance in practice e.g. MOC
 - (ii) the effect of appraisal on doctors' performance in practice
- c. thematic analysis of 24 in-depth interviews and 2 focus groups with key stakeholders in the revalidation process
- d. the CAMERA team's research expertise in regulatory assessment including the use of Cultural Historical Activity Theory (CHAT)

Thus the proposed Revalidation Evaluation Framework is a result of the triangulation of the core findings from the literature reviews and stakeholder interviews with the standards underpinning revalidation, as outlined by the GMC. Central to this process was CHAT as the underlying conceptual framework. CHAT allows for an understanding of complex activities within systems, appreciating that systems impact on activity, in this case revalidation activity in the complex provision of healthcare, and vice versa. Within this framework a series of important evaluation questions were established that could be explored by four overarching work streams.

Work Stream 1 – *supporting information* aims to capture the nature, quantities and uses of supporting information and proposes a literature review, a survey and semi-structured interviews with appraisees and appraisers, audio-recording appraisal meetings and a documentary analysis of appraisal forms. **These approaches will capture how GMC guidance is being applied, what information doctors actually collect, what information they value, in what ways they reflect and change their practice in light of this information and how this is evolving over time.**

Work Stream 2 – *appraisal* aims to explore the activity, quality and impact of appraisal through evaluative measures including a literature review, secondary data analysis of appraisal rates, a survey and semi-structured interviews with ROs, as well as a survey of and semi-structured interviews with appraisees and appraisers. **These approaches will capture what appraisal activity is happening, when and how GMC guidance (e.g. GMP framework for appraisal and revalidation) is being used, and how revalidation is shaping**

professionalism through the adoption of the values and principles of *Good Medical Practice*.

Work Stream 3 – *RO judgement-making* attempts to understand the craft of ROs as they make their judgements through a survey and semi-structured interviews with ROs and Employer Liaison Advisers and other relevant personnel, and a case control study involving doctors who have been referred to the FTP procedures. **These approaches will establish how revalidation supports the early identification of concerning practice, and how ROs make their decisions in relation to protocol, recommendation statements and processes.**

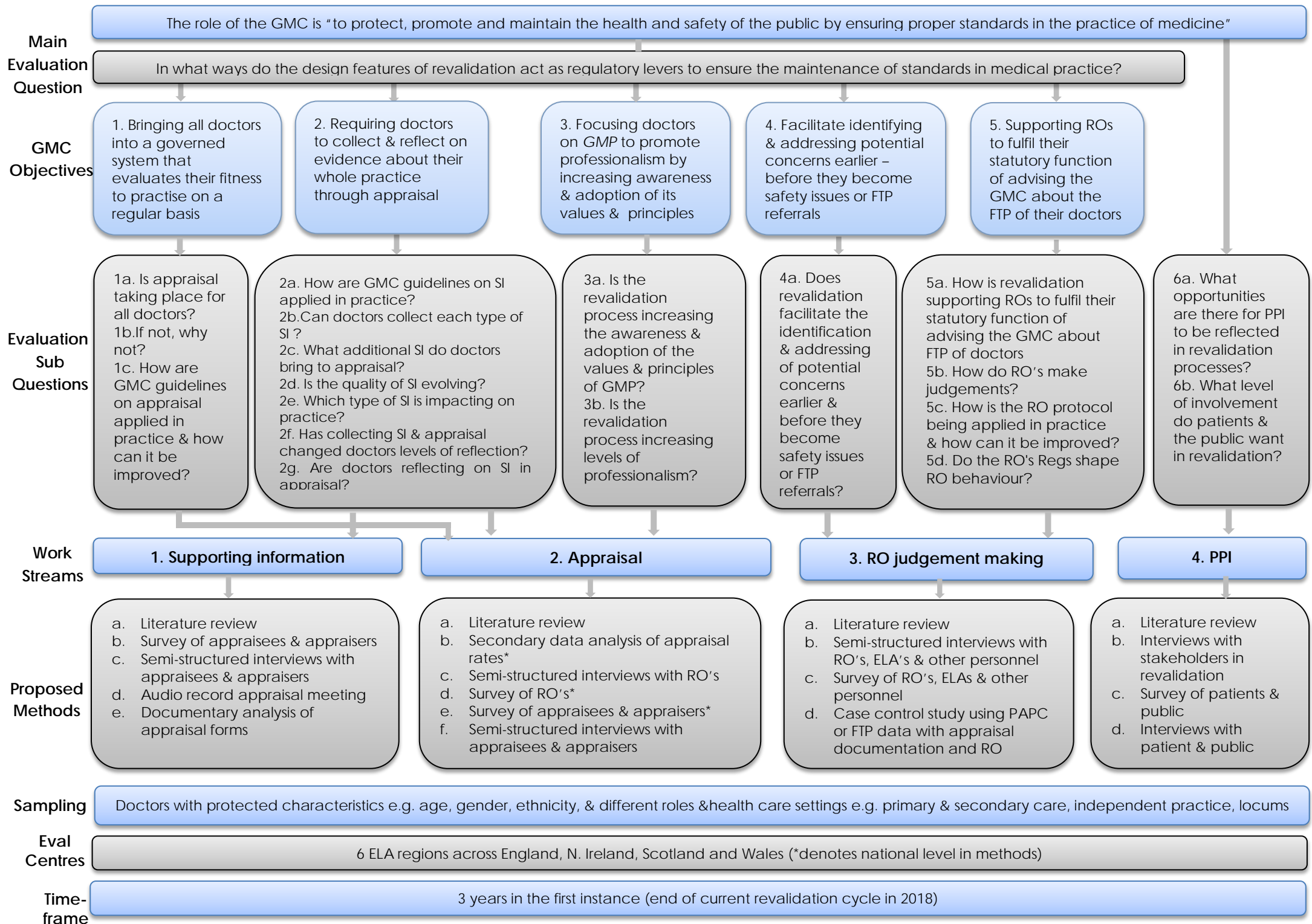
Work Stream 4 – *Patient and public involvement (PPI)* explores the role of patient and public involvement in revalidation and addresses the overall role of the GMC which is to “protect, promote and maintain the health and safety of the public”.²⁷ The proposed methods include a literature review, interviews with key stakeholders in revalidation, and a survey and in-depth semi-structured interviews with patients and the public. **These approaches will explore the opportunities for PPI to be reflected in revalidation processes and the level of involvement patients and the public want in revalidation.**

By developing a framework centred on these core areas of revalidation ‘activity’, the evaluation programme will generate a range of fundamental insights from different perspectives. In addition, a mixed-methods approach is suggested for each work stream in order to effectively address all of the proposed evaluation questions. A mixed-methods design would add a further layer of robustness to the framework as it would enable the triangulation of findings, particularly the different perspectives.

In order to explore issues of equity, diversity and inclusivity across practice settings and to assure transferable findings, it is proposed that the work streams could be implemented in for example six revalidation evaluation centres across the four nations (England, Northern Ireland, Scotland and Wales). Importantly, the sampling strategy should aim to capture the diversity of protected characteristics in the doctor population such as gender, age, ethnicity etc. as well as doctors in a variety of settings and roles, such as locum doctors, as well as doctors in primary and secondary care. The sampling should ideally be designed to explore if all doctors, from any background, in any role and any setting are experiencing revalidation similarly. A longitudinal design could be adopted such as a three year period in which different aspects of the framework could be developed. The proposed Revalidation Evaluation Framework is summarised in Figure 4 (following page).

The methods that were investigated by the CAMERA team for the framework, but were then excluded, are presented in Appendix 9. This includes recourse to the reasons as to why they were excluded from the framework.

Figure 4: Flow diagram of revalidation evaluation framework



4.2 Evaluation Questions

The GMC's regulatory function is to "protect, promote and maintain the health and safety of the public by ensuring proper standards in the practice of medicine".²⁷ It is anticipated that the design features of revalidation will add value to existing appraisal and clinical governance processes, in order to achieve this end. The proposed Revalidation Evaluation Framework is underpinned by an overarching evaluation question that addresses this function:

In what ways do the design features of revalidation act as regulatory levers to ensure the maintenance of standards in medical practice?

The main evaluation question encompasses the five areas based on the five objectives (or proper standards) the GMC designed revalidation to drive. Within these five areas are a series of sub-evaluation questions:

- 1. Bringing all doctors into a governed system that evaluates their fitness to practise on a regular basis**
 - a. Is appraisal taking place for all licensed doctors every year?
 - b. If appraisal is not taking place for all licensed doctors what are the reasons?
 - c. How are the GMC guidelines on appraisal (GMP framework for appraisal and revalidation) being applied in practice and how might they be enhanced to increase its impact?

- 2. Requiring doctors to collect and reflect on supporting information about their whole practice through appraisal (continuing professional development, quality improvement activity, significant events, feedback from colleagues, feedback from patients, review of complaints and compliments)**
 - a. How are the GMC guidelines on supporting information¹² being applied in practice and how might they be enhanced to increase its impact?
 - b. Can doctors collect each type of the required supporting information in their practice?
 - c. What additional supporting information do doctors bring to their appraisal?
 - d. To what extent is the quality of supporting information changing/evolving?
 - e. Which type(s) of supporting information has had the most impact on doctors' practice?
 - f. Has the process of collecting the supporting information and the appraisal process increased doctors' levels of reflection?
 - g. Are doctors reflecting on or supported by the supporting information in appraisal?

- 3. Focusing doctors on *Good Medical Practice* to promote professionalism by increasing awareness and adoption of its values and principles**
 - a. Is the revalidation process increasing the awareness and adoption of the values and principles of *Good Medical Practice*?
 - b. Is the revalidation process increasing professionalism?

- 4. Facilitate identifying and addressing potential concerns earlier – and before they become safety issues or fitness to practise referrals**
 - a. Does revalidation facilitate the identification and addressing of potential concerns earlier and before they become safety issues or fitness to practise referrals?

- 5. Supporting ROs to fulfil their statutory function of advising the GMC about the fitness to practise of their doctors.**
 - a. How is revalidation supporting ROs to fulfil their statutory function of advising the GMC about the fitness to practise of their doctors?
 - b. How do ROs make their judgements?
 - c. How is the RO protocol¹¹ being applied in practice and how might it be improved?
 - d. How do RO recommendation statements shape RO behaviour?

While exploring these five areas that the GMC designed revalidation to drive, there is a further important area that needs to be addressed. The overall aim of the GMC is to “protect, promote and maintain the health and safety of the public....”²⁷. Patients are already directly involved in the revalidation process in the provision of feedback to doctors about their medical practice, which doctors use as supporting information in their appraisal. However, arising from the GMC’s overall aim and the interview data, it would appear important that a wider evidence-base be established by investigating the overall role of patient and public involvement in revalidation. In particular to explore what the opportunities are for PPI in revalidation and what levels of involvement patients actually want in the revalidation process. Do they for example want to be directly involved in the process or does the public simply want to be assured that systems are in place for their protection? The following evaluation questions addressing these issues were therefore included in the framework:

- 6a. What opportunities are there for the patient perspective to be reflected in revalidation processes?

6b. What level of involvement do patients and the public want in the revalidation process?

These initial questions form the basis of the evaluation. They draw upon both process and outcome evaluation models. As with all research, it is likely that further questions will arise once the evaluative programme is initiated, drawing on early findings. It is therefore not a definitive or exhaustive list of questions but a comprehensive initial framework of evaluation questions for consideration by the GMC.

4.3 Work Streams

To address the overarching evaluation question and each of the individual sub-questions, we propose that a major evaluative programme be organised into four work streams based broadly on the three main areas of revalidation activity as well as the PPI dimension:

Work Stream 1 - Supporting information (addresses GMC objective 2 and GMC guidelines on supporting information requirements)

Work Stream 2 - Appraisal (addresses GMC objective 1, 2 & 3 and GMC guidelines on GMP framework for appraisal and revalidation)

Work Stream 3 - RO judgement-making (addresses GMC objective 4 & 5 and GMC guidelines on RO protocol)

Work Stream 4 – Patient and public involvement (addresses the overall role of GMC)

By developing a framework centred on these core areas of revalidation ‘activity’, the evaluation programme will generate a range of fundamental insights. This will assist the GMC in making strategic decisions about future design features for revalidation. A proposed approach to the commissioning and application of this framework is discussed in detail in the subsequent section of this report. In short, CHAT provides a conceptual framework that acknowledges the complexity of revalidation as an activity in a complex system, and highlights the need for a combination of evaluative approaches rather than a singular measure or a series of unrelated measures.

Each of the work streams are described in detail in the following sections.

4.3.1 Work Stream 1 - Supporting information (GMC Objective 2)

Work Stream 1 focuses broadly on evaluating the supporting information in the revalidation process and addresses evaluation questions (EQs) 2a-2f:

- 2a How are the GMC guidelines on supporting information¹² being applied in practice and how might they be enhanced to increase its impact?
- 2b Can doctors collect each type of the required supporting information in their practice?
- 2c What additional supporting information do doctors bring to their appraisal?
- 2d To what extent is the quality of supporting information evolving?
- 2e Which type(s) of supporting information has had the most impact on doctors' practice?
- 2f Has the process of collecting the supporting information and the appraisal changed doctor's levels of reflection?
- 2g Are doctors reflecting on or supported by the supporting information in their appraisal?

To answer these evaluation questions we suggest the following:

- (a) a review of literature
- (b) survey of appraisees and appraisers
- (c) conducting semi-structured interviews with appraisees and appraisers
- (d) audio-recording appraisal meetings
- (e) undertaking of a documentary analysis of appraisal forms
- (f) a mixed-methods approach
- (g) a sampling strategy

Each of these are outlined in detail in the following sections.

(a) Review of literature on supporting information

Rationale: The most rigorous research is underpinned by reviewing the literature. Synthesising the existing literature around each type of supporting information would inform decisions about the development of revalidation in relation to the implementation of the supporting information.

Method: Six literature reviews (systematic if appropriate) investigating the extent to which each type of supporting information (1. continuing professional development, 2. quality improvement activity, 3. significant events, 4. feedback from colleagues, 5. feedback from patients, 6. review of complaints and compliments) affect the

performance of doctors could be developed. To provide an indication of the inclusion criteria that would be utilised in the review continuing professional development is used as an example. The inclusion criteria for this review could be; empirical studies written in English, published from 1960 to present, investigating the effect of continuing professional development on the performance of doctors in any specialty using any method and any outcome measure. The search strategy would involve searching a variety of relevant databases e.g. Medline, EMBASE, grey literature, hand-searching key journals, searching trials registers, contacting experts in the field and citation searching. Data would be extracted and quality assessed. The data would be synthesised using qualitative, quantitative or mixed methods syntheses. In some cases there may be a need for a 'tertiary review' which is a systematic review of systematic reviews.¹²⁶ Tertiary reviews use the same methodology as a standard systematic literature review. It is potentially less resource intensive than conducting a new systematic review of primary studies but is dependent on sufficient systematic reviews of a high quality being available. This would probably be the case for the quality improvement activity review but the scope for a tertiary review would be decided when conducting the scoping searches to inform the systematic review design protocol for each review. The reviews should be conducted using established guidance for conducting systematic reviews e.g. the CRD¹²⁷, and for reporting systematic reviews e.g. PRISMA.¹²⁸

Methodological strengths and limitations: Literature reviews are vital pieces of research for organising and making accessible the major findings in an area of inquiry, and are often used to inform policy and future research directions. However the quality of a review (particularly a systematic review) is dependent on the quality of existing studies. Furthermore, while a literature review can answer some research questions they will not be able to answer all research questions and in this case further empirical research would be necessary. Literature reviews can also be quite time and resource-intensive and can go out of date quite quickly.

