Academic support for the Assessment and Appraisal Workstream of Health Education England's review of the ARCP

Final Report

Dr Katherine Woolf and Dr Michael Page

Research Department of Medical Education, UCL Medical School.

Contact: k.woolf@ucl.ac.uk

30th September 2017 amended 4th October 2017

Commissioned by Health Education England

UCL Medical School
RDME
Research Department of Medical Education

Glossary

ACGME	Accreditation Council for Graduate Medical Education
AoMRC	Academy of Medical Royal Colleges
ARCP	Annual Review of Competence Progression
BME	Black and Minority Ethnic
COPMeD	Council of Postgraduate Medical Deans of the United Kingdom
ESR	Educational Supervisor's Report
GMC	General Medical Council
GP	General Practice
GPC	Generic Professional Competencies
HEE	Health Education England
IMG	International Medical Graduate
NHS	National Health Service
PLAB	Professional and Linguistic Assessment Board
SLE	Supervised Learning Event
UK	United Kingdom
UKFPO	United Kingdom Foundation Programme Office

Table of Contents

1. Executive summary 4
2. Introduction 10
   2.1 Purpose of the report 10
   2.2 The ARCP is an assessment 10
   2.3 Using validity to frame the ARCP evidence 10
   2.4 Aim of the report 10
3. Methods 10
4. Results 12
   4.1 ARCP scoping review: description of studies found 12
   4.2 Purposes of the ARCP 13
   4.3 Summative aspects of the ARCP 13
      4.3.1 Assessment of trainee progression and performance 13
      4.3.2 Panel judgements against curricula 14
      4.3.3 Which evidence panels use to make decisions 15
      4.3.4 Ability of the ARCP to detect unsatisfactory performance and patient safety issues 18
      4.3.5 Reliability of the ARCP 21
      4.3.6 Fairness of the ARCP 24
      4.3.7 Impact of the ARCP on trainee learning 25
   4.4 The appraisal aspects of the ARCP 26
      4.4.1 Feedback given to trainees by the panel 27
      4.4.2 Feedback given to trainees to prepare them for the panel 29
   4.5 Revalidation and the ARCP 31
5. Summary of findings and suggestions 33
      5.1.1 How can the summative components of the ARCP process be standardised, and what measures can be taken to ensure that the process is robust, reliable and valid? 33
      5.1.2 How can the appraisal elements of the ARCP process be standardised, ensuring that appropriate formative feedback is provided to all trainees? 35
      5.1.3 How can ARCP processes be aligned to GMC revalidation requirements? 36
      5.1.4 What understanding do trainees and supervisors have of the educational principles that underpin the summative and formative elements of ARCPs? 37
6. Conclusions 37
7. References 39
8. Appendix 44

1. Executive summary

1.1 Background

This report was prepared for Health Education England (HEE) by the Research Department of Medical Education at UCL Medical School. Its purpose is to inform the recommendations of HEE's Annual Review of Competency Progression (ARCP) Review Assessment and Appraisal Workstream in relation to a number of aspects of the ARCP process.

This report aims to answer the following questions set by the Assessment & Appraisal Workstream:

- How can the summative components of the ARCP process be standardised, and what measures can be taken to ensure that the process is robust, reliable and valid?
- How can the appraisal elements of the ARCP process be standardised, ensuring that appropriate formative feedback is provided to all trainees?
- How can ARCP processes be aligned to GMC revalidation requirements?
- What understanding do trainees and supervisors have of the educational principles that underpin the summative and formative elements of ARCPs?

In terms of the summative elements of the ARCP we considered: the criteria against which panels make decisions, which evidence panels use to make decisions, the ability of the ARCP to distinguish between different levels of trainee performance and to identify patient safety concerns, the reliability and fairness of the ARCP, and the impact of the ARCP on trainee learning.

In terms of the appraisal elements of the ARCP we focused on the feedback given to trainees by the panel, and how trainees are prepared for the ARCP panel.

We also considered the relationship between the ARCP and medical revalidation, and trainee and supervisor perceptions of the ARCP.

1.2 Methods

We started with a scoping review of the peer-reviewed and grey literature on the ARCP to identify evidence regarding the validity of the summative and appraisal aspects of ARCP, the ARCP and revalidation, and trainee and supervisor perspectives on the ARCP.

We then reviewed the official national guidance relating to ARCPs, and some other local or specialty-specific guidance that we were aware of or had found via non-systematic Google searches.

We then drew together the literature findings and guidance with evidence from the wider medical education and mainstream education literature, to provide suggestions for increasing the validity, reliability, standardisation, and robustness of the ARCP as a tool to assess and drive trainee progress, to revalidate trainees, and to improve trainee and supervisor understanding of the ARCP.

1.3 Results

The overarching suggestion from this report is that the ARCP panel is recognised as a high-stakes assessment that is likely to have a significant impact on patient care. The ARCP panel should therefore be subject to the same scrutiny and psychometric considerations as other high-stakes assessments, such as medical school finals and postgraduate examinations.

Report findings and suggestions are summarised under headings relating to the research questions. Major findings and suggestions are presented first, followed by a summary of the more detailed findings and suggestions, as appropriate.

1.3.1 How can the summative components of the ARCP process be standardised, and what measures can be taken to ensure that the process is robust, reliable and valid?

1.3.1.1 Major findings and suggestions

Major finding: The guidance lacks consistency and detail about how panels should make decisions. This is a threat to the standardisation, reliability, and validity of the ARCP.

Major suggestion: The guidance to provide more detailed and practicable information about the information panels should use to make decisions, and how panels should go about making decisions in practice (including that panel decisions should be made without the trainee present). Decision-aids to be developed that are applicable nationally, and consistent in quality across specialties.

Major finding: There are concerns that ARCPs only identify very poorly performing trainees and fail to identify other trainees with specific or less serious performance issues. This may result partly from the 'failure to fail' phenomenon.

Major suggestion: Minimise failure to fail by having panels provide constructive feedback for all trainees post-decision; by providing time for panels to discuss trainee performance; by having panels consider multiple sources of evidence not just the Educational Supervisor's Report; by improving support for trainees who are failed; and by introducing lower-stakes formative pre-ARCP reviews.

Major finding: There is a lack of empirical research and published evaluation of the ARCP, and ARCP data provided for research has sometimes been of poor quality. This makes it difficult to assess the quality of the ARCP as an assessment.

Major suggestion: Collect and provide good quality data for rigorous research and evaluation.

1.3.1.2 Summary of more detailed findings and suggestions

Finding: The guidance is unclear about the basis upon which panels should make decisions, leading to inconsistency and threatening validity.

Suggestion: In order to develop decision-aids that are applicable nationally, comparable across specialties, and that help panels make valid and robust decisions, the following needs to be clarified in the guidance:

- Expectations around trainee progression and performance.
- How panels should weight different pieces of information (research is needed to discover what panels are doing currently, and the impact that weighting can have on outcomes).
- How panels should take into account any contextual and environmental factors that affected a trainee's performance, whether positively or negatively.
- The purpose of the additional information sought and the processes to be followed when a trainee is anticipated to receive an unsatisfactory outcome; and how the additional information should be used by panels to improve the validity of their decision-making.
- Expectation that assessments submitted in the portfolio are sampled across the curriculum and assessors.

- Requirement that Educational Supervisors be absent when panels are making decisions. If a trainee is present at panel, this will not influence panel decision-making (i.e. decisions to be made in absentia). Consider having all trainees submit written evidence stating whether they support or disagree with the Educational Supervisor's report to lessen the conflict between the Educational Supervisor's dual roles as mentor and assessor.

Suggestion: To aid the development of decision-aids, it may be helpful to map all UK local and specialty-specific ARCP guidance to draw out the similarities and differences in the competencies required and the information provided about how panels should make judgements. Following this, a Delphi process (or similar) could be performed to achieve consensus about how panels should make decisions in practice. This could be done by specialty or by families of specialities.

Suggestion: Newly-developed decision aids to be compared and reviewed to ensure consistency across specialties and locations, and then piloted to ensure they are practicable.

Finding: There is a lack of evidence about whether the ARCP is able to reliably distinguish between satisfactory and unsatisfactory performance, or between different levels of unsatisfactory performance. Trainees and trainers are sceptical about the ARCP's ability to identify anything other than extreme poor performance, feeling it is not able to identify patient safety concerns but may be overly harsh to trainees with protected characteristics. Medical education research from outside the UK demonstrates reliable judgements about a trainee's competence can be made on a basis of a several workplace based assessments sampled across curriculum areas and assessors.

Suggestion: Collect and provide good quality data for rigorous research and evaluation regarding the validity and reliability of the ARCP. Ensure research is published.

Suggestion: Put in place training and support for supervisors and panels to fail trainees if necessary. This includes: normalising 'remedial' training by providing constructive feedback for all trainees (including stretching feedback for trainees performing well); better information for supervisors, panels, and trainees about support provided to trainees receiving unsatisfactory outcomes, especially for those receiving an Outcome 4; providing sufficient time for panels to discuss each trainee's performance; ensuring panels consider multiple sources of evidence not just the Educational Supervisor's Report; having formative relatively low-stakes pre-ARCPs for all trainees and/or using other tools to help the early identification of problems.

Finding: Panels make more reliable judgements than individuals, but steps need to be taken to guard against problems that can arise from group decision-making.

Suggestion: Panels to agree on their roles and their expectations of trainees at the start (potentially including clearer role expectations and Terms of Reference for panel members); proformas to aid structured decision-making; sufficient time for panel members to discuss each trainee, with panel chairs regularly summarising and ensuring panel members contribute and share as much information as possible. More panel members are recommended (with the proviso that all should be fully engaged) with three being a minimum.

Finding: On average IMGs, male doctors, older doctors, and doctors from black and minority ethnic (BME) backgrounds are more likely to have an unsatisfactory ARCP outcome. It is unclear whether

this reflects panel decision-making or other factors. Unconscious bias training may not be sufficient to combat the potential for unfair bias.

Suggestion: Panels to explicitly state their commitment to ensuring decision-making is fair during the ARCP. Panel membership to be monitored, efforts made to ensure diversity, and monitoring undertaken to check for any unfair bias. Panels to have greater consideration of the impact of training environments or other external factors on trainee progression and performance.

Finding: Many trainees feel demotivated by the fact that it is impossible to achieve an ARCP outcome higher than 'Satisfactory' and see it as discouraging excellence.

Suggestion: Consideration of how the ARCP process can recognise excellence. Constructive feedback for all trainees, including 'stretching' feedback for those performing well.

1.3.2 How can the appraisal elements of the ARCP process be standardised, ensuring that appropriate formative feedback is provided to all trainees?

1.3.2.1 Major findings and suggestions

Major finding: At present, only trainees expected to receive an unsatisfactory outcome are required to talk to the panel to discuss their performance, although the limited literature suggests in practice most trainees receive some kind of feedback from the panel, and many receive this in person. There is little guidance about the format in which feedback should be provided or what feedback should be about (e.g. past performance or future performance). There appears to be considerable variation between specialties and geographic locations, which is likely to hinder the validity of the ARCP.

Major suggestion: Ensure all trainees receive constructive feedback to improve their learning and performance, including trainees who are progressing satisfactorily. Provide guidance to standardise the way in which trainees receive feedback across specialties and locations. Ensure trainees and panels know panel decision-making will not be influenced by a trainee's attendance.

Major finding: There is no national guidance relating to preparing trainees for ARCPs. In practice, it seems different LETB's/Deaneries provide different types of pre-ARCP meetings for different groups of trainees. This lack of standardisation is likely to hinder the validity of the ARCP.

Major suggestion: Ensure all trainees have a pre-ARCP meeting with their Educational Supervisor and another person, possibly an ARCP panel member, to check progress and provide feedback. Provide guidance to standardise the pre-ARCP meeting process.

1.3.2.2 Summary of detailed findings and suggestions

Finding: All trainees are likely to benefit from feedback. Research on appraisal suggests the benefits depend on appraiser and appraisee factors, but it is uncertain how relevant this research is to panels providing feedback. ARCP prioritises summative assessment in the form of the Educational Supervisor Report and assessments of trainees by seniors, but peer- and self-assessment is important for learning.

Suggestion: All trainees to receive constructive feedback post-ARCP panel, including stretching feedback for those performing well. Standardised guidance on feedback to be developed and evaluated. Panel members to be trained to provide feedback. Panels to consider peer and self-assessment as well as assessor by seniors.

Finding: Neither the Gold Guide nor the Foundation Programme Guidance describe in any detail how the trainee should be prepared for the ARCP. The literature provides various ways to prepare trainees for the panel, such as ensuring trainee portfolios meet up-to-date curriculum requirements and having interim/pre-ARCP panels.

Suggestion: Formative interim/pre-ARCPs for all trainees, which focus on providing constructive feedback and feedforward, and on identifying any problems a trainee is having, including any contextual or environmental factors affecting their progression. A degree of externality, possibly from a panel representative, is likely to be beneficial.

Suggestion: Explore the potential for developing a standardised tool (adaptable to local contexts), to help trainees and supervisors track achievement over the course of the year and map these to curricular requirements.

Finding: Lack of standardised support for trainees receiving an Outcome 4, despite many of these doctors continuing to practice medicine. Potentially risky for patient safety. Also likely to increase failure to fail.

Suggestion: Greater educational and career support for trainees receiving an Outcome 4. Trainees, panels, and supervisors to have more information about the support for trainees who are failed in order to combat failure to fail.

1.3.3 How can ARCP processes be aligned to GMC revalidation requirements?

1.3.3.1 Major findings and suggestions

Major finding: Lack of clarity about how revalidation decisions and progression decisions relate to one another. Lack of consistency between revalidation for trainees and non-trainees (e.g. consultants, staff grade doctors). The former is based on a summative ARCP and progression against a curriculum, the latter is based on multiple formative appraisals and considered against Good Medical Practice.

Major suggestion: Clarification in the guidance as to how the processes for determining progression and revalidation relate to one another. Trainees could receive feedback from the ARCP panel that is aligned with the major domains of Good Medical Practice, to help align revalidation during and after training.

1.3.4 What understanding do trainees and supervisors have of the educational principles that underpin the summative and formative elements of ARCPs?

1.3.4.1 Major findings and suggestions

Major finding: The limited literature suggests many trainees feel the an annual review is valuable in principle, but have serious criticisms of the ARCP, perceiving that it does not provide meaningful feedback, that it can discourage excellence, and that it is not sensitive enough to pick up anything but very poor performance.

Major suggestion: Communicate to trainees how the ARCP review aims to:

- **ensure all trainees receive constructive feedback** in preparation for the ARCP and after the ARCP, including feedback to stretch trainees who are already progressing satisfactorily to encourage excellence, and by finding ways for process to recognise excellence; and by providing

interim/pre-ARCPs and developing tools to help trainees and supervisors track a trainee's progression and identify problems early.

- **improve the reliability and validity of the ARCP** (and thereby contribute to patient safety) by standardising panel decision-making across specialties, grades, and locations; by increasing the panel's consideration of environmental and contextual impacts on a trainee's progression; by reducing the impact of a single Educational Supervisor's report on outcomes; by ensuring panels make decisions without the Educational Supervisor or the trainee present; by ensuring rigorous and transparent evaluation of the ARCP.

1.4 Conclusions

We found relatively little published research on the ARCP, and much of the evidence was small-scale. This reflects the fact that, despite its high-stakes nature, the ARCP panel is not officially an assessment and does not receive the same scrutiny as other high-stakes assessments in medical education. Our primary suggestion therefore that is the ARCP panel is officially recognised as a high-stakes assessment likely to have a significant impact on patient care and subject to the same scrutiny as other high-stakes assessments.

The lack of standardisation in ARCP processes is concerning, and we have suggested much greater clarity in the guidance in a number of areas to address this. To prevent the proliferation of local guidance and the concomitant threat to validity, we have suggested guidance is standardised and then piloted to determine feasibility in practice. We have also suggested that the relationship between the ARCP and revalidation, and the appraisal elements of the ARCP, are both more clearly defined and communicated.

The ARCP is generally perceived negatively by trainees, who feel it does not provide meaningful feedback, that excellence is not rewarded, and only the poorest performance is identified. We have suggested a number of changes designed to improve the validity of the ARCP by making it more meaningful. Communicating with trainees about the changes being made to the ARCP and the rationale for those changes is likely to be crucial to restore confidence in the assessment.

Finally, we firmly believe in the principle that 'assessment is not easy to develop and is only as good as the time and energy put into it' (1) (p.707). We recognise that investment in the ARCP is particularly challenging in the current circumstances, but we believe it is worthwhile. As Eva *et al* (2) point out: 'It seems antithetical to the very reasoning behind assessment (the protection of patients) to suggest that we should not think about how to improve current assessment practices, not only in terms of their role in gatekeeping but also in terms of their opportunities for shaping further learning' (p.907). We suggest investment in undertaking proper and continual evaluation of the process and outcomes of the ARCP, including any changes made, is essential to ensure the validity, reliability, robustness, and defensibility of the ARCP and its role in postgraduate training.

2. Introduction

2.1 Purpose of the report

2.2 The ARCP is an assessment

Throughout this report we consider the ARCP to be an assessment with *summative* and *formative* purposes[1]. Broadly speaking, the main summative purpose is to determine whether a trainee can progress to the next state of training or complete training, and the main formative purpose is to provide feedback to guide trainees' learning – this is also called the appraisal element of the ARCP. Other purposes of the ARCP, including the revalidation purpose, are discussed in Section 4.2.

2.3 Using validity to frame the ARCP evidence

We use a validity framework to consider the evidence for best educational practice in relation to ARCPs - validity being the most important attribute of assessment quality. Validity is particularly important since, as Goodyear, Wall and Bindal point out, the ARCP is 'a high-stakes assessment…as it must be passed to go on to the next year of training' (3) (p. 398). A trainee who receives an Outcome 4 at their ARCP has to leave their training programme, so in some ways the ARCP is higher stakes than a postgraduate examination, which can be repeated within limits.

Validity is often misconceived as a purely psychometric or mathematical answer to the question 'does this assessment measure what it is supposed to?' (4) (p.310). In fact, validity also requires an answer the question: 'should this assessment be used for the proposed purpose in the proposed way?'.(5) The latter addresses crucial issues around the impact or consequences of testing, which can be 'beneficial, or harmful … intended or unintended' (6), and in the case of the ARCP, are felt by trainees but also by patients, employers, and supervisors. Validity incorporates psychometric evidence (reliability and correlations with other tests) and philosophical and educational evidence about the assessment's purpose, implementation, fairness, and the impact on teacher and learner motivation and behaviour. Validity also incorporates formative aspects e.g. feedback trainees receive - or don't receive - from the ARCP panel and from supervisors in preparing for the ARCP.(7)

2.4 Aim of the report

This report aims to answer the following questions set by the Assessment & Appraisal Workstream:

- How can the summative components of the ARCP process be standardised, and what measures can be taken to ensure that the process is robust, reliable and valid?
- How can the appraisal elements of the ARCP process be standardised, ensuring that appropriate formative feedback is provided to all trainees?
- How can ARCP processes be aligned to GMC revalidation requirements?
- What understanding do trainees and supervisors have of the educational principles that underpin the summative and formative elements of ARCPs?