(b) Survey of appraisees & appraisers (EQ 2a-2e)

Rationale: A survey of appraisees and appraisers could also be utilised to address the evaluation questions a-e in this work stream. This would provide standardised data that could be compared amongst groups.

Method: In-depth questionnaire survey of appraisees and appraisers. A questionnaire could be designed to capture quantitative and qualitative data in a standardised format addressing the above research questions. Data would be analysed using descriptive statistics and a thematic analysis of free text.

Limitations: Surveys are generally less resource intensive and therefore cost less to implement. They can incorporate a broader/larger sample than, for example, conducting interviews or undertaking ethnographic studies, but they are generally more limited in their scope and do not facilitate the in-depth understanding of the nuances often required.

(c) In-depth semi-structured interviews with appraisees & appraisers (EQ2a-2e)

Rationale: In-depth semi-structured interviews could be used to address evaluation questions 2a-e and would provide a nuanced insight into what is collected in comparison to GMC guidelines. In order to understand activity in relation to supporting information and what facilitates and hinders its collection and the ability to reflect on it, this would be most effectively achieved by understanding the collective themes across the study cohort but also the discourses that underpin such activity. These analyses combined will allow for a greater understanding not only of the activity but also of the reasons behind it in order to develop a future strategy for what doctors should collect, what they and other stakeholders (including patients) might best benefit from and what might be the barriers to overcome practically and in terms of policy implementation.

Method: In-depth semi-structured interviews with appraisees and appraisers. Interview data would need transcribing, coding with the support of suitable software (e.g. NVivo) and then analysed or interrogated using thematic and discourse analyses.

The interviews could be conducted by patients as researchers in order to enhance the patient and public involvement (PPI) in the evaluation framework. A recent survey of ROs found that organisations need to invest in ways to capture patient feedback and experience if the quality of patient experience, care and outcomes are to be placed at the centre of revalidation.¹²⁹ NIHR guidelines on PPI advocate how patients and the public always offer unique, invaluable insights.¹³⁰ Their advice when designing, implementing and evaluating research invariably makes studies more effective, more credible and often more cost effective.

Methodological strengths and limitations: As with all research methods conducting qualitative interviews does have its limitations. While semi structured interviews will provide an important in-depth investigation, they will not provide breadth in terms of sampling. Interview findings are likely to be transferable but not generalizable. These terms should not be confused. On the whole, qualitative studies are not generalizable in the traditional sense.¹³¹ Partial generalizations may be possible to similar populations,

but this is not the primary concern of qualitative research. In many situations, a small sample size may be more useful in examining a situation in depth from various perspectives, whereas a large sample would be inconsequential. Small qualitative studies gain a more in-depth understanding of the phenomenon and how, importantly, the studied phenomena are experienced. The results often contribute valuable knowledge and understanding that could not simply be achieved by quantitative research using questionnaires. Interviews are, of course, more time consuming and resource intensive and therefore more expensive.

(d) Direct observation – audio recorded appraisals (EQ 2e-2g)

Rationale: Observation techniques could be used to address, in particular, evaluation questions 2e, 2f and 2g. Ultimately in order to “open the closed door on appraisal” and understand whether doctors are reflecting on their supporting information, the actual appraisal meetings need to be captured and analysed.

Method: Audio-recording the appraisal meeting. The audio recorded appraisal would be transcribed verbatim and analysed using thematic and conversation analyses. In addition to these analyses, levels of reflection in the appraisal meeting could be explored by analysing audio recordings of the appraisal meeting using the Boud et al.¹³² categorisation of reflection levels. This framework cites six stages of increasing depth of reflection: attention to feelings, association, integration, relationship seeking, validation, appropriation and outcome, which could be used to analyse reflection of doctors within the appraisal meeting. Levels of reflection could therefore be scored quantitatively over time.

Methodological strengths & limitations: A potential limitation of observational techniques is that the act of being observed may alter the behaviour of participants. It is possible that appraisers and appraisees who are audio-recorded could be influenced adversely and behave differently, changing what actually happens in the appraisal meeting. However research that explores the effects of video-recording consultations between doctors and patients has found a limited impact.^{133, 134} A study of the patient perspective revealed that 70% of patients forgot the video-recorder was in the room during the consultation and <5% felt that their GP had dealt with them in an ‘unusual’ way.¹³⁵ In addition, Pringle et al¹³⁶ conducted a study to determine whether GPs' awareness of being video-recorded influences their consulting behaviour and found no difference, albeit with a very small sample.

However in order to minimise the potential impact it would be important to stress during participant recruitment that an independent body i.e. the research team, would record the appraisals and that the GMC would not have access to any recorded material.

(e) Documentary analysis of appraisal documentation (EQ 2c, 2f & 2g)

Rationale: An analysis of the appraisal documentation from a sample of doctors at each evaluation centre could be used to demonstrate the presence or absence, and level of, reflective thinking when completing appraisal documentation. In addition, this method would also provide documentary evidence on the types of supporting information that doctors are bringing into their appraisal, the quality of the information and whether this is changing over time. By repeating this analysis it would provide information on whether levels of reflective practice are evolving.

Method: A documentary analysis of appraisal documentations e.g. Medical Appraisal Guide or MARS. Again this could use the Boud et al.¹³² categorisation of the six stages of increasing depth of reflection outlined previously. The data from the documentary analysis would need to be triangulated with the analysis of the audio-recorded appraisal to gain a more rigorous and in-depth understanding of levels of reflection in the revalidation process.

Methodological strengths & limitations: While documentary analysis is not very resource or cost intensive it provides very specific data on a topic. This method is strongest when it is combined and triangulated with other data i.e. interviews, questionnaires, observation.

(f) A mixed-methods approach

While each of the individual proposed methods will answer some of the evaluation questions, none of the individual methods will effectively address all of the evaluation questions. However adopting a mixed-methods approach combining two or more of the methods would help answer all of the evaluation questions and would facilitate the triangulation of results thus strengthening the robustness of the conclusions.

For example, combining observational data with interviews would allow the triangulation of findings. These could then be used to explore how appraisers and appraisees believe supporting documentation is collected and reflected on in appraisal (in comparison with each other) and then in comparison with what actually happens, using direct observation of appraisal meetings. Furthermore, the audio data could be used to stimulate recall in the interview by focusing the interview particularly around

the GMC guidelines, the different types of supporting information and whether doctors reflect on these. Adding another method to the equation by conducting a documentary analysis of appraisal forms would further strengthen this approach by providing written evidence of the types of supporting information doctors are bringing to appraisal and the level of reflection when completing the appraisal form. A survey could then be used to investigate the findings on a larger scale.

(g) Sampling strategy

In line with the GMC's Equality and Diversity strategy,¹³⁷ sampling in the evaluation framework should explore any potential issues by capturing the extent to which revalidation is impacting differently, or not, on different sections of the medical population; in particular whether any differential impact is associated with any of the prescribed protected characteristics such as age, gender, disability etc. In addition, further characteristics or demographics that emerged as important from the stakeholder interviews were differing specialisms and those that take leave (maternity, sabbatical, sick).

Where possible the sampling strategy should also take account of healthcare settings across the following groups: such as primary and secondary care, community services, clinical academics and the independent sector. Doctors in training could also be considered, as mentioned in the original tender documentation. However monitoring mechanisms are already well established for this group and in terms of managing the scope of any initial evaluation we would recommend this group should be part of a later separate evaluation.

While making these suggestions we are aware that the ability to assess the different groups may be limited by currently held data about participants, what they might be prepared to divulge and indeed who consents to taking part in any aspect of the evaluative research.

We suggest that the work stream methods might be carried out in six revalidation evaluation centres across the four nations (England, Northern Ireland, Scotland and Wales). The evaluation centres could be based within six of the fifteen ELA regions. Wales, Scotland and Northern Ireland are each an ELA region and would therefore each host one centre. In addition three revalidation evaluation centres could be chosen in England due to the significantly larger population of doctors. Sampling across the six sites within the four nations should consider the need for rural, urban and private practices to be included. This would not be required in all centres but having one of the

three centres in England in a rural location and one in an urban location could address this aspect.

If interviews were to be conducted with appraisees and appraisers, it would be most beneficial to conduct them until sample saturation is reached which is very participant dependent.¹³⁸ However a minimum of 38 appraisees is a reasonable estimate in each evaluation centre to generate valid and transferable results. Researching until saturation if achieved is a challenging approach because 'it forces the researcher to combine sampling, data collection, and data analysis, rather than treating them as separate stages in a linear process. It also means that it is impossible to specify the exact number of qualitative interviews necessary to complete the research at its inception'.^{138:5} Thus while many experts agree that saturation is ideal, it is often helpful for funders to have some numerical guidance. As an example, Baker advises sampling between 12 and 60, mean = 30.¹³⁸ In order to sample effectively across settings (nation, protected characteristics, specialty, role etc.) multiples of this guidance are likely to be required.

If a mixed-methods approach was adopted in this work stream then ideally each appraisee would consent to having their appraisal recorded and participating in a semi-structured interview. The appraiser of each of the appraisees would also be interviewed so both perspectives of the same appraisal could be compared. In addition, if the documentary analysis of appraisal forms was conducted then approximately half of the sample could also have their appraisal documentation analysed. If a survey was to be conducted with appraisees and appraisers to investigate the findings on a larger scale, it could be carried out in the 6 evaluation centres across the four nations.

We suggest that the work stream methods might be carried out over a three year period with three different options. Interviews could be carried out in (a) years 1 (b) years 1 and 3 or (c) years 1, 2 and 3. Carrying out the methods in year 1 would establish a baseline as a cross-sectional study. However in order to gather evidence over time (in a longitudinal study) the same sample (where possible) could be studied again in subsequent years either in years 1, 2 and 3, or in years 1 and 3. Thus for a one year cross-sectional study, a minimum of 228 appraisees and 228 appraisers would be interviewed, 228 appraisals recorded and 114 appraisees would have their appraisal documentation analysed. For a longitudinal study, the same participants (where possible) would be interviewed and observed again in year 2 and 3, or just year 3. A summary of a possible sampling strategy is provided in Appendix 10.

4.3.2 Work Stream 2: Appraisal (GMC Objective 1, 2 & 3)

Work Stream 2 focuses broadly on appraisal in the revalidation process and addresses the following evaluation questions:

- 1a Is appraisal taking place for all licensed doctors every year?
- 1b If appraisal is not taking place for all licensed doctors what are the reasons?
- 1c How are the GMC guidelines on appraisal (*GMP* framework for appraisal and revalidation) being applied in practice and how might they be enhanced to increase its impact?
- 3a Is the revalidation process increasing the awareness and adoption of the values and principles of *Good Medical Practice*?
- 3b Is the revalidation process increasing professionalism?

To answer these evaluation questions we suggest the following:

- (a) a review of literature on appraisal
- (b) secondary data analysis of appraisal rates
- (c) interviews with ROs
- (d) survey of ROs
- (e) survey of appraisees and appraisers
- (f) semi-structured interviews with appraisees & appraisers
- (g) a mixed-methods approach

Each of these methods is outlined in detail in the following section.

(a) A review of the literature on appraisal

Rationale: Synthesising the literature on appraisal would establish an evidence base on the use of appraisal in the revalidation process and would inform decisions about the development of revalidation. The literature review could also inform the implementation of the methods in the work streams e.g. the questionnaire design, the interview schedules.

Method: A literature review (systematic if appropriate) of the impact of appraisal on doctors' performance could be undertaken. A detailed review protocol should be developed. The following inclusion criteria could be applied: empirical studies written in English, published from 1960 to present, investigating the effect of appraisal on the performance of doctors in any specialty using any method and any outcome measure. The search strategy would involve searching a variety of relevant

databases (e.g. Medline, EMBASE), grey literature, hand-searching key journals, searching trials registers, contacting experts in the field and citation searching. Data would be extracted and quality assessed. The data would be synthesised using qualitative, quantitative or mixed methods approaches. The review would be conducted using established guidance for conducting systematic reviews e.g. the CRD¹²⁷, and for reporting systematic reviews e.g. PRISMA.¹²⁸

Methodological strengths & limitations: Literature reviews organise and make accessible the major findings in an area of inquiry. They are often used to inform policy and underpin future research directions. A major limitation of a review (particularly a systematic review) is that the reviews quality is dependent on the quality of existing studies. So if the quality of existing studies is low then the quality of the results of the systematic review will be low. While a literature review can answer some research questions they will not be able to answer all research questions. Literature reviews can also be quite time and resource-intensive and need to be updated regularly to remain current.

(b) Secondary data analysis (EQ 1a)

Rationale: To investigate whether appraisal is taking place for all licensed doctors, annual appraisal rates across the four nations should be examined. Analysing existing databases would provide national level data on appraisal rates on a yearly basis and could be undertaken to demonstrate annual descriptive statistics, e.g. annual appraisal rates, non-compliance, reasons for non-engagement (if available) and geographical comparative analyses etc. This would help determine whether revalidation has resulted in all doctors being part of a governed system.

Method: We suggest an analysis of secondary data on appraisal rates held in each of the four nations. In England, this data is currently being generated by the Revalidation Support Team (RST) through their benefit measures work or the Organisational Readiness Self-Assessment (ORSA). However we are aware that the RST is being disbanded in March 2014 and this might end. If this annual data collection ends, ROs would need to be surveyed to gather the data as currently the GMC only hold data about revalidation dates and completions and not the annual appraisal that informs them. In Wales, this data is held by the Welsh Deanery and is based on the Medical Appraisal and Revalidation System (MARS)¹ IT system. In Northern Ireland this data is not held centrally by a single organisation but is held by

¹<http://www.marswales.org/>

the different health and social care boards, which would need to be surveyed. In Scotland, appraisal rates are collected by Health Improvement Scotland and are based on the Scottish Online Appraisal Resource (SOAR) IT system.

Sample: All doctors across all settings in the four nations for each year of the evaluation (years 1-3). Where data are available, comparisons should be considered across the various groups of the medical profession with protected characteristics (i.e. age, gender, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, ethnicity, religion or belief, sexual orientation, and also their specialism and those who have taken a period of leave). In addition consideration should be given to doctor settings (i.e. primary care, secondary care, mental health, independent sector, locums, SAS, etc.) in order to explore the extent to which revalidation has impacted differently, or not, in different sections of the doctor population.

Methodological strengths & limitations: There would be no control over what data are contained in each provided submission. There may be differences in the information in each data set across, and in the case of England within, the four nations. Data may be incomplete or inaccurate. In addition, secondary data are often unreliable as they rely on data input and therefore the potential for human error. However data that are available do provide a potential source for relatively cheap and efficient interrogation.

(c) In-depth semi-structured interviews with ROs (EQ 1b, & 1c)

Rationale: Semi-structured interviews with a sample of ROs would provide an in-depth understanding of why appraisal is not taking place for all doctors, how the appraisal rates could be supported and how the GMC guidelines on appraisal might be enhanced.

Method: To address evaluation questions 1b, 3a and 3b we propose in-depth semi-structured interviews should be conducted with a sample of ROs. Interview data would be thematically analysed.