3. Methods

We started with a scoping review of the literature on the ARCP to draw out evidence regarding the validity of the summative and appraisal aspects of ARCP, the ARCP and revalidation, and trainee and supervisor perspectives on the ARCP. A scoping review is appropriate to summarise and

[1] Educationally-speaking the summative/formative distinction is not rigid - all assessments are formative in that they provide some kind of feedback that influences learning.

communicate complex and varied evidence about a broad topic, and in particular to 'describe in more detail the findings and range of research in particular areas of study, thereby providing a mechanism for summarising and disseminating research findings to policy makers, practitioners and consumers who might otherwise lack time or resources to undertake such work themselves'(8).

We consider the validity of the evidence presented to the panel, including the Educational Supervisor report, although a full review of the validity of workplace based assessments was outside the scope of the report. In terms of the appraisal elements of the ARCP, we focus on the feedback given to trainees by the panel and how trainees are prepared for the ARCP panel.

We reviewed the following official national guidance relating to ARCPs: Gold Guide (9), Foundation Programme Reference Guide,(10) UK Foundation Programme Office (UKFPO) ARCP Guide,(11) GMC and Academy of Medical Royal Colleges Generic Professional Capabilities Implementation guide,(12) GMC protocol for making revalidation recommendations: Guidance for Responsible Officers and Suitable Persons,(13) and the COPMeD guidance on making revalidation recommendations for doctors in postgraduate training.(14)

We included other resources we were aware of or found via non-systematic Google searches, including the Accreditation Council for Graduate Medical Education (ACGME) Guidance on Clinical Competency Committees (15) and local or specialty-specific UK guidance on ARCPs.

Finally, we drew together the evidence from the literature and the guidance, with evidence from the wider medical education and mainstream education literature, to provide suggestions for increasing the validity (including reliability, standardisation, and robustness) of the ARCP as a tool to assess and drive trainee progress and to revalidate trainees, and to improve trainee and supervisor understanding of the ARCP.

3.1 Literature review methods

The ARCP literature review was conducted by KW, using the following search parameters and inclusion and exclusion criteria agreed by KW and MP in advance:

Search string: "ARCP OR "Annual Review of Competence Progression" OR "Annual Review of Competency Progression".

Sources searched: Medline and PubMed databases; journals likely to produce hits (Medical Education, Medical Teacher, Advances in Health Sciences Education, BMC Medicine, BMC Medical Education, BMJ, Academic Medicine); the 'grey' literature (i.e. non-peer reviewed reports) and unpublished PhD research. We also performed backwards- and forward-citation searching.

Search period: January 2005 to August 2017. Publications outside of the stated timeframe were included if deemed to be particularly important – for example, seminal theoretical articles or large empirical studies. Searches were restricted to reports published in English.

Inclusion: Articles or reports containing qualitative or quantitative information about ARCP process or outcomes.

Exclusion: Articles or reports mentioning the ARCP but not containing information about the ARCP itself.

Articles were assessed for inclusion initially based on the title and abstract. The full-texts of articles selected for potential inclusion were read to make a final decision. Uncertainties about inclusion or exclusion of studies were discussed. All included studies were summarised in narrative form.

3.2 Ethics

The study was a review of published literature and did not require ethical approval – see https://ethics.grad.ucl.ac.uk/exemptions.php .

4. Results

The finding are presented in three main sections: Summative aspects of the ARCP (Section 4.3), Appraisal elements of the ARCP (Section 4.4), and Revalidation and the ARCP (Section 4.5). For each section we present a summary of the published guidance (how the ARCP should be done in theory); a summary of the literature (how the ARCP is done in practice, including how it is perceived by trainees and to an extent by trainers); and then bring in relevant evidence from the wider medical education literature to make suggestions for improving the ARCP.

4.1 ARCP scoping review: description of studies found

The searches gave 297 hits in addition to the 11 potentially relevant reports we already knew about. After removing duplicates and reports that were not relevant, we ended up with 30 reports for inclusion - see Appendix for full search results and details of the studies.

We identified five major themes covered in the ARCP literature (Table 1). Most addressed the summative aspects of the ARCP, although Theme 5 covered preparation for and feedback from the ARCP, and Theme 4 (stakeholder perceptions) also included perceptions of feedback.

Table 1: Main themes addressed by the ARCP literature, whether the theme relates to the summative or formative/appraisal aspects of the ARCP.

Theme	Elements of the ARCP
1. Relationship between ARCP outcomes and other measures.	Summative
2. Demographic differences in ARCP outcomes.	Summative
3. Stakeholder perceptions of ARCPs.	Summative and appraisal
4. A description and/or evaluation of ARCP process.	Summative
5. Preparation for the ARCP and feedback post-ARCP.	Appraisal

4.2 Purposes of the ARCP

There is agreement across the Gold Guide, the Foundation Programme and UKFPO ARCP guidance, and the GMC's Generic Professional Capabilities implementation guidance that the purposes of the ARCP are to:

1. Assess trainee achievement and learning, and suitability to progress, including identifying trainees who are not achieving or progressing satisfactorily:

    - Judge trainee performance to date (achievement of competencies) against curricula.
    - Judge trainee progression (achievement of competencies at a defined rate) to date against curricula.
    - Judge trainee suitability to progress to the next stage of training or complete training.
    - Provide advice to the Royal College/Faculty about completion of training.

2. Provide trainees with feedback to remediate poor performance:

    - Provide trainees with feedback to guide any additional training they may need to enable them to demonstrate performance and progress to the next stage.

3. Demonstrate rigour and fairness:

    - Be a fair, defensible, consistent, documented method of judging trainees.

4. Quality assure medical training and medical practice:

    - Provide advice to the Responsible Officer about revalidation.
    - Provide information to contribute to the quality assurance of training programmes.
    - Flag concerns about patient safety or fitness to practise.
    - Provide patients and employers with reassurances about trainee competence.

These purposes are explored in this review, with the exception of the quality-assurance of training and the provision of reassurances to patients and employers.

4.3 Summative aspects of the ARCP

4.3.1 Assessment of trainee progression and performance

The ARCP follows a competency based medical education model (e.g.(16)), in which trainees are considered *satisfactory* or *unsatisfactory* depending on whether they or not have achieved a minimum criterion-referenced standard; however the AoMRC/GMC guidance on General Professional Capabilities introduces the term *suboptimal* – the opposite of which is *optimal* which implies 'best' rather than 'minimally competent'. We discuss the implications of minimal competence on trainee motivation and learning in Sections 4.3.7 and 4.4.

The ARCP guidance differentiates between a trainee's *progression* (achievement of competencies at a particular rate) and *performance* to date (achievement of a specified number of competencies). Panels are also asked to judge a trainee's suitability to progress to the next stage i.e. their future progression and performance. There is an implication - not explicitly stated but reflected in the different types of unsatisfactory outcomes - that progression is more important than performance, since it reflects the trainee's 'own abilities, their determination' as well as 'their exposure to situations that enable them to develop the required competences' (9) (p.49).

In medical education, ability and determination are typically believed to be stable traits.(17) While there is good evidence that doctors' previous performance in assessments predicts their performance in later assessments, and this is partly underpinned by 'ability and determination', it is also the case that context-free intelligence or aptitude tests are relatively poor predictors of performance compared to achievement tests such as A levels.(18, 19) Furthermore, environmental and contextual factors are also important determinants of performance, and are considered partly responsible for differential attainment in medicine and elsewhere.(20-22) The ARCP does take some account of environmental factors, for example Outcome 5 provides trainees with the opportunity to collect and present evidence without additional training time, however the literature suggests that in practice an Outcome 5 is perceived by trainees as a criticism of their ability or motivation rather than a reflection of factors in the training environment that are beyond their control.(23)

The notion of testing trainee progression over time rather than their performance in one-off tests is increasing in popularity in medical education.(24) More and more medical schools are introducing progress tests. These are written tests of equivalent difficulty, administered to medical students several times over their undergraduate training. The results are combined to provide a 'growth curve' of learning, which is used to make judgements about suitability for progression that are more reliable and valid than those from one-off high stakes final assessments.(25) In principle, the ARCP provides an opportunity to use evidence about a trainee's progression over a year to predict their future performance. However, the ARCP is not a review of formal written progress tests, and, as explained in more detail below, a number of other factors potentially impact on the reliability and validity of the decisions made by panels. Furthermore, the ARCP is not currently perceived as taking into account contextual or environmental factors which can also affect trainee performance.

Suggestion: Greater clarity in the guidance about the importance of progression to date over performance to date for making judgements about future performance.

Suggestion: Greater recognition of contextual and environmental factors that can affect trainee progression and performance.

4.3.2 Panel judgements against curricula

The Gold Guide states trainee progression and performance should be judged against the relevant curriculum. It does not explain the process by which panels should go about making their judgements (and we discuss the implications of this further in Section 4.3.5). Because of the huge number of different specialty and sub-specialty curricula, the Gold Guide does not give details about which aspects of the curriculum are most important for panels to consider (see Section 4.3.3).

Curriculum-specific guidance has been issued by some medical royal colleges in the form of ARCP decision aids. These documents are developed by specialty-specific education committees within the colleges and describe the expectations for a minimally competent trainee at each stage of training. However, these decision aids do not exist across all specialties, and there is evidence that additional local guidance has been produced by Deaneries and Local Education and Training Boards even within specialties who have produced nationally-applicable guidance. This provides considerable potential for inconsistencies across specialties and locations.(26)

The literature suggests this situation undermines the ability of panels to make valid, reliable decisions. Trainers in Rothwell's qualitative study (27) reported finding it challenging to remember

what the curriculum requirements were for trainees at different levels, particularly when there was curriculum change; and trainees in Viney *et al*'s study reported trainers and systems being unable to keep up with changes in curricula.(23) Ntatsaki, Tugnet, Nadesalingham and Watson reported that the regular updates to the Rheumatology ARCP decision guide by the curriculum committee meant trainees and trainers needed additional information about what evidence should be reviewed at ARCP.(28) Trainees in Viney *et al* believed inconsistencies between specialties and grades was unfair.(23)

Suggestion: Mapping of all UK local and specialty-specific ARCP guidance to draw out the similarities and differences in the competencies required and the information provided about how panels should make judgements.

Suggestion: Delphi process (or similar) to achieve consensus about how panels should make decisions in practice, which could be done by specialty or by families of specialities. This process to inform the production of national ARCP decision aids, precluding the need for local decision aids.

Suggestion: Newly-developed decision aids to be compared to ensure consistency across specialties.

4.3.3 Which evidence panels use to make decisions

The Gold Guide notes that the following evidence should be reviewed by the panel: the Educational Supervisor's Report (but not the Clinical Supervisor's Report) workplace based assessments, examinations and self-reflective learning logs. It also states that logbooks, audit or quality improvement reports/projects, research activity and publications 'may be required to demonstrate progress'. As described above, the Gold Guide cannot provide a definitive list of, for example, the number and type of workplace based assessments that should be reviewed at ARCP. The UKFPO Guidance is able to be more specific, although the exact number of supervised learning events is not specified, and it is not clear what 'significant personal contribution to a quality improvement programme' looks like. Neither is there detail about what level of 'coverage of Foundation Professional Capability' is sufficient.

4.3.3.1 Weighting of evidence

The guidance is unclear as to whether any weighting should be applied to evidence. The Gold Guide implies, although it is not explicitly stated, that Educational Supervisor's Report (ESR) is the primary piece of evidence the panel should use. The UKFPO ARCP Guide does not imply this.

The limited published literature suggests that the Educational Supervisor's Report and the number of workplace based assessments are both important to panels, although their relative weighting may vary. McKee suggests the ESR is the primary evidence considered in surgical ARCPs.(29) Eynon-Lewis and Price identified the ESR and the number (not the quality) of workplace based assessments as the two most important pieces of evidence in the South West Peninsula Deanery.(30) Trainees in Viney *et al* described failing their ARCP because they had not submitted enough supervisor reports or workplace based assessments, suggesting the number of assessments is important.(23) In a first-hand account of receiving an Outcome 5, a surgical trainee described how his panel would not overlook missing assessments in favour of a positive Educational Supervisor's Report.(31) Similarly, a GP Educational Supervisor in Viney *et al*'s study described the negative impact on the trainee-trainer relationship when the ARCP panel overrode an unsatisfactory recommendation in an ESR.(23) By contrast, Dormandy and Laycock suggest many trainees in London Core Medical Training had passed their ARCP on the strength of a positive ESR despite not meeting e-portfolio requirements.(32)

Whether panels place emphasis on the quality as well as on the number of assessments is uncertain. Bedward, Davison, Burke and Thomas (33) found the average score for various workplace based assessments was a better predictor of ARCP outcome than how many of each assessment were undertaken in GP training; however that finding may reflect that trainees with higher scores also score better in other areas. Bodgener (34) found that case based discussions scored by GP assessors were a better predictor of ARCP outcomes for GP trainees than those scored by consultants or SAS doctors, perhaps suggesting ARCP panels weighted them more heavily.

It is uncertain whether the clinical supervisor's report, audits, and reflective logs are important to panels. Edwards and Petra (35) described how the three quality assurance criteria for ARCPs in General Practice are the quality of the ESR, a review of whether the evidence in the portfolio matches the outcome, and the quality of the clinical supervisor's report. Goodyear *et al* (3) noted that paediatric trainees with poor reflective logs tended to pass their ARCP, and that the poor quality of the reflection was generally not picked up by the supervisions; however of the 10 trainees in their study who had not completed any reflection in their log, all but one failed their ARCP, although they may have failed for another reason.

The medical education literature suggests that to make valid judgements, a panel needs to consider a variety of evidence reflecting different aspects of performance, rated by a variety of assessors,(36) and that both the number of assessments and the performance of the trainee are important.(37) The literature is unclear about what weighting should be applied to any pieces of evidence. In Section 4.3.5 we discuss in more detail the potential impact of inconsistent guidance about how panels should make decisions, including weighting of evidence. We also discuss the risks to reliability if the ARCP decision rests solely, or almost solely on the Educational Supervisor's Report.

Suggestion: More consistent and explicit national guidance about the variety of evidence that should be taken into account by panels.

Suggestion: Research to examine the impact of weighting on outcomes awarded.

4.3.3.2 Attendance of Educational Supervisor at the panel

The Gold Guide is vague about whether an Educational Supervisor should be a part of a panel considering their own trainee(s), stating that the Educational Supervisor should remove themselves if it is anticipated that the trainee will get an unsatisfactory outcome (i.e. they can be present otherwise). By contrast the UKFPO ARCP guidance explicitly states that 'supervisors cannot participate in the ARCP review panel of doctors under their own supervision'.

The issue of how much the supervisor should be involved in panel decision-making is much discussed by van der Vleuten and colleagues.(1, 24, 36, 38). While they believe supervisor input is very important to increase the accuracy of panel judgements ('the mentor knows the learner best' (38) (p.643), they are concerned that the trainee-supervisor (learner-mentor) relationship - which is crucial to learning - is likely to be compromised when the supervisor makes high-stakes decisions about trainee progression. This was described by a trainer in Rothwell:(27)

> *'It can be very difficult if you have to be both trainer and Educational Supervisor to someone that you're also having to give an outcome three to or maybe refuse an appeal, so there's a conflict of interest there. It can just get a bit messy' (p236)*

To protect the trainee-supervisor relationship, van der Vleuten and colleagues suggest the supervisor's mentor role is separated as much as possible from their assessor role. The value of this was described by a trainer in data from Woolf, Viney, Rich *et al*:(17)

> *'It can be quite useful to some to say [to the trainee] 'Well, look…' - especially if the trainee doesn't quite engage the way we want them to - so say 'Look, it's not me. That's the rules, it's not me that passes and fails you.''*

It may be that this separation is more important for supervisors than for trainees. Cho, Parry and Wade (39) surveyed and interviewed physician trainees and trainers in 2012 to gain their views about Supervised Learning Events, which are 'trainee-led formative assessments that aim to promote and facilitate learning' and that include feedback from supervisors. They asked whether SLEs should be considered by the ARCP panel, finding trainees thought they should, whereas trainers were split: a third felt they should, another third felt that if the SLEs were seen by the panel this might hinder supervisors giving feedback, and a third felt SLEs should contribute to the ARCP only for trainees who were underperforming.

Van der Vleuten and colleagues suggest a supervisor's (mentor's) input to a pass-fail decision could be 'limited for example, to a declaration of the mentor that the portfolio provides authentic evidence of the learner's progress' (1) (p.713). A step further is to ask the supervisor to 'make a recommendation on the performance decision, which can be amended by the learner' (24) (p.210). The ARCP goes one step further: the Educational Supervisor makes a recommendation that is not amended by the learner, and that decision in practice seems often to exert a heavy influence on the outcome (see Section 4.3.3.1 above). According to van der Vleuten *et al*(38) this may be a step too far since 'to eliminate bias in judgement and to protect the relationship with the learner, the mentor should not be responsible for final pass–fail decisions'(p.643), and it also potentially introduces a conflict of interest even if the supervisor steps out while the panel makes a decision.

Suggestion: Educational Supervisors to be absent when their trainee is being discussed by the panel.

Suggestion: Supervisor input to the process is valuable and should be in writing. While a supervisor report can include a recommendation about trainee progression, it may also be worth considering having all trainees submit written evidence stating whether they support or disagree with the Educational Supervisor's report, to lessen the conflict between the Educational Supervisor's dual roles as mentor and assessor.

Suggestion: Care be taken to ensure that the panel is not perceived as simply rubberstamping the Educational Supervisor's recommendation.

4.3.3.3 Attendance of trainees at the panel

The guidance states trainees should not routinely contribute to panel decision-making. The Gold Guide states trainees 'should not attend the panel', although 'HEE, NES, the Wales Deanery and NIMDTA may wish to have trainees present on the day to meet with the panel after its discussion of the evidence and agreement as to the outcome(s).'(9) (p.60). The Foundation Programme Reference Guide states that, even if a trainee is present on the day of the ARCP to discuss their training, they 'must not be present while the panel considers the outcome.'(10) It is also possible for an ARCP review not to take place because the trainee did not attend (see p.100 of the Gold Guide and p.35 of the UKFPO ARCP Guide), which implies that trainees are expected to attend.

As explained in Section 4.4.1 below, the literature suggests that in practice trainees often do attend panels, and feel their attendance affects the outcome they receive, which can contribute to making the experience stressful.(23, 27, 31) We were unable to find information in the wider literature about the educational benefits or otherwise of trainees attending panels; however it seems likely that having a trainee present could hinder a panel from awarding an unsatisfactory outcome and as such, we suggest trainees should not be present while panels are making decisions.

Suggestion: Panel decisions to continue to be made in absentia, although trainees may attend panels after a decision has been made, for example to receive feedback.