Sample: Four ROs across different settings should be sampled in each of the six evaluation centres across the four nations (24 in total). Interviews could be carried out in (a) years 1 (b) years 1 and 3 or (c) years 1, 2 and 3). Carrying out interviews in year 1 would establish a baseline (as a cross-sectional study) so this can start to

inform future policy decisions about why appraisal is not taking place, and the awareness and adoption of *Good Medical Practice*. However in order to gather evidence over time (in a longitudinal study), especially to address EQ3, the same sample (where possible) could be interviewed again in subsequent years. A summary of sampling numbers is provided in Appendix 10. The interviews could be conducted by patients as researchers to strengthen the PPI element in the evaluation framework.

Methodological strengths & limitations: While semi-structured interviews will provide an important in-depth investigation, they will not provide breadth in terms of sampling. Interview findings are likely to be transferable but not generalizable. Nonetheless, small-scale qualitative studies elicit a more in-depth understanding of phenomenon, how the studied phenomena are experienced and the results often contribute valuable knowledge and understanding that could not simply be achieved by quantitative research using questionnaires. On the downside, interviews are more expensive to conduct as they are more time consuming and resource intensive.

(d) Survey of ROs (EQ 1a - c)

Rationale: A survey of all ROs across the four nations would investigate the number of doctors in their jurisdiction that had not undergone appraisal, the reasons why appraisal had not taken place for all doctors and how might the GMC guidelines on appraisal be enhanced to increase its impact. This would provide standardised and comparable data from a large sample of ROs across the four nations.

Method: In-depth questionnaire survey of ROs. A questionnaire could be designed to capture the numbers (to verify nationally held data) and reasons why all doctors may not have undergone appraisal. This questionnaire would capture qualitative and quantitative data from a large sample of RO's in a standardised format about appraisal activity and could then be analysed using descriptive statistics and a thematic analysis of free text.

Sample: All ROs across the four nations. The questionnaire could be implemented once as a cross-sectional study or more than once (e.g. annually) as a longitudinal study to capture change over time.

Methodological strengths & limitations: Surveys can incorporate a broader/larger sample than, for example, conducting interviews or undertaking ethnographic

studies, and they are generally less resource intensive and therefore cost less to implement. However a major limitation of this method is that they are more narrow in their scope and do not facilitate the in-depth understanding of the nuances often required in the exploration of a topic.

(e) Survey of appraisees and appraisers (EQ 1c, 3a & 3b)

Rationale: In order to explore whether the revalidation process is increasing the awareness and adoption of the values and principles of *Good Medical Practice*, and whether it is increasing levels of professionalism, a survey of appraisees and appraisers could be implemented. This would provide standardised qualitative and quantitative data of the use of *Good Medical Practice* and professionalism across a large sample of appraisees and appraisers in the four nations. This questionnaire could also be utilised to examine how GMC guidelines are being utilised in practice and how they could be enhanced to increase its impact.

Method: In-depth questionnaire survey of appraisees and appraisers. A questionnaire would be designed to capture quantitative and qualitative data in a standardised format about the use of *Good Medical Practice*, professionalism and GMC guidance on appraisal. Data would be analysed using descriptive statistics and a thematic analysis of free text.

Sample: All appraisees and appraisers in each of the six evaluation centres across the four nations could be targeted. Where data are available, comparisons should be considered across the various groups of the medical profession with protected characteristics (i.e. age, gender, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, ethnicity, religion or belief, sexual orientation, and also their specialism and those who have taken a period of leave). In addition consideration should be given to doctor settings (i.e. primary care, secondary care, mental health, independent sector, locums, SAS, etc.) in order to explore the extent to which revalidation has impacted differently, or not, in different sections of the doctor population. The questionnaire could be implemented in one year or more years the obvious benefits of a longitudinal study in being able to investigate trends over time.

Methodological strengths and limitations: Surveys are generally less resource intensive and therefore cost less to implement. They can incorporate a broader/larger sample than, for example, conducting interviews or undertaking

ethnographic studies, but they are generally more limited in their scope and do not facilitate the in-depth understanding of the nuances often required.

(f) In-depth semi-structured interviews with appraisees and appraisers (EQs1c, 3a & 3b)²

Rationale: In order to explore whether the revalidation process is increasing the awareness and adoption of the values and principles of *Good Medical Practice*, and whether it is increasing levels of professionalism we propose conducting in-depth semi-structured interviews with appraisees and appraisers. This would provide a nuanced insight into what is actually happening on the ground in terms of the awareness and adoption of the values and principles of *Good Medical Practice* and whether levels of professionalism are changing. The interviews would also be utilised to get a more in-depth understanding of how GMC guidance on appraisal could be used to enhance its impact.

Method: In-depth semi-structured interviews with appraisees and appraisers. Interview data will need transcribing, coding with the support of suitable software (e.g. NVivo) and then analysed using thematic and discourse analyses. Again interviews could be conducted by patients as researchers. Having the patient perspective in this process may elicit important additional information and insight.

Sample: Appraisees and appraisers should be sampled across the four nations in different settings including primary care, secondary care, mental health consultants, independent practice doctors, primary care locums, secondary care locums, SAS or trust grade doctors and clinical academics. Where data are available, comparisons should be considered across the various groups of the medical profession with protected characteristics (i.e. age, gender, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, ethnicity, religion or belief, sexual orientation, and also their specialism and those who have taken a period of leave). In addition consideration should be given to doctor settings (i.e. primary care, secondary care, mental health, independent sector, locums, SAS, etc.) in order to explore the extent to which revalidation has impacted differently, or not, in different sections of the doctor population. Again a minimum of 38 appraisees and 8

²It should be noted that these interviews are clearly methodically similar to those proposed in WS 1. If both work streams were commissioned alongside each other these two areas of the evaluation would be conducted in the same interviews.

appraisers per centre are likely to be required to generate transferable and valid results. In total therefore approximately 228 interviews with appraisees and 38 interviews with appraisers is proposed. Again, we propose that the interviews are carried out in year 1 to establish a baseline (as a cross-sectional study) and repeated in years 2 and 3 (as part of a longitudinal study). A summary of sampling numbers is provided in Appendix 10. Interviews could be conducted by patients as researchers to enhance patient involvement in the evaluation framework.

Methodological strengths and limitations: While semi-structured interviews will effectively provide an in-depth investigation of a topic, they will not provide breadth in terms of sampling. Interview findings are transferable but not generalisable. A final limitation of interviews is that they can be time consuming and resource intensive.

(g) A mixed-methods approach

Each individual method proposed in this work stream has its own strengths and limitations and all of the evaluation questions cannot be addressed by any individual method. Similar to the first work stream, a mixed-method approach would overcome these issues. Two or three methods could be adopted and implemented in different sequences. For example, the survey of ROs would provide some level of explanation for the trends observed in the analysis of secondary data held at the national level. Interviewing the ROs afterwards would then provide an in-depth understanding of these trends. Or vice versa, interviews could be conducted with ROs to provide an in-depth explanation of the trends observed in the analysis of secondary data held at the national level. A survey could then be designed based on the findings of the interviews to explore the interview findings on a larger scale.

In terms of timescale, the proposed methods could be conducted once in the evaluation cycle or more depending on whether cross-sectional or longitudinal data was being sought.

4.3.3 Work Stream 3: RO judgement-making (GMC Objective 4 & 5)

Work Stream 3 focuses broadly on RO judgment-making in the revalidation process and addresses the following evaluation questions:

- 4a Does revalidation facilitate the identification and addressing of potential concerns earlier and before they become safety issues or fitness to practise referrals?
- 5a How is revalidation supporting ROs to fulfil their statutory function of advising the GMC about the fitness to practise of their doctors?
- 5b How do ROs make judgements?
- 5c How is the RO protocol being applied in practice and how can it be improved?
- 5d Do RO recommendation statements shape RO behaviour?

To answer these evaluation questions we suggest the following methods:

- (a) a review of literature on medical regulation
- (b) semi-structured interviews with ROs, ELA's and other relevant personnel
- (c) a survey of RO's, ELA's and other relevant personnel
- (d) a case control study using PAPC or FTP data with appraisal documentation and RO interviews
- (e) a mixed-methods approach

Each of these methods is outlined in detail in the following sections.

(a) Literature review on medical regulation

Rationale: In order to support the GMC's role as a regulator in the revalidation process a review of the literature into how medical regulation is shaped across the world would be beneficial. This review could inform decisions about the development of revalidation and future empirical research in the field.

Method: A narrative literature review investigating how medical regulation is shaped in a variety of different countries around the world. The search strategy would involve searching a variety of relevant databases e.g. Medline, EMBASE, grey literature, hand-searching key journals, searching trials registers, contacting experts in the field and citation searching. The literature would be synthesised using a narrative approach.

Methodological strengths & limitations: Literature reviews identify and make accessible the major findings in a research field, inform policy development and underpin future

research directions identify gaps in the research. Literature reviews can be quite time and resource-intensive.

(b) Semi-structured interviews with ROs³, ELAs & other relevant personnel (EQ 4a & 5a-d)

Rationale: To address evaluation questions 4a and 5a-d qualitative interviews with a sample of ROs and ELAs could be undertaken to explore their experiences of the revalidation process in terms of whether it supports the identification of fitness to practice concerns. Similarly, how ROs are identifying fitness to practise concerns; how revalidation is supporting ROs to fulfil their statutory function of advising the GMC about the fitness to practise status of their doctors; how ROs go about making their judgements; how the RO protocol is being applied in practice; how it can be improved; and whether the RO recommendation statements are shaping RO behaviour. These interviews would enable the GMC to identify how revalidation fits with other improvement and accountability processes in the system, and how it might be used by, for example, employing organisations, professional bodies and system regulators. In addition to interviewing RO's and ELAs; CEO's, Chairs, Human Resources personnel and quality managers within the same organisation as the RO could also be interviewed in order to get a wider systems' perspective.

Methods: In-depth semi-structured interviews with a sample of ROs and ELAs. Interview data will need transcribing, coding with the support of suitable software (e.g. NVivo) and then analysed using thematic and discourse analyses.

Sample: Four ROs could be sampled in each of the six evaluation centres across the four nations (24 ROs in total). In addition, a CEO, Chair, Human Resources personnel and quality managers from the same organisation as the RO could be interviewed. Interviews could be carried out in (a) years 1 (b) years 1 and 3 or (c) years 1, 2 and 3. Carrying out interviews in year 1 would establish a baseline (as a cross-sectional study). However in order to gather evidence over time (in a longitudinal study) the same sample (where possible) could be interviewed again in subsequent years. A summary of potential sampling numbers is presented in Appendix 10.

³It should be noted that these interviews are clearly methodologically similar to those proposed in WS 2. If both work streams were commissioned these two areas of the evaluation would be conducted in the same interviews.

Methodological strengths and limitations: While semi structured interviews will effectively provide an in-depth investigation of a topic, they will not provide breadth in terms of sampling. Qualitative interviews are time and resource intensive. They do not generate generalizable findings but findings are transferable.

(c) Survey of RO's, ELAs & other relevant personnel (EQ 4a & 5a-d)

Rationale: A survey of RO's, ELAs and other relevant personnel could be utilised to explore whether revalidation facilitates the identification and addressing of potential concerns earlier and before they become safety issues or FTP referrals. They could help to establish how revalidation is supporting ROs to fulfil their statutory function of advising the GMC about FTP of doctors, how ROs make judgements, how the RO protocol is being applied in practice and how it can be improved and whether the RO regulations shape RO behaviour. This would provide standardised quantitative and qualitative data which could be compared and contrasted across regions and sampling groups.

Method: In-depth questionnaire survey of appraisees and appraisers. A questionnaire should be designed to capture quantitative and qualitative data in a standardised format addressing the above research questions. Data would be analysed using descriptive statistics and a thematic analysis of free text.

Sample: In order to maximise response rates, all RO's, ELAs and other personnel in each of the proposed evaluation centres across the four nations could be targeted. The survey could be carried out in (a) years 1 (b) years 1 and 3 or (c) years 1, 2 and 3 of the 3 year evaluation cycle. Carrying out interviews in year 1 would establish a baseline (as a cross-sectional study). However in order to gather evidence over time (in a longitudinal study) the same sample (where possible) could be interviewed again in subsequent years.

Methodological strengths & limitations: The key strength of a survey methodology is that they can target a broader/larger sample than, for example, conducting interviews or undertaking ethnographic studies. Furthermore, surveys are generally less resource intensive and thus cost less to implement. The main limitation of survey methodology is that they are more limited in their scope and do not enable the in-depth understanding of a topic.

(d) Case control study (EQ 4a)

Rationale: To address evaluation question 4a a qualitative case control comparative study could be utilised. The outcome of RO judgements within revalidation is either a recommendation to revalidate, to defer or to highlight non-engagement. However revalidation is ultimately an evaluation of a doctor's fitness to practise. A case control study would examine the differences between doctors who have been referred to FTP procedures and controls to see if particular parts of the revalidation process are more instrumental in detecting these cases than others. It could for example be hypothesised that some types of supporting information may trigger more referrals than others or that other events or evidence outside of revalidation in fact trigger FTP referrals.

Method: A prospective qualitative case control comparative study. This study would explore differences between appraisal and revalidation documentation between those who had been referred to the GMC Fitness-to-Practise (FTP) procedures and a group of controls. During the evaluation period any doctor referred to the GMC (from Wales or Scotland) would be matched to a 'control'. Controls would be chosen and matched to cases by demographic similarities, such as age, gender, clinical area, practice size etc. The appraisal documentation would be analysed from both the referred doctor and the matched control using a content analysis. In addition an in-depth semi-structured interview could be undertaken with the relevant RO. The null hypothesis would be that ROs do not use revalidation to inform FTP decisions.

Sample: We propose that this could, in the first instance be piloted in Scotland and Wales as these nations have nationally held computer based appraisal datasets including historical appraisal data. This study could recruit over the three years of the evaluation cycle so that hopefully enough data on FTP referrals would be available for the analysis.

Methodological strengths & limitations: As this is a qualitative case control study, the findings would be transferable but not generalizable. In addition, there may be difficulties in gaining enough FTP data to conduct a feasible study.

(e) A mixed-methods approach

Interviews or surveys with RO's, ELAs and other relevant personnel would address all of the questions to varying degrees in this work stream. However with a survey, breadth of sampling would be achieved at the cost of an in-depth understanding; while interviews would achieve depth and a richer understanding but with a loss of breadth. Implementing both methods would offset these limitations and enable triangulation of results thus increasing the robustness of the conclusions. The data generated from the

semi-structured interviews could also be used to design the questionnaire for the large-scale survey of RO's etc. or vice versa.

4.3.4 Work Stream 4: Patient & Public Involvement (addresses role of GMC)

Work Stream 4 focuses broadly on the role of PPI in the revalidation process and addresses the following evaluation questions:

6a. What opportunities are there for the patient perspective to be reflected in revalidation processes?

6b. What level of involvement do patients and the public want in the revalidation process?

To answer these evaluation questions we would suggest:

- (a) a review of literature on PPI
- (b) conducting semi-structured interviews with stakeholders in revalidation
- (c) conducting a survey of patients and the public
- (d) conducting semi-structured interviews with patients and the public
- (e) a mixed-methods approach

Each of these methods is outlined in detail in the following sections.

(a) Literature review

Rationale: A review of the literature to establish the processes and rationale for PPI in medical regulation, and more specifically in revalidation, would provide a sound evidence-base to inform empirical research on the role of PPI in revalidation and the development of revalidation around PPI.

Method: A narrative review of the literature investigating the role of PPI in international medical regulation and in particular in revalidation. The search strategy would involve searching a variety of relevant databases e.g. Medline, EMBASE, grey literature, hand-searching key journals, searching trials registers, contacting experts in the field and citation searching. The literature would be synthesised using a narrative approach. The CAMERA research team are currently carrying out a study of patient and public involvement in medical revalidation for the Revalidation Support Team (RST).¹³⁹ The RST aim to use the findings to develop guidance on patient and public involvement in medical revalidation. This work stream would build on this research.