Suggestion: Ensure trainees and panels know all decisions about progression will be made on the basis of evidence submitted to be panel before the panel meets, not information provided by a trainee on the day of the panel, should they attend.

4.3.4 Ability of the ARCP to detect unsatisfactory performance and patient safety issues

4.3.4.1 Distinguishing between different levels of unsatisfactory

The ARCP needs to be able to accurately distinguish between satisfactory and unsatisfactory performance and progression, and to distinguish between different levels of unsatisfactory. It is not currently required to distinguish different levels of satisfactory.

Increasing the amount of information about a trainee's performance will generally improve the accuracy of judgements. In that regard, the Gold Guide's recommendations about the additional processes and information required if a trainee is anticipated to receive an unsatisfactory outcome could contribute to the accuracy of judgements about different levels of unsatisfactory performance (see Box 1). It is however uncertain whether the purpose of these additional processes is to improve the accuracy of decision-making or to protect the panel against appeals, or both. If it were to aid decision-making, it would be helpful to provide guidance to panels about the types of information they are seeking and how they should use that information.

Suggestion: Clearer guidance about the purpose of the additional processes and information required when trainees are anticipated to get an unsatisfactory outcome, including information about what purpose(s) the additional information will serve.

Box 1: Additional processes and information required by the Gold Guide for panels considering giving a trainee an unsatisfactory outcome. We suggest greater clarity about the reasons for these processes and how panels should use the additional information to make decisions:

- The Postgraduate Dean or nominated deputy must attend the panel (p.59)
- An external advisor must review the panel decision (p.70)
- The Postgraduate Dean should 'liaise with the Medical Director and the educational lead (e.g. the Director of Medical Education) or the general practice training and Training Programme Director where the trainee is employed/working […] to investigate and consider whether further action is required' (p.71).
- The trainee must be made aware of the concerns prior to the panel meeting, should have discussed them with the Educational Supervisor, the concerns should be documented in the educational portfolio, and the discussion should be documented (p.56).
- The Educational Supervisor should "withdraw temporarily from the [ARCP panel decision-making] process while their trainee is being considered" (p59).

The fact that additional information about a trainee is only sought if concerns are raised by the Educational Supervisor or training programme director makes it important that those people can easily detect, raise, and investigate any concerns. This is potentially problematic since it is well established that supervisors are reluctant to raise concerns about their trainees, partly because of the additional work involved.(27, 40) This makes it unlikely they will raise a concern about trainees whose performance is borderline or even fairly poor, which in turn makes it unlikely the panel will receive additional information about any but the worst-performing trainees.

The new AoMRC/GMC Generic Professional Capabilities Implementation guidance (12) goes some way to recognising and trying to disrupt the failure to fail problem, stating supervisors should be given 'time, training and support and be empowered to act if trainees are judged not to be making satisfactory progress. This principle is fundamental to the success of any system introduced to identify, support and successfully remediate trainees with suboptimal performance in generic professional capabilities.'(p.13).

Yepes-Rios *et al*'s review on failure to fail (40) supports the idea that training and supporting supervisors will reduce failure to fail. They also identified having a 'strong assessment system with established criteria' (p.1097) and having 'opportunities for trainees after failing' (p.1098) as important. As such, improving the ARCP guidance to make it clearer to panels and supervisors how they should be assessing trainees; evaluating and publishing data on the validity of the ARCP to improve the perception of the ARCP as a valid assessment; and making it more obvious to supervisors (and trainees) how trainees can be supported after receiving an unsatisfactory outcome, especially an Outcome 4, are all likely to be beneficial.

The review found supervisors find it easier to fail trainees when they had the opportunity to discuss a trainee's performance with others. This highlights the importance of having panels, not individuals, make pass-fail decisions, and the importance of giving panels enough time to discuss each trainee's performance (also supported by the evidence on group decision-making that longer discussion time leads to more valid decisions - see Section 4.3.5). Weighting evidence from a variety of sources

rather than placing so much emphasis on the Educational Supervisor's Report may help since Educational Supervisors' personal connection to a trainee can stop them failing the trainee.(40)

Other suggestions for reducing failure to fail include minimising the stigma of additional ('remedial') training by providing constructive feedback for all trainees - including stretching feedback for those performing well - and picking up on problems before the ARCP when the stakes are much lower. Both are described further in Section 4.4.

Suggestion: Put in place actions to improve the validity of ARCP outcomes by reducing the impact of failure to fail, including:

- training and support for supervisors and panels to fail trainees if necessary;
- greater clarity of guidance around panel expectations of trainee performance and progression (includes panels agreeing a shared mental model of what they are looking for before making judgements), probably via decision-aids;
- publication of evidence about the validity of the ARCP (demonstrating the strength of the ARCP as an assessment);
- better information for supervisors, panels, and trainees about support provided to trainees receiving unsatisfactory outcomes, especially for those receiving an outcome 4;
- ensuring panels consider and weight a variety of evidence about trainee performance and progression rather than relying on the Educational Supervisor's Report and recommendation.
- ensuring panels have time to fully discuss trainee performance before making a decision;
- normalising 'remedial' training by providing constructive feedback for all trainees, including stretching feedback for trainees performing well.
- interim-ARCPs and/or other tools to help the early identification of problems.

4.3.4.2 Distinguishing between satisfactory and unsatisfactory performance, and identifying patient safety concerns

To be able to distinguish between satisfactory and unsatisfactory performance, panels need to have a clear definition of 'unsatisfactory'. We have already described how the lack of standardisation across curricula can compromise the reliability of such judgements. In Section 4.3.5.1 below we additionally describe how the lack of specific guidance as to how panels should make decisions *as a group* can also hinder this process, and provide suggestions to address this.

The ARCP's role in ensuring patient safety is fairly vague in the Gold Guide. Although the ARCP is the tool for revalidating trainees, and revalidation emphasises patient safety (see Section 4.5 for more on revalidation), the 'What is the purpose of the ARCP' section in the Gold Guide (p.52) contains no explicit statement that the ARCP should be used to identify patient safety concerns. Section 7.49 of the Gold Guide (p.57) explains the Training Programme Director may need to make an additional report if a trainee has had a negative evaluation from their ES, and that the report can indicate 'that there may be a risk to patients arising from the trainee's practice', in which case 'this risk needs to be shared immediately with the Postgraduate Dean, the current employer and the LEP'. The Gold Guide also provides some information about what panels should do if they become aware of 'issues and concerns such as clinical safety or perceive undermining within the LEP' (p.58), although it's implied that the ARCP panel is not designed to routinely pick up such problems.

By contrast, the UKFPO ARCP Guidance explicitly states the ARCP 'will contribute to improving patient safety and the quality of care given by doctors.'(p.4). Similarly, the guidance on the implementation of Generic Professional Capabilities (12) describes how the introduction of GPCs is the 'regulatory response to the most common contemporary concerns about patient safety and fitness to practise within the medical profession' (p.5) and that 'the ARCP provides the existing and logical opportunity for the collation and overall assessment of professional capability' (p.14).

We found no data linking ARCP outcomes with patient safety or fitness to practise outcomes. We found six studies providing evidence that ARCP outcomes are positively correlated with performance in other assessments, including Royal College Examinations (33, 41), PLAB,(42) and tests for recruitment into specialty training.(43-45) This suggests that the ARCP is, to some degree, able to distinguish between different levels of performance, although there is not currently sufficient data to know how sensitive it is. This reflects the conclusion of Barrett *et al*'s recent review that there is not enough evidence to know whether workplace based assessments can identify and remediate poor performance.(46) Those authors also noted there was no evidence that workplace based assessments could distinguish between general underperformance and underperformance in specific areas e.g. communication skills – something which is important for ARCP panels in determining which unsatisfactory outcome to award.

The literature suggests trainees and supervisors have concerns that the ARCP measures clerical rather than clinical ability, that is does not reliably identify anything other than extreme poor performance, that it is not able to identify patient safety concerns, or that is measures the ability of a trainee to persuade a senior to sign off their assessments, which may be more difficult for trainees working less than full time or who were unfamiliar with the UK training system.(23, 27) (see also Section 4.3.6).

Obtaining good quality quantitative evidence for the reliability and predictive validity of the ARCP is problematic since ARCP data can be poor quality,(45) and because 'poor performance' and 'patient safety concerns' are not well defined or easy to measure – for example data are generally not available to link patient outcomes with individual trainee behaviour in the UK. That being said, an increasing number of studies have found a link between examination performance and GMC sanctions, (41, 42, 47), and it is in theory possible to determine whether the same is true for ARCP outcomes.

Suggestion: Improve the quality of ARCP data collection and reporting.

Suggestion: Conduct studies to quantify the ability of the ARCP to distinguish between satisfactory and unsatisfactory and between different levels of unsatisfactory, and to identify trainees who may pose a significant risk to patient safety.

4.3.5 Reliability of the ARCP

The reliability of an assessment can be thought of as its reproducibility: if the same trainee were to submit the same evidence to a different ARCP panel on a different day in a different part of the country, what is the probability that the trainee would get the same outcome? The reliability of the ARCP depends on the reliability of the panel decision-making, and on the reliability of the evidence submitted to the panel. We consider both below.

4.3.5.1 The reliability of the ARCP panel decision

The reliability of the ARCP is not explicitly mentioned in the guidance. There is reference in the Gold Guide and the Foundation Programme Reference Guide to panels making 'systematic' judgements, but there is limited detail about how this should happen in practice. The Foundation Programme guidance does provide some information about how many panel members should review a trainee's portfolio and what should be done if two panel members disagree. Furthermore, both the Gold Guide and the UKFPO ARCP Guide reproduce forms for recording some aspects of the panel decision-making (e.g. see Appendix 3 of the Gold Guide). It is likely these forms contribute to standardising the decision-making process and thereby improving its reliability.

We found no published numeric estimates of the reliability (inter-rater or test-retest) of ARCP outcomes. Nor did we find evidence about the number of appeals or success rates. One study into ARCPs in General Practice stated 'there is a consistency in [ARCP] outcomes across the deaneries and over time';(35) however no figures are provided. Our own descriptive analysis of ARCP outcome data published online by the GMC showed the proportion of unsatisfactory ARCP outcomes varies by specialty and region (see Figure 1), reflecting qualitative reports from trainees that the requirements of ARCP panels vary 'across specialties, regions and training grades' (23) (p.113). Without a large-scale longitudinal multilevel analysis it is uncertain how much these differences are due to variability in trainee ability, in curricula requirements, and/or in panel decision-making.

Suggestion: Research to examine the reliability of the ARCP panel decisions.

Suggestion: Guidance/decision-aids to specify assessments should be sampled across curricula and assessors.

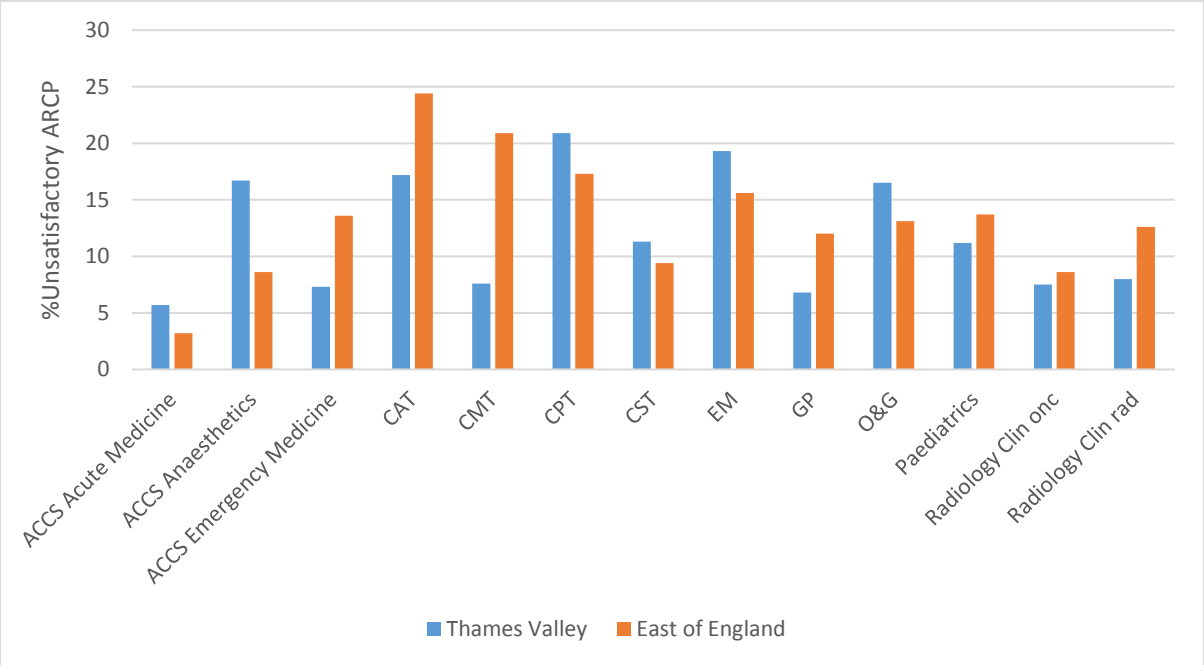


Figure 1: Variability in the percentage of 'unsatisfactory' outcomes awarded to trainees from 2010 to 2016 across 13 specialties in two Health Education England Local Education and Training Boards (ACCS= Acute Care Common Stem; CAT=Core Anaesthetics Training; CMT=Core Medical Training; CPT=Core Psychiatry Training; CST=Core Surgical Training; EM=Emergency Medicine; GP=General Practice; O&G = Obstetrics & Gynaecology; Clin onc= Clinical Oncology; Clin rad=Clinical Radiology). Data from http://www.gmc-uk.org/education/14105.asp

Recent publications from the United States have examined how psychological factors may influence the reliability and validity of decision-making of Clinical Competency Committees, which measure trainees' progressive attainment of competence.(48-50) Hauer *et al*'s (49) literature review on group decision-making made recommendations about improving the decision-making of Clinical Competency Committees. The Accreditation Council for Graduate Medical Education has also produced an evidence-based guidebook (15) which includes detailed guidance on the purpose of Clinical Competence Committees, their structure and membership, meeting preparation and running post-meeting documentation and follow-up, and legal issues and considerations. It includes a bibliography of relevant research. We have adapted Hauer *et al*'s table of recommendations for the ARCP, influenced also by the ACGME guidance (Table 2).

Suggestion: Review of the ACGME guidance on CCCs to assess its use in providing more detailed evidence-based guidance about how panels should be convened and run to maximise the quality of their decision-making.

Table 2: Suggestions for ARCP Panel members based on group decision-making literature. After 'Table 2: Recommendations for Clinical Competency Committees Based on Study Findings and Literature on Group Decision Making'.(15, 48)

Topic	Recommendation
Group composition	
Membership	Include new or rotating members as much as possible. Panel members assigned to represent disparate perspectives.
Size	Larger panels are better, so long as all members acquire relevant knowledge and demonstrate commitment. Minimum of three panel members (excluding lay members).
Group process	
Group understanding of its work	Members to agree on the role of the panel and their expectations of trainees at the start of panel meetings. Could include role descriptions and/or Terms of Reference for panel members.
Information sharing	Panel members to be encouraged to share information, particularly information not known to all panel members.
Sharing written information	Panels should rely primarily on written evidence and data when making decisions, and not on panel member memory.
Structuring discussions	Structured panel discussions should facilitate relevant information sharing. Can include solicitation of varied perspectives, members speaking in predetermined order, explicit weighing-up of alternatives.
Group leader soliciting perspectives	Panel chairs should encourage all members to share information.
Group leader encouraging elaboration and exchange	Chairs should repeat and summarise, ask about additional information, and encourage information exchange.

4.3.5.2 Reliability of the evidence presented to the panel

Workplace based assessments are the backbone of the ARCP. While it is well established that individual assessments typically have low reliability,(51) there is an increasing consensus that it is

possible to make reliable judgements about a trainee's overall competence on the basis of less reliable individual assessments, (1, 52) so long as judges have access to a large number of assessments and narrative reports sampled across curriculum areas and assessors.(36, 53)

The evidence suggests however that, in practice, assessments presented to the Educational Supervisor and in the portfolio are not always well sampled, for example because substandard training environments make it difficult for trainees to collect evidence, or because the high-stakes nature of the ARCP means trainees are incentivised to select assessors or cases they believe will show them in a positive light.(23)

As mentioned in Section 4.3.3 ARCP panel judgements are probably often heavily influenced by Educational Supervisor Reports. We found no evidence as to their reliability of Educational Supervisor Reports; however two studies from the literature indicated that a significant number may be poor quality.(30, 32, 35)

The Gold Guide refers to Educational Supervisor Reports as 'structured' and provides a general overview of the information that should be contained in the reports although a template - which is likely to improve the reliability of the report - is not provided. A non-systematic Google search revealed several templates in existence, and these varied considerably. It seems likely that the quality and content of ESR reports – and therefore of the ARCP decisions based on them - varies considerably across specialties and locations. For example, the Royal College of Ophthalmology template required a comment only when a trainee had not met a particular competency, whereas the Wessex Deanery Medicine Template required comments without requiring the supervisor to state whether a competency had been achieved.

The literature suggests that basing an ARCP decision on one person's expert judgement (e.g. the Educational Supervisor's recommendation) is likely to compromise the quality of the outcome, and states high-stakes judgements about trainee competence should be made by expert panels not individuals.(1, 35, 49) The Royal Colleges of Physicians has attempted to tackle this issue by introducing the Multiple Consultant Report, which provides the Educational Supervisor with a triangulated perspective on the trainee from a minimum of four consultants[2] who have worked closely with the trainee over the course of the training year.

Suggestion: Standardise the Educational Supervisor's report template, to ensure consistency within and between specialties.

Suggestion: Ensure panels consider a variety of evidence, not just the Educational Supervisor's recommendation.

Suggestion: Consider piloting the Multiple Consultant Report in non-physician specialties.

4.3.6 Fairness of the ARCP

The ARCP guidance mentions panel members should have training in equality and diversity; however, the literature shows that on average international medical graduates (IMGs), male doctors, older doctors, and doctors from black and minority ethnic (BME) backgrounds are more likely to have an unsatisfactory ARCP outcome.(27, 33, 42, 54) The exact reasons for these

[2] Some small specialties have been given special dispensation to include only 2 or three consultants in the MCR, on account of the difficulty in achieving four independent consultant views of a trainee within a year.

differences is unclear; although it is likely that it reflects the additional risk to achievement some groups experience during training.(20, 27) As mentioned in Section 4.3.1 above, greater consideration of contextual and environmental factors that influence a trainee's performance may improve the validity of the ARCP, and may also help improve the fairness of ARCP outcomes for trainees from groups who face additional risks during training.