Methodological strengths & limitations: The key strength of literature reviewing methods is that they identify and make accessible the major findings in a research area, they inform policy development and they underpin future research directions by identifying the gaps in the research. However, an in-depth rigorous literature review can be time and resource-intensive.

(b) In-depth interviews with key stakeholders in revalidation (EQ 6a)

Rationale: In order to explore the potential opportunities for the patient perspective to be reflected in the revalidation process, a variety of key stakeholders in the revalidation process could be interviewed to find out where they see potential opportunities in their particular area of the process.

Method: In-depth semi-structured interviews with key stakeholders in revalidation. The data generated by the interviews would be transcribed, coded and then analysed using thematic and discourse analyses.

Sampling: A sample of participants with a stake in the revalidation process in each of the six evaluation centres across the four nations: responsible officers, patients, patient groups, relevant GMC departments, CEO's, Chairs, Human Resources personnel, quality managers, the Care Quality Commission, the Royal Colleges, the British Medical Association, Ethnic Minority groups and the Departments of Health. Participants should be sampled until data saturation is reached but approximately 20 participants are likely to be sufficient in each centre. Interviews could be carried out in (a) years 1 (b) years 1 and 3 or (c) years 1, 2 and 3 of the 3 year evaluation cycle.

Methodological strengths & limitations: While semi structured interviews will effectively provide an in-depth investigation of a topic, they will not provide breadth in terms of sampling. Qualitative interviews are time and resource intensive. They do not generate generalisable findings, but findings are transferable.

(c) Survey of patients and the public (EQ 6b)

Rationale: A survey of patients and the public across the four nations could investigate the level of involvement that patients and the public would like in the revalidation process.

Method: In-depth questionnaire survey of patients and the public. The questionnaire would generate both quantitative and qualitative data and would provide standardised

and comparable data across the four nations and would be analysed using descriptive statistics and a thematic analysis of free text.

Sample: Patients and the public are terms often used interchangeably in the context of PPI but they are not the same. Patient refers to the relationship between a person requiring medical care (active healthcare users)¹⁴⁰ and their individual doctor or medical team while the public encapsulates the collective of patients as a stakeholder group which would also include 'well' patients e.g. Healthwatch. The survey should sample across these different groups in each of the six evaluation centres in the four nations. The survey could be carried out in (a) years 1 (b) years 1 and 3 or (c) years 1, 2 and 3, of the 3 year evaluation cycle.

Methodological strengths & limitations: Surveys are generally quite limited in scope and do not facilitate an in-depth understanding of a topic.

(d) Semi-structured interviews with patients and the public (EQ 6a & b)

Rationale: Interviews with patients and the public would provide the opportunity to conduct an in-depth exploration of the level of involvement the patient and public would like in the revalidation process in order to inform the development of PPI.

Method: In-depth semi-structured interviews with a sample of patients and the public. Interview data would need transcribing, coding with the support of suitable software (e.g. NVivo) and then analysed or interrogated using thematic and discourse analyses. Using this method would build on the work that has been commissioned by the RST and is currently being carried out by the CAMERA research team on strengthening PPI in revalidation.

Sampling: A sample of patients and the public in each of the six evaluation centres across the four nations. Participants including patients and the public i.e. patient groups, should be sampled until data saturation is reached but approximately 20 participants would be sufficient in each centre. Interviews could be carried out in (a) years 1 (b) years 1 and 3 or (c) years 1, 2 and 3 of the 3 year evaluation cycle.

Methodological strengths & limitations: While semi structured interviews will effectively provide an in-depth investigation of a topic, they will not provide breadth in terms of sampling. Qualitative interviews are time and resource intensive. They do not generate generalisable findings but findings are transferable.

(e) A mixed-methods approach

A combination of the proposed methods would address all of the evaluation questions in this work stream. The interviews with key stakeholders would identify the potential opportunities for patient and public involvement. These opportunities could then be presented to patients and the public in interviews in order to elicit their opinions on whether they would be interested in being part of these initiatives and to gain an understanding of the level of involvement they would like in general in the revalidation process. The interview findings could then be used to generate a questionnaire which would capture a large sample of respondents in centres across the four nations. These individual studies could be initiated sequentially as outlined above or at numerous points in the evaluation cycle in order to generate longitudinal data on the changes over time.

4.4 Summary of Framework

The proposed Revalidation Evaluation Framework is underpinned by the GMC's regulatory function to protect, promote and maintain the health and safety of the public by ensuring proper standards in medical practice. Addressing the five proper standards outlined by the GMC and the role of the patient and public in revalidation in the framework, generated a series of important evaluation questions that could be addressed by four work streams focusing broadly on; supporting information, appraisal, RO judgement-making and PPI.

Work Stream 1 addresses supporting information and includes a literature review, a survey and semi-structured interviews with appraisees and appraisers, audio-recording appraisal meetings and documentary analysis of appraisal forms. Work Stream 2 addresses the appraisal and includes a literature review, secondary data analysis of appraisal rates, a survey and interviews with ROs, and a survey and interviews with appraisees and appraisers. Work Stream 3 addresses RO judgement-making and includes a literature review, a survey and semi-structured interviews with ROs and ELAs and other relevant personnel, and a case control study. Work Stream 4 addresses PPI and includes a literature review, interviews with key stakeholders in revalidation, and a survey and interviews with patients and the public.

In summary, the work streams propose a mixed-methodological approach using both quantitative and qualitative techniques, and drawing on both process and outcome evaluation models. They are also multi-perspective, collecting data from a variety of stakeholders in the revalidation process including appraisees, appraisers, ROs, ELAs, Chairs, CEOs, Human Resource Managers, patients, patient groups, quality managers, the Care

Quality Commission, the Royal Colleges, the British Medical Association, Ethnic Minority groups and the Departments of Health. Patients and the public are also involved in the data collection as researchers conducting interviews with appraisees, appraisers, ELAs and ROs. By organising the evaluation into work streams centred on four key revalidation processes, the framework will generate a variety of fundamental information that will help the GMC make strategic decisions about the design features of revalidation.

The framework proposes conducting the work streams in six revalidation evaluation centres across the four nations (England, Northern Ireland, Scotland and Wales). The evaluation centres are based on six of the fifteen ELA regions, although some methodologies draw on national level data collection and national level surveys. The proposed sampling strategy is wide-ranging aiming to capture the diversity of protected characteristics in the doctor population such as gender, age, ethnicity etc. as well as doctors in a variety of settings and roles, such as locum doctors, as well as doctors in primary and secondary care. This wide-ranging sampling strategy would enable the investigation of whether doctors from different backgrounds, settings and nations are experiencing revalidation similarly. The use of CHAT as a mapping tool will allow for an overall analysis which highlights the impact of any variation in process or population within the activity system of revalidation.

A longitudinal design has been suggested in the framework with each work stream collecting data over a three year period. However the evaluation design is both formative and summative. Interim results/formative feedback would be available at the end of the first year so that the process can evolve. The summative output of the evaluation will be available to inform potential changes to the revalidation process before the start of the second revalidation cycle in 2018.

There is significant overlap in terms of methods in the four work streams which has implications for how the work streams could be commissioned by the GMC. The potential commissioning options are presented in the following section.

5. Commissioning Considerations

Strategic decisions will need to be made by the GMC in terms of the parts of the Evaluation Framework that will be commissioned and the time frame. Funding is always a limitation in any research endeavour and this will have a big impact on the commissioning and implementation of the evaluation. In the following section we present how the commissioning of work streams should be considered in relation to each other and in relation to the conceptual framework of CHAT. This is followed by a section on commissioning in practice.

5.1 The use of Cultural Historical Activity Theory (CHAT) to help shape commissioning

Revalidation is a complex intervention in an even more complex system of healthcare delivery. To this end our evaluative framework, for consideration by the GMC, has been underpinned by a conceptual framework (CHAT) which maps together the different dimensions of revalidation for evaluative data collection and data analysis. CHAT seeks to move thinking away from isolated research activities to a more robust multi-method approach and will help the GMC draw simple messages out of complex systems.

CHAT stresses the importance of capturing information (data collection) about as many of the revalidation processes as possible to enable a complete understanding of what is really happening as revalidation is implemented. The commissioning of any of the proposed methods in the work streams will provide useful insights into particular aspects of each of the revalidation 'activity' systems (components of revalidation). However it is only with a longitudinal, systematic, mixed-methods approach across all four work streams, will the GMC be able to generate evidence across the complexity and begin to understand what is happening within and between revalidation activity systems as they play out and interact with each other over time.

Additionally the application of the findings (data analysis) of all four work streams to a CHAT model for the components of revalidation will enable the GMC to map the diverse experiences of PPI, appraisees, the different approaches of ROs to judgement-making, and the multiple processes that inform the use of supporting information. These can then be modelled, as required, to examine the implications of variation in the revalidation system (for example; different conditions, changes in activity, levels of participation, use of specific artefacts and so on). Such analysis will be useful to considering where the value-added of revalidation can be maximised, or where shortcomings can be identified, thus providing evidence for further development of its processes.

5.2 Commissioning in practice

It is unlikely that all of the framework could be initiated and sustained throughout a longitudinal study due to the cost implications. However the GMC working with academic partner(s) could make strategic decisions about what to commission when and what should then follow. In addition economies of scale could be realised by covering topics in different workstreams using the same methods. For example in work stream 1 and work stream 2 interviews are proposed with appraisees and appraisers. If both work streams were commissioned these two areas of the evaluation could be conducted in the same interviews.

By mapping results from initial research onto the framework informed by CHAT it will be possible to ascertain where the gaps in knowledge are over time. By discovering for example that doctors (subjects) struggle with provided eSolutions (artefacts) that are meant to support revalidation by supporting information and the administration of the appraisal processes, further research could be commissioned to explore and compare different solutions across England and or between devolved nations of the UK. This might focus on the interactions between eSolutions (artefacts) and other aspects of the activity systems such as the wider community and rules relating to eSolutions (rather than working directly with doctors as subjects) in order to understand the wider dimensions.

6. Conclusions

The Revalidation Evaluation Framework was developed in collaboration with the GMC, the views of stakeholders in the revalidation process and the related literatures. Combined with the CAMERA team's research experience in regulatory assessment and knowledge of research methodology, the framework addresses the needs of the stakeholders using innovative rigorous methods ensuring that the evaluation framework could be used to explore all the key questions of revalidation policy in action. The framework suggests a range of methods to understand revalidation at four key points of inquiry across revalidation activity. These are the uses of supporting information, the process of appraisal, the judgement-making decisions of ROs and the role of patients and the public.

The mixed-methods, multi-perspective approach suggested would support the ability to triangulate results, adding rigour to the transferability of findings into future policy and practice.

The use of CHAT as a conceptual framework acknowledges the complexity of revalidation as activity in a complex system and the need for a combination of evaluative approaches overtime rather than singular measures or a series of unrelated measures.

The involvement of patients as researchers potentially strengthens the patient and public dimension of the framework as guidance on PPI indicates that patients and the public always offer unique and invaluable insights. Their advice when designing, implementing and evaluating research invariably makes studies more effective, more credible and often more cost effective.

This framework will allow the GMC to better understand the process and early regulatory impact of revalidation in practice and will establish opportunities for the effectiveness and efficiency of revalidation processes and design features to be improved in the future.

7. References

1. Department of Health. *Trust, Assurance and Safety; the Regulation of Health Professionals in the 21st Century*. 2007; Available from: <http://www.official-documents.gov.uk/document/cm70/7013/7013.pdf>.
2. Archer, J., et al. *Revalidation: In policy*. 2012; Available from: <http://www1.plymouth.ac.uk/peninsula/research/camera/revalidation/Documents/Stage-One-Report.pdf>.
3. Department of Health. *Good doctors, safer patients: Proposals to strengthen the system to assure and improve the performance of doctors and to protect the safety of patients*. 2006; Available from: <http://www.stateclaims.ie/ClinicalIndemnityScheme/publications/GoodDoctors.pdf>.
4. General Medical Council. *Consultation Document: Revalidating Doctors. Ensuring standards, securing the future*. 2000.
5. House of Commons Health Committee. *Revalidation of Doctors, Fourth Report of Session 2010-2011, Volume 1: Report together with formal minutes and oral and written evidence*. 2011; Available from: www.publications.parliament.uk/pa/cm201011/cmselect/cmhealth/.../557.pdf: 7.
6. United Kingdom, *The Medical Act 1983 (Amendment) Order 2002.*, in *SI 2002/31352002*: United Kingdom.
7. General Medical Council. *Starting revalidation*. 2013; Available from: <http://www.gmc-uk.org/doctors/revalidation/13273.asp>.

8. General Medical Council, *Research Specification: Evaluating the strategic impact of medical revalidation: building an evaluation framework*, 2013, General Medical Council: London.
9. *The Medical Profession (Responsible Officers) Regulations in SI 2010/28412010*: United Kingdom.
10. *The Medical Profession (Responsible Officers) Regulation (Northern Ireland) 2010*, in *SR 2010/2222010*: United Kingdom.
11. General Medical Council. *Making revalidation recommendations: the GMC responsible officer protocol*. 2012; Available from: http://www.gmc-uk.org/FINAL_GMC_RO_Protocol_03.12.2012.pdf_50553987.pdf.
12. General Medical Council. *Ready for revalidation: Supporting information for appraisal and revalidation*. 2012; Available from: http://www.gmc-uk.org/static/documents/content/Supporting_information_for_appraisal_and_revalidation.pdf.
13. General Medical Council. *The Good Medical Practice framework for appraisal and revalidation*. 2012; Available from: http://www.gmc-uk.org/static/documents/content/GMC_Revalidation_A4_Guidance_GMP_Framework_04.pdf.
14. General Medical Council. *Good Medical Practice*. 2013; Available from: http://www.gmc-uk.org/static/documents/content/GMP_2013.pdf_51447599.pdf.
15. Smith, R., *Profile of the GMC: The day of judgment comes closer*. *BMJ*, 1989. **298**(6682): p. 1241-1244.

16. Brennan, N., et al., *Current practice for supporting the maintenance of professional competence in practising doctors: Strategic review of the Medical Council of Ireland's arrangements for the maintenance of professional competence.*, 2013, CAMERA, Plymouth University, Peninsula Schools of Medicine and Dentistry.
17. RAND Europe. *International Comparison of Ten Medical Regulatory Systems: Egypt, Germany, Greece, India, Italy, Nigeria, Pakistan, Poland, South Africa and Spain.* 2009; Available from:
http://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR691.pdf.
18. RST. *Independent Evaluation of Medical Revalidation Pathfinder Pilot.* 2011 [cited http://www.revalidationsupport.nhs.uk/files/DE0270-00-DH-MedRev-Summary_Report_FINAL.pdf].
19. RST. *Medical Appraisal Guide Pilot.* 2012; Available from:
<http://www.revalidationsupport.nhs.uk/CubeCore/.uploads/documents/pdf/MAGpiloreport201112v1.1.pdf>.
20. RST. *Organisational readiness self-assessment (ORSA) Report.* 2012; Available from:
http://www.revalidationsupport.nhs.uk/responsible_officer/Organisational_management/responsible_officer_orsa.php.
21. Galvin, K., *Navigating a qualitative course in programme evaluation*, in *Qualitative Research in Health Care*, I. Holloway, Editor. 2005, Open University Press: Maidenhead. p. 229-249.
22. Stufflebeam, D., ed. *Foundational models for 21st century program evaluation.* Evaluation models: Viewpoints on educationla and human services evaluation, ed. D.