There is also evidence that some trainees believe ARCP panels can be biased against minority ethnic and/or pregnant trainees.(20, 23, 55) Similarly, a trainee commenting on an ARCP article in a British Medical Association online forum stated that, 'If you are non-white trainee, you are at risk and your portfolio should be rock solid' https://www.bma.org.uk/connecting-doctors/b/work/posts/how-does-the-arcp-work-for-you. This emphasises the need to ensure and demonstrate panel decision-making is free from unfair bias. Ensuring panel members have had equality and diversity training, while a positive step, does not guarantee that panel decision-making will not be unfair. Indeed, Ahmed (56) has argued that this approach focuses the requirement on doing the training rather than on enacting the principles being taught, and as such training can in some cases actually conceal rather than guard against discrimination. Explicit discussion of equality and diversity during decision-making may remind assessors of their commitment to fairness as they make decisions.

Despite the ARCP being a high-stakes assessment, it has been subjected to relatively little scrutiny in contrast to some postgraduate exams, where the characteristics of examiners is monitored and subject to research (e.g. (57, 58)). As Tiffin *et al* (42) point out:

> *[ARCPs] cannot be considered free from cultural influences and opportunities for assessor bias, which would seem to be a priority area for future research. The extent to which IMGs are represented as ARCP panel members or what effect, if any, this may have on the outcomes of assessments is currently unknown. Certainly, monitoring the world region of qualification [i.e. region in which medically-qualified panel members obtained their primary medical qualification] as well as the ethnicity of postgraduate assessors and examiners may be important in research relating to potential bias or discrimination.*

Suggestion: Greater recognition of contextual and environmental factors that can affect trainee progression and performance.

Suggestion: Explicit discussion of equality and diversity principles in panel decision-making.

Suggestion: Monitoring characteristics of panel members.

Suggestion: Research examining the potential impact of the ARCP process and panel composition on outcomes for trainees from diverse groups.

4.3.7 Impact of the ARCP on trainee learning

The impact of the ARCP on teaching and learning is not discussed in the Gold Guide or Foundation Programme guidance; however, the literature shows that many trainees find the fact that it is impossible to achieve an ARCP outcome higher than 'Satisfactory' demotivating and see it as discouraging excellence.(23, 30-32) Furthermore, the several different types of unsatisfactory

outcome and the manner in which these categories have been used has created a perception of the ARCP as negative, capricious and bureaucratic, and detrimental to learning.(23, 27, 32)

There is much discussion in the medical education literature about the model of minimal competency upon which UK postgraduate medical training and assessment, including the ARCP, is largely based[3]. Eva, Bordage, Campbell and colleagues (2) argue that minimal competency is underpinned by incorrect assumptions: firstly that a trainee who can perform a task well in one context can perform it equally well in all contexts, and secondly that once competence at a task has been achieved and 'ticked off' a trainee no longer needs to work on it. Both assumptions risk encouraging trainees to learn just enough to get competency ticked off but no more, and to never revisit a competency once they have had ticked off. This can hinder learning, result in poor performance, and is potentially harmful to patients. A minimal competency approach can also make it difficult for educators to address concerns about trainees who 'bump along the bottom' - scraping passes but performing fairly poorly in practice. These trainees are considered equivalent to trainees performing at an extremely high level making it difficult to justify actions to help them improve their performance, or to put in place safeguards to prevent them slipping up in a way that could compromise patient care.

One approach to dealing with this would be to create new levels of 'satisfactory' - such as 'pass', 'good', 'excellent'. This would require developing criteria for each level, determining 'cut scores' (or equivalent qualitative standards) for each level, and ensuring that assessment standards were set and maintained nationally both within and between years and specialties. There would still be the potential for significant and relevant additional achievements (e.g. publications and leadership) to fall outside the scope of the assessment criteria and thus go unrecognised.

Eva *et al* (2) suggest that the provision of high quality constructive and relevant feedback to all trainees - including feedback that aims to stretch highly-performing trainees – may help combat some of the negative impacts of the minimal competency model. It may be that this is sufficient without the development of new levels of 'satisfactory'. We discuss feedback further in Section 4.4.

Suggestion: Consideration as to how panels might recognise and reward excellence.

Suggestion: Clearer guidance to panels about how evidence about important additional activities, such as leadership, teaching or research, are taken into account.

4.4 The appraisal aspects of the ARCP

We focus on the appraisal elements that are proximal to the ARCP event, namely the provision of feedback to trainees as after the panel decision has been made, and the preparatory discussions that typically occur between trainees and supervisors in the short period prior to the ARCP. Consideration of educational appraisal discussions between supervisors and trainees at other times (e.g. at the beginning of a rotation) were beyond the remit of the report.

[3] Interestingly, recruitment in postgraduate medical education is not entirely based upon a minimal competency model. For example recruitment into Foundation training includes the Educational Progress Measure, which is a decile ranking of a medical graduate in relation to their cohort at medical school. It is as yet uncertain whether or how the UK Medical Licensing Assessment will alter this.

4.4.1 Feedback given to trainees by the panel

At present, only trainees who receive an unsatisfactory ARCP outcome are mandated to attend the panel to 'discuss the recommendations for focused or additional remedial training if these are required.' (9) (p.60). This discussion is not officially called 'feedback' in the Gold Guide but in practice it is likely that it includes discussion of the trainee's previous performance as well as plans to help them improve their future performance ('feedforward'(59)). The Gold Guide leaves it to the discretion of HEE, NES, the Wales Deanery or NIMDTA as to whether all trainees under their remit meet the panel after their ARCP to discuss 'next steps and their future training requirements' (p.60), and it is unclear whether this should form an official part of the ARCP. Also at the discretion of HEE/NES/Wales Deanery/NIMDTA is whether 'for practical and administrative reasons' (p.60) the trainee's presence may be required to 'discuss other issues (e.g. the trainee's views on their training or planning of future placements)' (p.60). The Gold Guide is clear that this discussion is *not* part of the 'the review of evidence and the judgement arising from the panel' (p.60) and 'must be kept separate' (p.60).

The UKFPO ARCP Guidance suggests all trainees should have feedback about 'targeted learning, areas for improvement and/or areas of demonstrated excellence.'(p18) The implication is that the feedback is in writing, and only trainees who are anticipated to have a non-satisfactory outcome are asked to attend a post-ARCP meeting 'to provide feedback and discuss the particulars of supporting the doctor or possibly the exiting process, depending on which outcome is assigned.' The Foundation Programme Guidance also explains that if a foundation trainee has an *unanticipated* unsatisfactory outcome a meeting should be scheduled with the trainee, the Foundation Training Programme Director, and potentially other members of the panel. This is not specified in the Gold Guide.

Suggestion: Standardise ARCP guidance to include procedures for when a trainee has an unexpected unsatisfactory outcome.

We found no evidence about what proportion of trainees actually attend a meeting with the panel, although the literature suggests it probably isn't everyone: Bindal *et al* (60) explained that because of a shortage of consultants it was not possible for all Paediatric trainees in the West Midlands to attend a post-ARCP meeting (called an Annual Planning Meeting) in 2011. Also a group of GP trainers from England in Viney *et al* (23) explained that only trainees who were likely to receive an unsatisfactory outcome attended the panel, and Foundation Trainees in the study seemed not to have attended.

For trainees who do attend panel, the literature suggests that the separation of the ARCP (i.e. the panel's review of evidence and judgement) from feedback and other discussions may not always be clear, and trainees do not always perceive the process as useful.(3) (61) Where feedback is provided, it can be perceived as negative and even confrontational.(30, 62, 63) A GP trainer in Viney *et al* (23)described how recommending a trainee attend panel was equivalent to 'sending them to the headmaster's office for a telling off' (p.114), i.e. infantilising and punitive.

This literature does not mean that trainees do not want feedback. Indeed, many trainees know feedback is crucial to learning (e.g.(27, 32, 59)). Instead the findings support the wider literature that the benefits arising from feedback depend on the nature of the feedback, who it is delivered by, and how and when it is delivered. A recent review of how appraisal of doctors works (64) suggests that

positive learning and change can occur at an annual review, but it depends on having a skilled appraiser, a supportive work environment, and the appraisee having 'the right attitude'. An annual review with an appraiser, who presumably knows the doctor well, is different from an annual review with a panel who don't know the trainee. Research is needed to determine how Brennan *et al*'s findings translate to feedback provided by the ARCP panel.

Feedback from the ARCP panel cannot and should not supplant regular feedback from colleagues and supervisors, but having an ARCP system that provides feedback for all trainees - not just those considered not to be progressing or performing poorly – is educationally important. Panels providing training with constructive feedback and goals for improvement is likely to encourage a culture of learning and development, help trainees aspire to excellence, and enhance patient safety, and could help remove the stigma attached to additional ('remedial') training.(2) The ACGME guidance for Clinical Competence Committees (15) states 'All residents/fellows should receive timely feedback after CCC meetings, not just those for whom the CCC has concerns.' (p.18) and 'even the best residents/fellows can grow professionally and improve, so be careful not to short-change your more talented learners who will also benefit from robust feedback and the committee's providing them with some 'stretch goals.'(p.17). The guidance provides information how feedback can be collated and delivered usefully (see ps. 20-22).

Suggestion: ARCP panels to provide post-panel constructive feedback to all trainees, following best educational practice. Feasible mechanisms for doing so would need to be explored, but could include feedback being provided in writing, or being delivered face-to-face by a representative of the panel.

Suggestion: Panels be issued with a proforma for the sorts of feedback information that should be provided to trainees.

Suggestion: Panel members who have not received feedback training by other means (e.g. Educational Supervisor training) should be trained to provide feedback (as appropriate, whether face-to-face or in writing).