- Stufflebeam, G. Madaus, and T. Kellaghan. 2000, Kluwer Academic Publishers: Boston, MA.
23. Hansen, H.F., *Choosing Evaluation Models: A Discussion on Evaluation Design*. Evaluation, 2005. **11**(4): p. 447-462.
 24. SAMHSA. *Using process evaluation to monitor program implementation*. 2013; Available from: <http://captus.samhsa.gov/access-resources/using-process-evaluation-monitor-program-implementation>.
 25. Community Interventions for Health. *Process evaluation*. 2013; Available from: http://www.oxha.org/cih_manual/index.php/process-evaluation.
 26. James Bell Associates. *Evaluation Brief: Whats the difference? Understanding process and outcome evaluation*. 2007; Available from: <http://www.jbassoc.com/reports/documents/understanding%20process%20and%20outcome%20evaluation.pdf>.
 27. General Medical Council, *Revalidation evaluation framework: Regulatory questions*, 2013, General Medical Council: London.
 28. Popay, J., et al. *Guidance on the conduct of narrative synthesis in systematic reviews*. ESRC Methods Programme 2006; Available from: <http://www.lancs.ac.uk/shm/research/nssr/research/dissemination/publications.php>.
 29. Schlosser, R.W., et al., *Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: the application of traditional and comprehensive Pearl Growing. A review*. International Journal of Language & Communication Disorders, 2006. **41**(5): p. 567-582.

30. Overeem, K., et al., *Doctor performance assessment in daily practise: does it help doctors or not? A systematic review*. Medical Education, 2007. **41**(11): p. 1039-1049.
31. Boylan, O., T. Bradley, and A. McKnight, *GP perceptions of appraisal: professional development, performance management, or both?* British Journal of General Practice, 2005. **55**(516): p. 544-545.
32. Bruce, D., et al., *Revalidation for general practitioners: randomised comparison of two revalidation models*. BMJ, 2004. **328**(7441): p. 687-691.
33. Campbell, C., et al., *A study of the factors that influence physicians' commitments to change their practices using learning diaries*. Academic Medicine, 1999. **74**(10): p. S34-6.
34. Campbell, C.M., et al., *Study of physicians' use of a software program to create a portfolio of their self-directed learning*. Academic Medicine, 1996. **71**(10): p. S49-51.
35. Challis, M., et al., *Portfolio-based learning: continuing medical education for general practitioners — a mid-point evaluation*. Medical Education, 1997. **31**(1): p. 22-26.
36. Dornan, T., C. Carroll, and J. Parboosingh, *An electronic learning portfolio for reflective continuing professional development*. Medical Education, 2002. **36**(8): p. 767-769.
37. Flood, S., *Using qualitative self-evaluation in rating physician performance*. Family Practice Management, 1998. **5**(22-34).
38. Lewis, M., G. Elwyn, and F. Wood, *Appraisal of family doctors: an evaluation study*. British Journal of General Practice, 2003. **53**(491): p. 454-460.

39. Mathers, N.J., et al., *Portfolios in continuing medical education – effective and efficient?* Medical Education, 1999. **33**(7): p. 521-530.
40. McKinstry, B., H. Peacock, and J. Shaw, *GP experiences of partner and external peer appraisal: a qualitative study*. British Journal of General Practice, 2005. **55**(516): p. 539-543.
41. Melville, C., et al., *Portfolios for assessment of paediatric specialist registrars*. Medical Education, 2004. **38**(10): p. 1117-1125.
42. CDC. *Framework for program evaluation in public health*. 1999; Available from: <ftp://ftp.cdc.gov/pub/Publications/mmwr/rr/rr4811.pdf>.
43. Potter, J. and M. Wetherell, *Natural order: why social psychologists should study (a constructed version of) natural language, and why they have not done so*. Journal of Language and Social Psychology, 1995. **14**(1-2): p. 216-222.
44. Kitzinger J, *Qualitative Research: Introducing focus groups*. BMJ, 1995. **311**(7000): p. 299-302.
45. Morgan D, *Focus Groups in Qualitative Research*. 2nd ed. Qualitative Research Methods. Vol. 16. 1997, London: Sage.
46. Jupp, V., *The Sage Dictionary of Social Research Methods*. 2006, London: Sage.
47. Saldarna, J., *The Coding Manual for Qualitative Researchers*. 2nd Edition ed. 2013, London: Sage.
48. Walt G and Gilson L, *Reforming the health sector in developing countries: the central role of policy analysis*. Health Policy and Planning, 1994. **9**(4): p. 353-370.

49. Vygotsky LS, *Mind in society: The development of higher psychological processes*. 1978, Cambridge, MA: Harvard University Press.
50. Roth WM and Lee YJ, "Vygotsky's Neglected Legacy": *Cultural-Historical Activity Theory*. *Review of Educational Research*, 2007. **77**(2): p. 186-232.
51. Blackler F, Crump N, and McDonald S, *Organising processes in complex activity networks*. *Organization*, 2000. **7**: p. 277-300.
52. Engestrom Y, *Developmental studies of work as a testbench of activity theory: The case of primary care medical practice*, in *Understanding practice: Perspectives on activity and context*, Chaiklin S and Lave J, Editors. 1993, Cambridge University Press: Cambridge UK. p. 64-103.
53. Engestrom Y, *Activity theory and individual and social transformation*, in *Perspectives on activity theory*, Engestrom Y, Miettinen R, and Punamaki RL, Editors. 1999, Cambridge University Press: Cambridge, UK. p. 19-38.
54. Engestrom Y, *Learning by expanding: An activity-theoretical approach to developmental research*. 1987, Helsinki, Finland: Orienta-Konsultit.
55. Mugweni, K., S. Kibble, and M. Conlon, *Benefits of appraisal as perceived by general practitioners*. *Education for Primary Care*, 2011. **22**(6): p. 393-398.
56. Wright, C., et al., *Multisource Feedback in Evaluating the Performance of Doctors: The Example of the UK General Medical Council Patient and Colleague Questionnaires*. *Academic Medicine*, 2012. **87**(12): p. 1668-1678
57. Murphy, D.J., et al., *Insightful practice: a reliable measure for medical revalidation*. *BMJ Quality & Safety*, 2012. **21**(8): p. 649-656.

58. Ivers, N., et al., *Audit and feedback: effects on professional practice and healthcare outcomes*. Cochrane Database Syst Rev, 2012. **6**.
59. Conway Ph and C.C. K., *Engaging physicians and leveraging professionalism: A key to success for quality measurement and improvement*. JAMA, 2012. **308**(10): p. 979-980.
60. Iglehart, J.K. and R.B. Baron, *Ensuring physicians' competence — Is maintenance of certification the answer?* New England Journal of Medicine, 2012. **367**(26): p. 2543-2549.
61. American Board of Medical Specialties. *MOC Competencies and Criteria*. 2013; Available from: https://www.abms.org/maintenance_of_certification/MOC_competencies.aspx.
62. American Board of Medical Specialities. *ABMS member board certification: Myths and Facts - Research findings from the ABMS evidence library*. 2013 [cited 2013 7th of June]; Available from: http://www.abms.org/MOC_Myths_And_Facts/#page-1.
63. Duffy, F.D., et al., *Self-assessment of practice performance: development of the ABIM Practice Improvement Module (PIM)*. J Contin Educ Health Prof, 2008. **28**(1): p. 38-46.
64. Hagen, M.D., et al., *Maintenance of certification for family physicians (MC-FP) Self Assessment Modules (SAMs): The first year*. The Journal of the American Board of Family Medicine, 2006. **19**(4): p. 398-403.
65. Hagen, M.D., W. Sumner, and H. Fu, *Diuretic of Choice in ABFM Hypertension Self-Assessment Module Simulations*. The Journal of the American Board of Family Medicine, 2012. **25**(6): p. 805-809.

66. Hess, B.J., et al., *The Association Between Physicians' Cognitive Skills and Quality of Diabetes Care*. *Academic Medicine*, 2012. **87**(2): p. 157-163
67. Hess, B.J., et al., *Setting a fair performance standard for physicians' quality of patient care*. *Journal of general internal medicine*, 2011. **26**(5): p. 467-73.
68. Holmboe, E.S., et al., *Promoting physicians' self-assessment and quality improvement: the ABIM diabetes practice improvement module*. *J Contin Educ Health Prof*, 2006. **26**(2): p. 109-19.
69. Holmboe E.S, W.Y.M.T.P. and et al., *Association between maintenance of certification examination scores and quality of care for medicare beneficiaries*. *Archives of Internal Medicine*, 2008. **168**(13): p. 1396-1403.
70. Mandel K and Kotagal UR, *Pay for performance alone cannot drive quality*. *Archives of Pediatrics & Adolescent Medicine*, 2007. **161**(7): p. 650-655.
71. Miles, P.V., *Maintenance of Certification: The Profession's Response to Physician Quality*. *The Annals of Family Medicine*, 2011. **9**(3): p. 196-197.
72. Miller, M.R., et al., *Decreasing PICU catheter-associated bloodstream infections: NACHRI's quality transformation efforts*. *Pediatrics*, 2010. **125**(2): p. 206-213.
73. Turchin, A., et al., *Effect of Board Certification on Antihypertensive Treatment Intensification in Patients With Diabetes Mellitus*. *Circulation*, 2008. **117**(5): p. 623-628.
74. Sharp, L.K., et al., *Specialty board certification and clinical outcomes: the missing link*. *Academic Medicine*, 2002. **77**(6): p. 534-542.

75. Adamson, J., *Five-year history of the American Society of Plastic and Reconstructive Surgeons, 1969-1973*. *Plastic and Reconstructive Surgery*, 1975. **55**(4): p. 445&hyphen.
76. Brook, R.H., et al., *Carotid endarterectomy for elderly patients: predicting complications*. *Annals of Internal Medicine*, 1990. **113**(10): p. 747-753.
77. Dale Tussing, A. and M.A. Wojtowycz, *The effect of physician characteristics on clinical behavior: cesarean section in New York State*. *Social Science & Medicine*, 1993. **37**(10): p. 1251-1260.
78. Haas, J.S., E.J. Orav, and L. Goldman, *The relationship between physicians' qualifications and experience and the adequacy of prenatal care and low birthweight*. *American journal of public health*, 1995. **85**(8_Pt_1): p. 1087-1091.
79. Heck, D.A., et al., *Patient outcomes after knee replacement*. *Clinical orthopaedics and related research*, 1998. **356**: p. 93-110.
80. Kelly, J.V. and F.J. Hellinger, *Physician and hospital factors associated with mortality of surgical patients*. *Medical care*, 1986: p. 785-800.
81. Kelly, J.V. and F.J. Hellinger, *Heart disease and hospital deaths: an empirical study*. *Health Services Research*, 1987. **22**(3): p. 369.
82. Morrison, J. and P. Wickersham, *Physicians disciplined by a state medical board*. *JAMA: the journal of the American Medical Association*, 1998. **279**(23): p. 1889-1893.
83. Nelsen Jr, D.A., et al., *The use of new technologies by rural family physicians*. *The Journal of family practice*, 1994. **38**(5): p. 479.

84. Pearce, W.H., et al., *The importance of surgeon volume and training in outcomes for vascular surgical procedures*. Journal of vascular surgery, 1999. **29**(5): p. 768-778.
85. Ramsey, P.G., et al., *Predictive validity of certification by the American Board of Internal Medicine*. Annals of Internal Medicine, 1989. **110**(9): p. 719-719.
86. Schwartz, W. and D. Mendelson, *Physicians Who Have Lost Their Malpractice Insurance. Their Demographic Characteristics and the Surplus-Lines Companies That Insure Them*. Survey of Anesthesiology, 1990. **34**(2): p. 111-112.
87. Sloan, F., et al., *Medical Malpractice Experience of Physicians: Predictable or Haphazard?* Survey of Anesthesiology, 1990. **34**(4): p. 262-263.
88. Walshe, K., *Medical revalidation in England: developing a research agenda*, 2013, Manchester Business School: Unpublished Manuscript.
89. Greenhalgh, T. and G. Wong, *Revalidation: a critical perspective*. The British Journal of General Practice, 2011. **61**(584): p. 166.
90. American Board of American Specialties. *Facts about the 2008 ABMS consumer study: lifelong learning and other qualities in choosing a doctor*. 2008; Available from: www.abms.org.
91. American Board of American Specialties. *Facts about the 2010 ABMS consumer study: lifelong learning and other qualities in choosing a doctor*. 2010; Available from: www.abms.org.
92. Brennan T.A., et al., *The role of physician specialty board certification status in the quality movement*. JAMA, 2004. **292**(9): p. 1038-1043.

93. Freed, G.L., K.M. Dunham, and L.A. Althouse, *Characteristics of general and subspecialty pediatricians who choose not to recertify*. *Pediatrics*, 2008. **121**(4): p. 711-717.
94. Freed, G.L., et al., *Perspectives and preferences among the general public regarding physician selection and board certification*. *The Journal of pediatrics*, 2010. **156**(5): p. 841-845. e1.
95. Anson, J., *MOCA saves a life: letters to the editor*. *ASA Newsletter*, 2013(77).
96. Freed, G.L., et al., *Characteristics of general and subspecialty pediatricians who choose not to recertify*. *Pediatrics*, 2008. **121**(4): p. 711-717.
97. Holmboe, E.S. and C. Cassel, *Continuing medical education and maintenance of certification: essential links*. *The Permanente Journal*, 2007. **11**(4): p. 71.
98. Holmboe ES, L.R., Greiner A, *Assessing quality of care: Knowledge matters*. *JAMA*, 2008. **299**(3): p. 338-340.
99. Levinson W, H.E., *Maintenance of certification in internal medicine: Facts and misconceptions*. *Archives of Internal Medicine*, 2011. **171**(2): p. 174-176.
100. Lipner, R.S., et al., *Who is maintaining certification in internal medicine—and why? A national survey 10 Years after initial certification*. *Annals of Internal Medicine*, 2006. **144**(1): p. 29-36.
101. Miles, P.V., *The future of maintenance of certification: A reaction to the paper by Kevin B. Weiss, MD*. *The Journal of the American Board of Family Medicine*, 2010. **23**(Supplement): p. S42-S45.

102. Puffer, J.C., et al., *Engagement of family physicians in maintenance of certification remains high*. The Journal of the American Board of Family Medicine, 2012. **25**(6): p. 761-762.
103. Xierali, I.M., et al., *Family physician participation in maintenance of certification*. The Annals of Family Medicine, 2011. **9**(3): p. 203-210.
104. Drazen, J.M. and D.F. Weinstein, *Considering recertification*. New England Journal of Medicine, 2010. **362**(10): p. 946-947.
105. Kritek, P.A. and J.M. Drazen, *American Board of Internal Medicine Maintenance of Certification Program — Polling Results*. New England Journal of Medicine, 2010. **362**(15): p. e54.
106. Davis, D., et al., *The case for knowledge translation: shortening the journey from evidence to effect*. BMJ, 2003. **327**(7405): p. 33-35.
107. Oyler, J., et al., *Teaching internal medicine residents quality improvement techniques using the ABIM's practice improvement modules*. Journal of General Internal Medicine, 2008. **23**(7): p. 927-30.
108. Sheehan, J., et al., *Identification of Knowledge Gaps in Neurosurgery Through Analysis of Responses to the Self-Assessment in Neurological Surgery (SANS)*. World Neurosurgery, (0).
109. Steadman, R.H. and Y.M. Huang, *Simulation for quality assurance in training, credentialing and maintenance of certification*. Best Practice & Research Clinical Anaesthesiology, 2012. **26**(1): p. 3-15.

110. Thomson, et al., *Continuing education meetings and workshops: effects on professional practice and health care outcomes*. Database Systematic Reviews, 2001.
111. Colthart, I., et al., *What do doctors really think about the relevance and impact of GP appraisal 3 years on? A survey of Scottish GPs*. The British Journal of General Practice, 2008. **58**(547): p. 82.
112. Wakeling, J., N. Cameron, and M. Lough, *What factors impact on general practitioners' engagement with appraisal? A qualitative evaluation in Scotland*. Education for Primary Care, 2008. **19**(6): p. 615-623.
113. Zolle, O., M. Rickenbach, and E. Gorrod, *The formative impact of general practice appraisals: a post-appraisal questionnaire survey*. Education for Primary Care, 2009. **20**(6): p. 445-454.
114. Finlay, K. and S. McLaren, *Does appraisal enhance learning, improve practice and encourage continuing professional development? A survey of general practitioners' experiences of appraisal*. Quality in Primary Care, 2009. **17**(6): p. 387-395.
115. Nayar, V., *A qualitative study of general practitioners' views of the appraisal process*. Education for Primary Care, 2005. **16**(6): p. 672.
116. Conlon, M., et al., *Using appraisal as the vehicle for revalidation: squeezing a quart into a pint pot?* Education for primary care: an official publication of the Association of Course Organisers, National Association of GP Tutors, World Organisation of Family Doctors, 2009. **20**(2): p. 77.
117. Rughani, A., *GP appraisal and revalidation based on the personal development plan*. The Journal of Clinical Governance, 2001. **9**(4): p. 175-179.

118. Lewis, M. and K. Evans, *Quality assurance of GP appraisal: a two-year study*. Education for Primary Care, 2006. **17**(4): p. 319-333.
119. Conlon, M., et al., *Appraisal: experiences, attitudes and impact. An evaluation of the appraisal process for general practitioners in England*. Clinician in Management, 2006. **14**(1): p. 5-22.
120. Williams, S., K. Evans, and P. Pownceby, *Appraisal as a vehicle for quality improvement: a review of the literature on the impact of appraisal*. 2008.
121. Chamberlain, J.M., *Portfolio-based performance appraisal for doctors: A case of paperwork compliance*. Sociological research online, 2010. **15**(1): p. 8.
122. Chamberlain, J.M., *The Sociology of Medical Regulation: An Introduction* 2013, New York and Amsterdam: Springer.
123. Lynn, N. and S.J. Lea, *Civil Disputes and Crime Recording Refusals, Disinterest And Power In Police Witcraft*. British Journal of Criminology, 2012. **52**(2): p. 361-380.
124. Rosenthal, M., *The incompetent doctor: behind closed doors*. 1995, Buckingham: Open University Press.
125. Edwards, D., *Script Formulations An Analysis of Event Descriptions in Conversation*. Journal of language and Social Psychology, 1994. **13**(3): p. 211-247.
126. Kitchenham, B. and S. Charters, *Guidelines for performing Systematic Literature Reviews in Software Engineering: EBSE Technical Report*, 2007, University of Keele & University of Durham.

127. Centre for Reviews & Dissemination. *Systematic Reviews: CRDs guidance for undertaking reviews in health care*. 2009; Available from: http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf.
128. Moher, D., et al., *Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement*. *Annals of Internal Medicine*, 2009. **151**(4): p. 264-269.
129. Nath, V., *Revalidation: The early experiences and views of responsible officers from London*, 2013: London.
130. NIHR. *PPI - Information for members of the public*. 2013; Available from: http://www.crncc.nihr.ac.uk/about_us/ccrn/cdtv/PPI/PPI_Public.
131. Myers, M., *Qualitative research and the generalizability question: Standing firm with proteus*. *The Qualitative Report*, 2000. **4**(3/4).
132. Boud, D., R. Keogh, and D. Walker, *Reflection: turning experience into learning*. 1985, London: Kogan Page.
133. Coleman, T., *Using video-recorded consultations for research in primary care: advantages and limitations*. *Family Practice*, 2000. **17**(5): p. 422-427.
134. Themessl-Huber, M., et al., *Audio-visual recording of patient–GP consultations for research purposes: A literature review on recruiting rates and strategies*. *Patient Education and Counseling*, 2008. **71**(2): p. 157-168.
135. Martin, E. and P. Martin, *The reactions of patients to a video camera in the consulting room*. *The Journal of the Royal College of General Practitioners*, 1984. **34**(268): p. 607.

136. Pringle, M. and C. Stewart-Evans, *Does awareness of being video recorded affect doctors' consultation behaviour?* The British Journal of General Practice, 1990. **40**(340): p. 455.
137. General Medical Council. *Equality and diversity strategy 2010-2013*. 2013; Available from: http://www.gmc-uk.org/ED_Strategy_2010_2013_36700815.pdf.
138. Baker, S. *How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research*. 2012; Available from: http://eprints.ncrm.ac.uk/2273/4/how_many_interviews.pdf.
139. Rimmer, A. *Patients to be surveyed on their involvement in revalidation*. BMJ Careers 2013; Available from: http://careers.bmj.com/careers/advice/view-article.html?id=20015486&utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+careers%2Fnews+%28Latest+BMJ+Careers+news%29.
140. Florin, D. and J. Dixon, *Public involvement in health care*. BMJ: British Medical Journal, 2004. **328**(7432): p. 159.
141. Dale D, *Recertification in Internal Medicine - The American Experience*. Ann Acad Med Singapore, 2007. **36**: p. 894-7.
142. Dauphinee D, *Revalidation of doctors in Canada*. BMJ, 1999. **319**(7218): p. 1188-1190.
143. Newble D, Paget N, and McLaren B, *Revalidation in Australia and New Zealand: approach of the Royal Australasian College of Physicians*. BMJ, 1999. **319**: p. 1185-8.
144. Lynch M, *Commentary on Shepherd and Cameron and continuation of the revalidation debate*. Journal of Evaluation and Clinical Practice, 2010. **16**: p. 651-654.

145. Mann, K., J. Gordon, and A. MacLeod, *Reflection and reflective practice in health professions education: a systematic review*. *Advances in Health Sciences Education*, 2009. **14**(4): p. 595-621.
146. Mamede, S. and H.G. Schmidt, *The structure of reflective practice in medicine*. *Medical Education*, 2004. **38**(12): p. 1302-1308.
147. AAMC, *Assessment of professionalism project*, 2013, Association of American Medical Colleges.
148. Wilkinson, T.J., W.B. Wade, and L.D. Knock, *A blueprint to assess professionalism: results of a systematic review*. *Academic Medicine*, 2009. **84**(5): p. 551-558.
149. Papadakis, M.A., et al., *Disciplinary action by medical boards and prior behavior in medical school*. *New England Journal of Medicine*, 2005. **353**(25): p. 2673-2682.
150. Veloski, J.J., et al., *Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002*. *Academic Medicine*, 2005. **80**(4): p. 366-370.
151. Humphrey, H.J., et al., *Promoting an environment of professionalism: the University of Chicago "Roadmap"*. *Academic Medicine*, 2007. **82**(11): p. 1098-1107.
152. Hall, M.A., et al., *Trust in physicians and medical institutions: what is it, can it be measured, and does it matter?* *Milbank Quarterly*, 2001. **79**(4): p. 613-639.
153. Silber, C.G., et al., *Do global rating forms enable program directors to assess the ACGME competencies?* *Academic Medicine*, 2004. **79**(6): p. 549-556.

154. De Haes, J., et al., *Assessment of medical students' communicative behaviour and attitudes: Estimating the reliability of the use of the Amsterdam attitudes and communication scale through generalisability coefficients*. Patient Education and Counseling, 2001. **45**(1): p. 35-42.
155. Blackall, G.F., et al., *Professionalism in medical education: The development and validation of a survey instrument to assess attitudes toward professionalism*. Medical teacher, 2007. **29**(2-3): p. e58-e62.
156. Brown, C. and R. Lilford, *The stepped wedge trial design: a systematic review*. BMC Medical Research Methodology, 2006. **6**(1): p. 54.
157. London Deanery. *The NHS appraisal scheme for doctors*. 2013; Available from: <http://www.faculty.londondeanery.ac.uk/e-learning/appraisal/the-nhs-appraisal-scheme-for-doctors>.
158. Archer, J., et al. *Understanding the rise in fitness to practise complaints from members of the public*. 2013.
159. General Medical Council, *2011 Annual Statistics: Fitness to Practise*, 2011, General Medical Council: London.
160. Singh, R. and K. Mukhopadhyay, *Survival analysis in clinical trials: Basics and must know areas*. Statistics, 2011. **2**(4): p. 145-148.
161. Ramsay, C.R., et al., *Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies*. International journal of technology assessment in health care, 2003. **19**(04): p. 613-623.

162. Wagner, A., et al., *Segmented regression analysis of interrupted time series studies in medication use research*. Journal of clinical pharmacy and therapeutics, 2002. **27**(4): p. 299-309.
163. Zhang, S., J. Cao, and C. Ahn, *Sample size calculation for studies comparing binary outcomes using historical controls*. Biometrical Journal, 2013. **55**(2): p. 190-202.

Appendix 1: Interview Schedule

Section 1: Background & understanding of revalidation

1. Tell me a little bit about your background and your links with revalidation?
2. What do you think revalidation is for?
3. How do you think revalidation might change or shape healthcare?

Section 2: Evaluation of revalidation & potential measures

4. If you were to evaluate revalidation, as a regulatory mechanism to ensure proper standards in the practice of medicine, what would you measure (using words as well as numbers)?
5. What measures do you know already exist that could be used to evaluate revalidation's impact on ensuring proper standards in the practice of medicine?
6. What measures do you think could be or should be developed to evaluate revalidation?
7. Are you aware of any historical datasets that might be helpful in understanding the impact of revalidation?
8. The GMC designed revalidation to drive the following. How would you evaluate each of these:
 - (a) Bringing all doctors into a governed system that evaluates their fitness to practise on a regular basis
 - (b) Requiring doctors to collect and reflect on evidence about their whole practice through appraisal
 - i. Continuing professional development
 - ii. Quality improvement activity
 - iii. Significant events
 - iv. Feedback from colleagues
 - v. Feedback from patients
 - vi. Review of complaints and compliments
 - (c) Focusing doctors on *Good Medical Practice* to promote professionalism by increasing awareness and adoption of its values and principles
 - (d) Facilitate identifying and addressing potential concerns earlier – and before they become safety issues or fitness to practise referrals
 - (e) Supporting ROs to fulfil their statutory function of advising the GMC about the fitness to practise of their doctors.
9. Revalidation is likely, amongst other things, to be a driver for change in two key areas; the supporting documentation doctors bring to appraisal and appraisal itself. What do you think? How might this be best evaluated?
10. ROs make important decisions in the revalidation process and which relates to fitness to practise procedures. How might these decisions be evaluated?
11. Specifically we are interested in evaluating the heart of how revalidation might bring about potential performance change in healthcare to ensure proper standards in the practice of medicine. We see this as potentially happening through appraisal. We are interested in capturing appraisal activity either through recording appraisals (to see what happens first hand) and or reviewing appraisal documentation, both what goes in as well as what comes out. Do you think this might help evaluate revalidation's impact? Is it feasible? How might it be supported?

12. How do you think the revalidation process might differ for different types of doctors including
 - (a) Doctors in different settings
 - (b) Minority ethnic groups

Section 3: Prioritising ideas & evaluation timeframe

13. Taking your main ideas, what would you prioritise?
14. If we are interested in demonstrating change over time and revalidation's impact in terms of outcomes how long do you think an evaluation would need to run for? How long would each of your ideas need to happen for do you think?

Appendix 2: Ethics approval letter



13 May 2013

CONFIDENTIAL

Dr Julian Archer
Plymouth University Peninsula Schools of Medicine & Dentistry
Plymouth University
C408 Portland Square

Dear Julian

Application for Approval by Faculty Research Ethics Committee

Reference Number: 12/13-122

Application Title: *Evaluating the strategic impact of medical revalidation*

I am pleased to inform you that the Committee has granted approval to you to conduct this research.

Please note that this approval is for three years, after which you will be required to seek extension of existing approval.

Please note that should any MAJOR changes to your research design occur which effect the ethics of procedures involved you must inform the Committee. Please contact Claire Butcher on (01752) 585337 or by email claire.butcher@plymouth.ac.uk

Yours sincerely

Professor Michael Sheppard, PhD, AcSS
Chair, Research Ethics Committee -
Faculty of Health, Education & Society and
Peninsula Schools of Medicine & Dentistry

Faculty of Health, Education and Society
Plymouth University
Drake Circus
Plymouth PL4 8AA

T +44 (0)1752 585337
F +44 (0)1752 585328
E claire.butcher@plymouth.ac.uk
W www.plymouth.ac.uk

Professor Michael Sheppard
CQSW BSc MA PhD AcSS
Chair, Faculty Research Ethics
Committee

Appendix 3: Ethics approval letter – Focus Group Amendment



21 June 2013

CONFIDENTIAL

Dr Julian Archer
Plymouth University Peninsula Schools of Medicine & Dentistry
Plymouth University
C408 Portland Square

Dear Julian

Amendment to Approved Application

Amendment Reference Number: 12/13-153

Original application Reference Number: 12/13-122

Application Title: Evaluating the strategic impact of medical revalidation

I am pleased to inform you that the Committee has granted approval to you for your amendment to the application approved on 13 May 2013.

Please note that this approval is for three years, after which you will be required to seek extension of existing approval.

Please note that should any MAJOR changes to your research design occur which effect the ethics of procedures involved you must inform the Committee. Please contact Claire Butcher on (01752) 585337 or by email claire.butcher@plymouth.ac.uk

Yours sincerely

Professor Michael Sheppard, PhD, AcSS
Chair, Research Ethics Committee -
Faculty of Health, Education & Society and
Peninsula Schools of Medicine & Dentistry

Faculty of Health, Education and Society
Plymouth University
Drake Circus
Plymouth PL4 8AA

T +44 (0)1752 585337
F +44 (0)1752 585328
E claire.butcher@plymouth.ac.uk
W www.plymouth.ac.uk

Professor Michael Sheppard
CQSW BSc MA PhD AcSS
Chair, Faculty Research Ethics
Committee

Appendix 4: Ethics application

PLYMOUTH UNIVERSITY
FACULTY OF HEALTH, EDUCATION AND SOCIETY

• Application
No:

(for FREC use)

**RESEARCH
WITH
PLYMOUTH
UNIVERSITY**

**Faculty Research Ethics Committee
APPLICATION FOR ETHICAL APPROVAL OF RESEARCH**

Title of research: Evaluating the strategic impact of medical revalidation

1. Nature of approval sought (Please tick relevant box)

(a) PROJECT*: (b) PROGRAMME*:

If (a) then please indicate which category:

- Funded research project
- MPhil/PhD project
- Other (please specify):

Y

**Note: In most cases, approval should be sought individually for each project. Programme approval is granted for research which comprises an ongoing set of studies or investigations utilising the same methods and methodology and where the precise number and timing of such studies cannot be specified in advance. Such approval is normally appropriate only for ongoing, and typically unfunded, scholarly research activity.*

2. Investigators/Supervisors

Principal Investigator:

Name: Dr Julian Archer

Address for correspondence:

Dr Julian Archer

Director of the Collaboration for the Advancement of Medical Education Research & Assessment (CAMERA), Plymouth University Peninsula Schools of Medicine & Dentistry, C408 Portland Square, Plymouth University, Plymouth, PL4 8AA

Tel No + 44 (0)1752 586750

Email: Julian.archer@pms.ac.uk

Other staff investigators:

Dr Sam Regan de Bere, Dr Suzanne Nunn, Dr Lee Coombes, Dr Marie Bryce

Please indicate Department of each named individual, including collaborators external to the Faculty:

Co-applicant:

Dr Sam Regan de Bere, Collaboration for the Advancement of Medical Education Research & Assessment (CAMERA), Plymouth University Peninsula Schools of Medicine and Dentistry (PU PSMD)

Research Fellow: Dr Nicola Brennan, CAMERA, PU PSMD

Research Fellow: Dr Suzanne Nunn, CAMERA, PU PSMD

Research assistant: Dr Marie Bryce, CAMERA, PU PSMD

Psychometric/statistical support: Dr Lee Coombes, CAMERA, PU PSMD

**Note: Principal investigators are responsible for ensuring that all staff employed on projects (including research assistants, technicians and clerical staff) act in accordance with the University's ethical principles, the design of the research described in this proposal and any conditions attached to its approval.*

3. Funding body (if any) and duration of project/programme with dates*:

General Medical Council 'Invitation to tender'

March 2013 – 30th September 2013

**Approval is granted for the duration of projects or for a maximum of three years in the case of programmes. Further approval is necessary for any extension of programmes.*

4. Research Outline:

Background

The approach to improving patient care by encouraging individual doctors to maintain their clinical knowledge, skills and professional behaviours through on-going assessment is growing in popularity across many countries. In the US, maintenance of certification (MOC) sponsored by the American Board of Medical Specialties (ABMS) and its 24 member boards, requires most certified specialists to seek recertification periodically (mostly every 10 years).⁶⁰

In the UK, revalidation was launched on 3rd December 2012. Similar to the MOC, revalidation asks doctors to provide documentation across four domains; knowledge skills and performance, safety and quality, communication partnership and teamwork, and

maintaining trust.¹³ Documentation focuses on continuing professional development (CPD), quality improvement activities such as clinical audit, significant events analysis, feedback from colleagues and patients, as well as complaints and compliments.¹²

Revalidation is unique in a number of important ways. Firstly, it is the only system in the world where all doctors must successfully take part in order to retain their licence to practise. This includes all specialists, general practitioners and all doctors in training; although in the case of trainees this is linked to pre-existing training programmes. Secondly, revalidation has been informed by the Royal Colleges, the UK's professional medical bodies, but has been implemented by the UK's national regulator the General Medical Council (GMC). Thirdly, decision making is initially devolved locally, structured around the NHS appraisal process combined with information from the clinical governance systems in each organisation. Documentation feeds into annual appraisals when, every five years, the responsible officer (RO, normally the most senior doctor regionally), makes the recommendation to the GMC as to whether a doctor should be revalidated or not. Whilst the system for revalidation is operational, little is known about the role of professional regulation in terms of its actual impact on practice. While MOC and relicensure are well established in the US^{141, 142} - and to a lesser extent in Canada and Australasia¹⁴³, there has been a dearth of focused research that evaluates its introduction, or its impact on professional practice, quality improvement and patient safety. With the introduction of revalidation in the UK comes a unique opportunity to explore its implementation, in terms of best practice and deeper analysis to provide new insights into revalidation as a regulatory lever and the development of a new professionalism.¹⁴⁴

Research into the impact of revalidation as a lever for change requires investigation, for as well as desired impacts or consequences there can also be unintended consequences. Unintended consequences are significant because over time they can become significant drivers, and not just products, of intervention (p.8).¹⁵ Identifying both positive and more challenging unintended consequences may provide the opportunity to develop evidence-based gateways to better practice.

Aim:

- To develop a framework for the future evaluation of the impacts of revalidation, whether positive or negative. This will be used as the basis of a future, GMC commissioned, longitudinal study to monitor impact

Objectives:

- The framework should enable the GMC to identify how far revalidation is meeting its regulatory objectives of promoting proper standards of medical practise in the UK
- The framework should identify the ways in which revalidation supports change for improvement

Research Design

Conceptualising what revalidation is about and how it might achieve its declared aims is central to any evaluation. Drawing on policy impact and social discourse theories, we propose a framework that explores revalidation as a potential unifying system, which brings together two inter-related but historically parallel systems of appraisal and clinical governance.

Methods to develop and establish a new evaluative framework

In order to establish an evaluative framework we will need to take a multi-methodological approach to reviewing the available or potential data that could develop meaning for the components and their interactions. The focus of this tender response is to identify possible data sources, existing or with potential for development, and establish the various methodologies and methods that will be required to test their applicability within revalidation as a system for change.

To achieve this, we propose to undertake three key approaches and triangulate the data to develop the framework; first a brief selective literature review, second a review of the current available published data and lastly interviews or focus groups with major stakeholders. We are seeking ethical approval for the third stage.

The views of the majority of stakeholders will be captured through interview however where a number of people in different departments of the one organisation need to be interviewed a focus group (group interview) will be used.

Stage 3 - Thematic analysis of stakeholder interviews/focus groups

Data from the available literature and data review will be triangulated with data from a series of interviews or focus groups undertaken with the major stakeholders. These include patients and the public; doctors in a variety of different practice settings, such as locums, the four UK Departments of Health, employers/contractors; and medical royal colleges, faculties or specialty associations and other health regulators, such as the Care Quality Commission and Monitor. We will use our existing network of connections in the field through our previous research into revalidation in policy (where we interviewed leading members of the medical and legal professions including royal colleges, the four national health departments and employers), and through a current GMC tender exploring the FTP procedures (interviewing patient groups and others) to establish a purposeful data sample of approximately twenty interviewees. We will seek to minimise the impact of interviews or focus groups on the participants by meeting them at their convenience.

Our expectation is that the available measures and stakeholder views and expertise will shape the framework, the theoretically underpinning and future data collection. Drawing on improvement science methodologies we will 1) involve user stakeholders to evaluate current information and understanding, 2) clarify the intentions of revalidation as an intervention, 3) begin to explore the *reality* of revalidation as an intervention as it is implemented, 4) explore any needed changes in revalidation activities, 5) explore alternative evaluation designs, and 6) agree on evaluation priorities and intended uses of evaluative information. This process is often cyclical returning to stakeholders with areas of disagreement between them and between them and the literature however this will be limited in the short timescales offered within the funding.

By triangulating the data from the three methods, we seek to identify the best quantitative and qualitative measures (*formative* and *summative* measures) that could then be used to feasibly evaluate revalidation as a lever for change.

The triangulation of the data will be presented within an overall framework of Cultural Historical Activity Theory (CHAT). CHAT provides an accessible and flexible framework with which to identify and examine interactions, as well as components of, a system as it is implemented. Importantly it, 1) includes the cultural and historical aspects in which systems

are working that will allow us to explore how the systems are currently shaped and understood and 2) CHAT theorises that people continually shape, and are shaped by, their social and cultural contexts and this will be important in terms of demonstrating cultural shifts (for example, revalidation shaping day to day practice for doctors) as well as structural changes. CHAT will therefore enable us to frame revalidation in the wider system of healthcare and explore how revalidation might impact on the culture of healthcare provision by supporting change for improvement in the future. The CHAT framework will be used to triangulate and synthesise the data as well as guide the data collection.

5. Where you are providing information sheets for participants please INSERT a copy here. The information should usually include, in lay language, the nature and purpose of the research and participants right to withdraw:

Qualitative research will be conducted with various stakeholder groups, including: The Patients Association and other patient interest groups, the four UK Departments of Health, regulatory bodies, the Medical Royal Colleges and doctors in a variety of practice settings across all four UK countries. The sampling strategy for this research is purposeful in nature. Through our previous research focusing on revalidation both in policy and practice we have developed a considerable list of relevant contacts. We will also identify key personnel through the literature. Having identified individuals we will contact them directly via email in partnership with the GMC.

Human participants are only involved at the interview/focus group stage. Researchers will arrange a time and location convenient to the participant and travel as necessary to conduct interviews in surroundings that the participant finds conducive. Informed consent will be confirmed. Interviews will focus on the viewpoints and perspectives of the participants rather than personal characteristics.

Interviews will last approximately 1 hour.

Flexible questioning will allow participants to frame their own narratives on a number of issues presented to them.

We are interviewing participants in their capacity as spokespeople for their organisation, for example the likely representative of a patient advocacy group will be their salaried complaints manager. The research does not touch on controversial issues; none of our interviewees are vulnerable people. We therefore do not intend to offer anonymity.

6. Ethical Protocol:

Please indicate how you will ensure this research conforms with each clause of Plymouth University's *Principles for Research Involving Human Participants*. Please attach a statement which addresses each of the ethical principles set out below. Please note: you may provide the degree of detail required. Each section will expand to accommodate this information.

(a) Informed consent:

Please indicate if a consent form is to be used. YES

Participants will be required to sign a consent form.

(b) Openness and honesty:

The team members are experienced researchers. They are fully aware of University policy relating to honesty and openness and used to adhering to University and, where stipulated, guidelines published by funders.

Note that deception is permissible only where it can be shown that all three conditions specified in Section 2 of Plymouth University's Ethical Principles have been made in full. Proposers are required to provide a detailed justification and to supply the names of two independent assessors whom the Sub-Committee can approach for advice.

(c) **Right to withdraw:**

Participants' right to withdraw is explained in the participant information and is included for agreement in the consent form.

(d) **Protection from harm:** N/A

Indicate here any vulnerability which may be present because:

- *of the participants (they may be children or have mental health issues)*
- *of the nature of the research process. Indicate how you shall ensure their protection from harm.*

Please note - researchers contacting children as an aspect of their research must be subject to CRB checks. These can be arranged through Human Resources.

Does this research involve:

Please tick

<i>Vulnerable groups</i>	
<i>Sensitive topics</i>	
<i>Permission of a gatekeeper for initial access</i>	
<i>Subjects being academically assessed by the researcher</i>	
<i>Deception or research which is conducted without full and informed consent</i>	
<i>Research that will induce psychological stress, anxiety or humiliation or cause minimal pain</i>	
<i>Intrusive intervention (e.g., the administration of drugs, vigorous physical exercise or hypnotherapy)</i>	

(e) **Debriefing:** Participants will be offered the opportunity to review their interview transcript to check for accuracy. Participants may also request the deletion of comments but may not alter the content of the transcript.

(f) **Confidentiality:** Digital audio recordings of interviews will be sent to an external transcriber using an encrypted memory stick. The transcriber will be bound by a confidentiality agreement. Hard copies will be kept in a secure cabinet and locked at all times. Electronic data is stored on a shared hard drive on University servers these are encrypted and password protected.

Personal data collected is in the form of names and occupational details only.

All data will be destroyed after 10 years in line with University policy.

Participants are aware that should the results of the research be published in a peer reviewed journal any quotations used will be attributed to them and their organisation will

also be named.

- (g) **Professional bodies whose ethical policies apply to this research:**
General Medical Council

7. Researchers Safety

- (a) **Are there any special considerations in relation to researchers' safety?**
No

- (b) **If so what provision has been made (*for example the provision of a mobile phone, or a clear recording of movements*)**

8. Declaration:

To the best of our knowledge and belief, this research conforms to the ethical principles laid down by Plymouth University and by the professional body specified in 6 (g).

Principal Investigator:

Dr Julian Archer



Signature

Date 21/06/2013

Appendix 5: Information sheet for participants



Evaluating the strategic impact of medical revalidation

Information for participants

[v3 21st of June 2013]

Thank you for showing an interest in this research project. Please read this information sheet carefully before deciding whether or not to participate. If you decide to participate we thank you. If you decide not to take part there will be no disadvantage to you of any kind and we thank you for considering our request.

What is the aim of the project?

To develop a framework for the future evaluation of the impacts of revalidation, whether positive or negative. This will be used as the basis of a future, GMC commissioned, longitudinal study to monitor impact

What type of participants are needed?

We have contacted you personally along with others as we are interested in speaking to people who represent stakeholders involved in revalidation. We are looking to recruit participants from a broad spectrum including: The Patients Association and other patient interest groups, the four UK Departments of Health, the Medical Royal Colleges, regulatory bodies and doctors in a variety of practice settings across all four UK countries in order to develop a framework for the future evaluation of the impacts of revalidation.

What will participants be asked to do?

Should you agree to take part in this research, you will be asked to complete and return the accompanying consent form. One of the research team will then contact you to arrange a convenient time and place for them to interview you either face to face or on the telephone, or to attend a focus group.

Time commitment

Approximately 1 hour for interview or 2 hours for a focus group

Can participants change their mind and withdraw from the project?

You may withdraw from participation in the project at any time and without any disadvantage to yourself of any kind. You are not required to give a reason for your decision to withdraw.

What data/information will be collected and what use will be made of it?

This research involves an open-questioning technique where the precise nature of the questions which will be asked have not been determined in advance, but will depend on the way in which the interview/focus group develops. In the event that a line of questioning does evolve in such a way that you feel hesitant or uncomfortable you are reminded of your right to decline to answer any particular question(s) and also that you may withdraw from participation in the research at any time and without any disadvantage to yourself of any kind. Interviewees may also be asked to 'draw' how they understand the process as another way of capturing data.

Individual interviews/focus groups will be recorded and transcribed. Digital audio tapes will be sent to the transcriber using an encrypted memory stick. The transcriber is bound by a confidentiality agreement. Your interview transcript will be combined with those of the other participants and the dataset will be analysed as a whole.

Participants will be provided with a copy of the transcript of their interview with a member of the research team on request in order to check for accuracy and request omissions but not to alter the content. The data collected will be securely stored. Hard copies of data will be kept in a secure cabinet and locked at all times. Electronic data is stored on a shared hard drive on University servers these are encrypted and password protected.

Results of this project may be published in a peer review journal. The data collected will be used as primary research material for a research report *Evaluating the Strategic Impact of Revalidation* to be submitted to the GMC. Participants will not be anonymised in the report or any publications in peer reviewed journals.

Why me?

You have been approached as we are interested in speaking to people who represent stakeholders involved in revalidation.

What if participants have any questions?

If you have any questions about our research, either now or in the future, please feel free to either contact:

<p>Dr Julian Archer</p> <p>Director of the Collaboration for the Advancement of Medical Education Research & Assessment (CAMERA)</p> <p>Tel No: 01752 586750</p> <p>julian.archer@pms.ac.uk</p>	<p>or</p>	<p>Dr Sam Regan de Bere</p> <p>Lead for Medical Humanities</p> <p>CAMERA</p> <p>Tel No: 01752 586777</p> <p>S.Regandebere@plymouth.ac.uk</p>
---	-----------	--

Complaints

If you have any complaints about the way in which this study has been carried out please contact the principle investigator Dr Julian Archer in the first instance, this may be followed by a complaint to the administrator of the Faculty Human Ethics Committee.

This project has been reviewed and approved by the University of Plymouth Faculty of Health, Education & Society Research Ethics Committee

Appendix 7: Invitation letter to participate in research



GMC Logo

08 October 2013

Dear _____,

We are writing to you regarding research into the General Medical Council's revalidation process for doctors. The aim of the research is to develop a framework for the future evaluation of the impacts of revalidation. This research was commissioned by the GMC and is being undertaken independently by the Collaboration for the Advancement of Medical Education Research and Assessment (CAMERA) at the Plymouth University Peninsula Schools of Medicine and Dentistry.

Due to your expertise in the area we would like to invite you, or a nominated individual from your organisation, to participate in this research. Participation would involve a short interview between yourself and a researcher, and will focus on your views and experiences regarding revalidation. The interview will take place on the phone or one of our researchers would travel to meet you. The interview would last for about an hour. It will be recorded and then transcribed. You will have the opportunity to view the transcription before it is analysed in order to check its accuracy and you may delete but not edit passages as you see fit. We will combine your interview transcript with that of other participants for analysis, but please note participants will be identified by name in the report and any subsequent publications.

You will have the right to withdraw from the research at any time without giving a reason. We would be very grateful if you would consider undertaking this interview with us. We have attached an information sheet and consent form. If you are happy to proceed then please return a signed consent form to:

Dr. Julian Archer, NIHR Career Development Fellow, Clinical Senior Lecturer & Director of the Collaboration for the Advancement of Medical Education Research & Assessment (CAMERA), Plymouth University Peninsula Schools of Medicine & Dentistry, C408 Portland Square, Plymouth University, Drake Circus, Plymouth, PL4 8AA or electronically to julian.archer@pms.ac.uk

Please include the best contact details for you or for your assistant on the consent form as we will then contact you to arrange a convenient time and place to conduct the interview.

101

If you have any questions in the meantime or require more information then please do not hesitate to contact us: julian.archer@pms.ac.uk or 01752 586750. If you would like to speak to someone at the GMC about this research, please contact Divya Patel dpatel@gmc-uk.org or 0207 189 5195.

We look forward to hearing from you in due course,
Yours sincerely



Dr Julian Archer
Plymouth University
Peninsula Schools of Medicine and Dentistry

Jon Billings
General Medical Council

Appendix 8: Completed Interviews and Focus Groups

<p>Interviews Completed (24)</p>	<ol style="list-style-type: none"> 1. NALM Patient Association 21/6/13 2. Health Policy & Management Research 21/6/13 3. Revalidation Implementation Advisory Board (RIAB) & Patient Representative 2/7/13 4. RIAB England 2/7/13 5. Academy of Royal Medical Colleges 3/7/13 6. British International Doctors Association 4/7/13 7. RIAB Northern Ireland 5/7/13 8. Responsible Officer (RO) 10/7/13 9. Appraiser 12/7/13 10. Royal College of Anaesthesia 12/7/13 11. Royal College of Psychiatrists 16/7/13 12. Appraisee 17/7/13 13. RIAB England 17/7/13 14. Wales Deanery 18/7/13 15. BMA 24/7/13 16. Appraiser in Primary Care 24/7/13 17. Independent Sector 24/7/13 18. RO Independent Sector 1/8/13 19. Locum Sector 2/8/13 20. Health Education England 14/8/13 21. Care Quality Commission 14/8/13 22. RIAB Wales 28/8/13 23. NHS England 2/9/13 24. RIAB Scotland 12/9/13
<p>Focus groups completed (2)</p>	<p><u>GMC Focus Group 2nd of July 2013</u></p> <ol style="list-style-type: none"> 1. Assistant Director, Registrations 2. Head of Regional Liaison Service 3. Responsible Officer 4. Assistant Director, Fitness to Practise 5. Strategic Relationships Manager 6. Head of Equality and Diversity 7. Assistant Director, Employer Liaison Service <p><u>Equality & Diversity Focus Group 2nd of July 2013</u></p> <ol style="list-style-type: none"> 1. Progressive Muslims Forum 2. Medical Women's Federation 3. GIRES

Appendix 9: Evaluation methods that were investigated but not included in the framework

As part of developing the evaluation framework, the CAMERA research team investigated the potential utility of a number of evaluation methods which were not ultimately included in the final framework. Each of these methods is outlined here as well as the rationale for excluding them.

(a) Reflection Scale

Rationale: One of the evaluation questions that emerged when developing the framework was whether revalidation would result in a change in doctors' level of reflection. There are a variety of different methods proposed in the literature to measure reflection.¹⁴⁵ A reflection scale could potentially provide a measure of whether the process of collecting the supporting evidence and the appraisal had resulted in a self-perceived change in doctors' levels of reflection.

Method: A reflection scale could be given to the proposed sample of appraisees before they started collecting the supporting information and after their appraisal meeting throughout the revalidation cycle in order to detect a change. There are a number of validated scales that measure personal reflection in medical practice. Mamede and Schmidt¹⁴⁶ developed an instrument to understand the nature of reflection in medical practice. Using an 87-item questionnaire, of which 65 questions were related to reflective practice, they identified a multidimensional, five-factor model of reflective practice. The factors and reliability of each were: deliberate induction ($\alpha = 0.83$); deliberate deduction ($\alpha = 0.81$); testing and synthesizing ($\alpha = 0.79$); openness for reflection ($\alpha = 0.86$); and, meta-reasoning ($\alpha = 0.68$).

Reason for Exclusion from Framework: It was concluded that this measure was not objective as it is based on self-report and thus only measures doctors' perceived level of reflection. There were no published objective instruments with a robust literature for large scale implementation without significant piloting in their own right. In addition the likelihood of demonstrating a statistical significant change across an appraisal cycle or indeed revalidation cycle was considered unlikely due to anticipated ceiling effects. Furthermore, the lack of a control group (as all doctors are participating in revalidation) means that it would be very difficult to establish causality.

(b) Professionalism Scale

Rationale: An important evaluation question that emerged when developing the framework was whether revalidation was resulting in a change in the awareness and adoption of the values and principles of *Good Medical Practice* and thus professionalism as defined by the GMC through this key document. There are a variety of different methods proposed in the literature to measure professionalism.¹⁴⁷ A professionalism scale could potentially provide a quantitative measure of whether the process of revalidation and the appraisal meeting have resulted in a change in doctors' perceived levels of professionalism. To investigate whether professionalism levels have increased might therefore be addressed by giving a professionalism scale to a sample of doctors before and after the appraisal throughout the revalidation cycle to measure if levels of professionalism are increasing.

Method: There are a number of possible professionalism scales¹⁴⁸⁻¹⁵⁰ administered by patients e.g. FACE cards¹⁵¹, Wake Forest Physician Trust Scheme¹⁵², or administered by supervisors/peers e.g. global rating forms¹⁵³, Amsterdam attitudes and communication scale¹⁵⁴ or self-administered e.g. Penn State College of Medicine Professionalism Questionnaire¹⁵⁵. A scale could be completed by doctors twice in the yearly appraisal cycle – before they started collecting their supporting information and after the appraisal. The scale could be given to doctors over a number of years to investigate change over time. Another option could be to analyse the MSF data that doctors collect as part of their appraisal process this would provide an assessment of doctors' professionalism from the patient perspective.

Reason for Exclusion from Framework: Again sadly the professionalism scales would not provide an objective measure as they are based on reported change and only measure levels of professionalism as perceived by doctors or patients.. The possibility of measuring change in MSF data, following an intervention, has already been rejected as part of a national funding application to the NIHR by CAMERA in collaboration on the grounds that the scores are already skewed and measured change would be unlikely to reach statistical significance without unrealistic sampling.

In addition professionalism is a highly contested concept as demonstrated in the stakeholder interviews. As with reflection, the lack of a control group (as all doctors are participating in revalidation) means that it would be very difficult to establish causality.

(c) Stepped Wedge Design Trial

Rationale: An important evaluation question that emerged in the development of the framework was whether the introduction of revalidation had directly resulted in a change in the activities of appraisal and clinical governance, and their components. The use of a stepped-wedge design could potentially help alleviate concerns over causality arising from the analysis of survey data over time (e.g. reflection, professionalism, awareness of *GMP*, supporting evidence collection over annual cycles).

Methods: Stepped wedge randomised trial designs involve sequential roll-out of an intervention to participants (individuals or clusters) over a number of time periods. By the end of the study, all participants have received the intervention, although the order in which participants receive the intervention is determined at random.¹⁵⁶ In this case the intervention is revalidation. The fact that doctors have different revalidation dates could be used to cluster the doctors into groups with various outcome measures.

Reason for Exclusion from Framework: Revalidation is not a 'point in time' intervention. The key mechanism of revalidation is the appraisal meeting which was introduced into the NHS in 2002.¹⁵⁷ Revalidation was of course introduced in December 2013, so all doctors should have already engaged with some aspects of the process, meaning that there is no clear-cut before and after scenario.

In a stepped wedge design the most plausible outcome measure would be referrals to Fitness to Practise, comparing referral rates between doctors who have been revalidated and those yet to go through their first full cycle. Revalidation you might hypothesise should reduce the proportion of doctors being referred. However in 2011; 8,781 enquiries were reported regarding doctors fitness to practise.^{158, 159} As a prospective trial, these numbers would not be sufficient to meet statistical power requirements. Alternative outcome measures such as professionalism and reflection scales are not appropriate as outlined above.

(d) 'Survival' analysis

Rationale: In order to explore whether revalidation has facilitated the identification and addressing of potential concerns earlier and before they become safety issues or fitness to practise referrals the research team explored the possibility of a 'survival' analysis.

Method: Survival analyses involve the modelling of time to event data; in the context of the current study, a doctor's referral is considered the event.¹⁶⁰ Using population data over 5-10 years one could compare the 'survival' (i.e. non-referral) of doctors between a) 2008-2013,

b) time since 2013 to revalidation date and c) time since revalidation (up to next revalidation). Comparing a) and b) may identify if the appraisal process is working e.g. we would expect a constant proportion of doctors referred each year 2008-2013, but a fall in the proportion each year 2013-revalidation date. If revalidation is working, we would expect this proportion to drop to almost zero in the post-revalidation period. In terms of survival, we would therefore hypothesise that $c > b > a$.

Reason for Exclusion from Framework: This method was rejected because of sample sizes. The difficulty arising as there are not enough 'cases' to be able to power the study. Too few doctors are referred to FTP procedures annually compared to the national cohort.

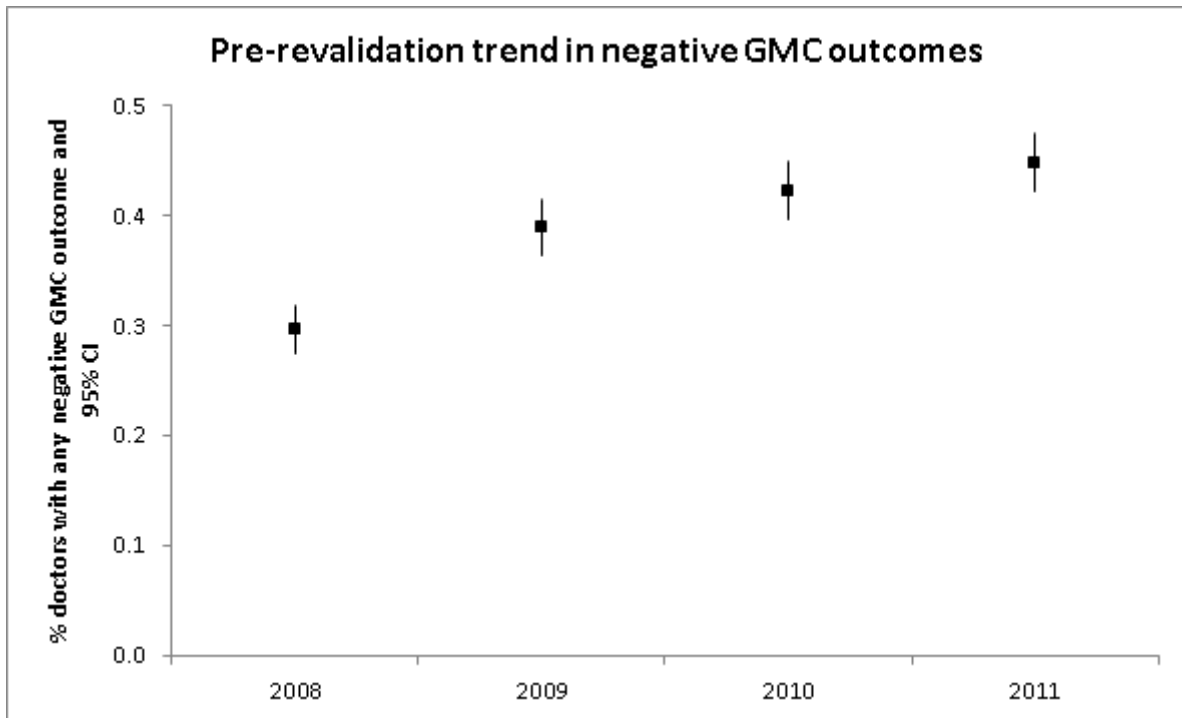
(e) Interrupted Time Series:

Rationale: In order to explore whether revalidation has facilitated the identification and addressing of potential concerns earlier and before they become safety issues or fitness to practise referrals the research team explored an interrupted time series design.

Method: In an interrupted time series (ITS) design, data are collected at multiple instances over time before and after an intervention to detect whether the intervention has an effect significantly greater than the underlying secular trend.^{161, 162} In the current study the introduction of revalidation is the interruption. ITS are strong designs when experimental approaches are not possible.^{111, 112} In an ITS, the rate of GMC sanctions (i.e. any "negative" outcome) for each year can be plotted. Two separate regression lines can then be estimated, one pre- and one post- revalidation (with a lag if thought appropriate) and we can then evaluate whether there is a change in level and/or slope between the two periods. It would be hypothesised that both changes are possible.

Reason for Exclusion from Framework: This method was not included in the framework due to the sample sizes required. The crucial inputs are the number of time points and the expected effect size.¹⁶³ As we would not have that many time points, we would need an effect size of at least 1. Looking at the standard deviation of the proportion of doctors with adverse FTP decisions in 2008-11, this means the change in level and change in rate would need to sum to at least 6%, which would be impossible (Figure 5). This is without controlling for dependency of observations across time points (i.e. mostly the same doctors each year) and an unbalanced number of time points pre- and post- intervention.

Figure 5: Pre-revalidation trend in negative outcomes



(f) Ethnographic Study of Reflection in Practice

Rationale: In order to explore doctors' reflection in practice the CAMERA team discussed the possibility of conducting an ethnographic study.

Method: An ethnographic study observing doctors in their daily practice could be conducted. This would involve researchers either taking field notes or recording (digital or audio) doctors on a typical day in their practice over a period of time and qualitatively analysing the data generated.

Reason for Exclusion from Framework: This would be an extremely expensive study to carry out. In addition, whether one could truly capture doctors' reflection in their daily medical practice is questionable as reflection is a cognitive process and therefore not 'observable'.

Appendix 10: Likely minimum sample sizes for qualitative interviews and audio-recorded appraisals if conducted annually for 3 years to achieve data saturation

	Analysis of Appraisal Forms	Appraisal meeting (audio-recorded)	Interview with Appraisee	Interview with Appraiser	Interview with RO	Interview with ELA
Primary Care	4	8	8	8		
Secondary care	4	8	8	8		
Mental Health consultants	1	2	2	2		
Independent practice doctors	2	4	4	4		
Primary Care Locums	2	4	4	4		
Secondary Care Locums	2	4	4	4		
SAS or trust grade doctors	2	4	4	4		
Clinical Academics	2	4	4	4		
Total per individual centre	19	38	38	38	4	1
Annual total for all 6 centres	114	228	228	228	24	6
Proposed minimum totals across a 3 year study period	342	684	684	684	72	18
Likely minimum total data points across the study: 2142						