It is also important to consider what happens to trainees who receive an Outcome 4 and are released from training. Although all doctors receiving an Outcome 4 are required to meet with the panel or relevant person there is no mandated career guidance or feedback. Instead, the Gold Guide states that 'they may wish to seek further advice from the Postgraduate Dean or their current employer about future career options, including pursuing a non-training, service-focused career pathway.' Outcome 4s are rare and by definition those receiving that outcome are likely to be performing very poorly or pose another risk to patients. It makes sense educationally and for patient safety that these individuals are provided with structured support to develop their career in a different direction rather than having most educational support withdrawn. Further, as described in Section 4.3.4.1 above, greater support for those receiving an Outcome 4 (and making trainees, supervisors, and panels more aware of this support) can help combat a 'failure to fail' culture.

Suggestion: Greater clarity regarding the educational and career support available for trainees receiving an Outcome 4.

Finally, the ARCP prioritises the Educational Supervisor Report and assessments of trainees by seniors, with relatively little consideration of peer- and self-assessment other than that which occurs

in multi-source feedback assessments. By contrast the literature emphasises the importance of both peer- and self-assessment to learning e.g. (7, 65-67).

Suggestion: Panels to place greater emphasis on peer and self-assessment as well as assessment by seniors.

4.4.2 Feedback given to trainees to prepare them for the panel

The Gold Guide indicates that the ARCP is separate from 'the important processes of educational review and programme planning, which should respectively precede and follow from the ARCP process.'(p.52). Neither the Gold Guide nor the Foundation Programme Guidance describe in any detail how the trainee should be prepared for the ARCP, although since 'The Educational Supervisor is jointly responsible with the trainee for the trainee's educational agreement', the implication is that this extends to ARCP preparation. Preparation for the ARCP can include identifying problems in time for them to be rectified, and planning to ensure a trainee obtains the relevant experience and undertakes the relevant assessments.

The literature suggests that in the earlier years of the ARCP, many trainees did not feel prepared (3, 31) although it is uncertain whether trainees feel more prepared now. We found three studies that reported on tools to help trainees prepare for the ARCP by tracking achievement of competencies mapped to curricular requirements throughout the year.(28, 68, 69) These tools were found to help trainees manage the practical difficulties of ensuring their portfolios meet up-to-date curriculum requirements. They could also be used to alert trainees as to where they need to focus their efforts in order to prepare for the ARCP, alert trainees and trainers to potential problems, and prompt supervision meetings. We also found one article describing how life coaching to address affective and attitudinal problems (rather than knowledge and skills problems) can help trainees with persistent poor performance.(70)

Suggestion: Explore the potential for developing a standardised tool (adaptable to local contexts), to help trainees track their achievement over the course of the year and map these to curricular requirements. This information could be used by trainees and supervisors to identify early any problems in the training environment or trainee approaches to learning.

4.4.2.1 Interim or Pre-ARCP panels

In seeking to provide additional support to trainees preparing for the ARCP, we found examples of several Local Education and Training Boards who have introduced interim ARCPs. These varied in whether they were mandatory for all trainees or only for those identified as struggling, and how formative or summative they were e.g.:

- HEE Kent Surrey and Sussex provide all Core Anaesthetics Trainees with an interim ARCP around 5 months pre-ARCP. It is very similar to the ARCP except the trainee is required to attend. It is implied that the ARCP is summative since trainees without 'good evidence of competencies' will be required to attend the subsequent full ARCP panel. http://www.ksseducation.hee.nhs.uk/core-anaesthetics-interim-reviews-and-arcps/
- All Core Medical Trainees in the West Midlands have two interim ARCP checks at months 8 and 16 and two formal ARCPs at months 12 and 23. Details of what is discussed or who should be

present are not provided. https://www.westmidlandsdeanery.nhs.uk/Specialty-Schools/Postgraduate-School-of-Medicine/ARCP

- HEE Wessex requires GP trainees who had an unsatisfactory ARCP to have an interim ARCP 6 months afterwards to 'check on progression' http://www.wessexdeanery.nhs.uk/gp__primary_care/arcp/when_is_my_arcp.aspx .
- Orthopaedics trainees at The Mersey Deanery have an interim ARCP 'at the end of a trainee's first 6 months, if there has been an unsatisfactory ARCP outcome, or if there any concerns highlighted to the TPD' or 'to get a trainee back in step with their CCT date if the CCT date has changed for some reason'. They specify that 'most' interim ARCPs are 'formative rather than summative', although they do not explain how. http://www.merseyortho.org/wp-content/uploads/2011/03/ARCP-guide-1.docx

Pre-ARCPs that are mandatory for all trainees to attend may be similar to the penultimate year assessment (PYA) which has long been a feature of physician training (e.g.(71)). The PYA occurs approximately 12–18 months before a trainee's Certificate of Completion of Training (CCT) date. At the meeting targets are set and a final CCT date is agreed. Trainees must attend in person, and it is conducted by a single trainer rather than a panel, and the trainer being external to the trainee's Deanery/HEE area (e.g. https://www.jrcptb.org.uk/training-certification/penultimate-year-assessment.)

We found no evaluations of interim ARCP meetings, although we did find one 2011 study of paediatric trainees and trainers in the West Midlands who were surveyed about their experiences of attending a pilot Annual Planning Meeting (APM) three months pre-ARCP.(60) The most common topics discussed were future career and training plans, in contrast to the retrospective ARCP focus. The APM appeared formative rather than summative: trainees found it encouraging, non-confrontational and supportive, and liked that it did not rely on paperwork. Interestingly, trainers felt that only trainees in difficulty should have the meetings, but trainees who were progressing satisfactorily felt they gained from it, and as such the authors explained they planned to make them compulsory and a part of consultants' appraisal. No other evaluation of the APMs was presented.

As described above, ensuring all trainees have an interim ARCP that is a) less high-stakes than a final ARCP and b) designed to pick up and deal with problems early in a constructive manner, could help reduce problems related to 'failure to fail'. It may also help identify any patient safety concerns. Interim reviews for all trainees are likely to improve trainee motivation and performance if they are designed to be an opportunity for feedback and feedforward in a constructive manner, from a panel member as well as a supervisor, rather than simply being perceived as another 'tickbox' hurdle. Indeed, externality on the review panel may help address any problems with educational supervision, but even if there are no problems with supervision, having a panel (or at a minimum, more than one person) review a trainee's performance is likely to provide a more reliable and valid assessment of a trainee's progress at a stage when there is the opportunity to rectify problems.

It should be noted that the Accreditation Council for Graduate Medical Education in the US requires Clinical Competency Committees to review trainee evaluations semi-annually.(15) In practice, this may consist of the programme director (or designee) meeting each trainee to 'to conduct a formal written evaluation of performance', see for example: http://www.ucdenver.edu/academics/colleges/medicalschool/education/graduatemedicaleducation/GMEDocuments/Documents/Semi%20annual%20evaluation%20procedure.pdf .

We are aware that convening additional panel meetings is costly, however the interim review could possibly be conducted by a representative of the panel plus the Educational Supervisor. Educational Supervisors already hold mid-year educational appraisals, which could provide the basis for a small-scale, ARCP-focused interim review. A degree of input from an experienced ARCP panel member would help ensure that the supervisor's view is aligned with that of the panel, as supervisors have at times reported panel decisions conflicting with their own view of the trainee (23), and is likely to improve the validity and reliability – and therefore the utility - of the interim review.

The question of whether an interim review should have a summative as well as a formative purpose is important. As described above, aggregating the results of regular testing over the course of a year provides a more reliable assessment than performance in a single high-stakes assessment. This suggests it could be appropriate to feed the result of an interim ARCP into the result of the final ARCP, as happens in HEE KSS. A problem with this is that it will increase the stakes of the interim ARCP, which could increase the likelihood of gaming and failure to fail. It also makes it more likely that the interim ARCP will focus on the number of competencies achieved (and have a tickbox feel) rather than providing constructive feedback and feedforward. Further, if the Educational Supervisor is one of the key decision-makers in a summative interim review, this can lead to conflicts of interest. Consequently, a formative review is more likely to have a beneficial educational impact.

Suggestion: Formative interim ARCPs for all trainees, which focus on providing constructive feedback and feedforward, and on identifying any problems a trainee is having including any contextual or environmental factors affecting their progression. A degree of externality, possibly from a panel representative, is likely to be beneficial.

4.5 Revalidation and the ARCP

Revalidation is 'the process by which all licensed doctors are required to demonstrate on a regular basis that they are up to date and fit to practise in their chosen field and able to provide a good level of care. This means that holding a licence to practise is…an indicator that the doctor continues to meet the professional standards set by the GMC'.(13) Licensed doctors revalidate every five years, based on a recommendation made by their responsible officer (RO) to the GMC.

One of the main functions of the ARCP is to provide 'advice to the RO about revalidation of the trainee to enable the RO to make a recommendation to the GMC when required' (9) (p.53). The guidance as to how this mechanism operates for trainees is dispersed throughout several sections of the Gold Guide, but it is clear that ARCP panels should provide information to the RO based on the evidence presented to them by the trainee. This is confirmed in The GMC Guidance for Responsible Officers and Suitable Persons,(13) and the CoPMED revalidation guidance.(14) The evidence is contained within the trainee's educational e-portfolio (9) (p.54). Additional evidence may be required when the trainee is involved in a complaint or a serious untoward incident, in which case the trainee is expected to record a relevant reflection in their portfolio once the issue has been resolved.

Revalidation recommendations for doctors not in training (e.g. consultants or specialty doctors) is based on five annual appraisals – which are (in theory) *formative* events – whereas revalidation for trainees is based on the ARCP – which is a *summative* event. It is not clear from the guidance whether trainees should have an annual revalidation-orientated appraisal discussion of the type undertaken

by their non-trainee colleagues. In fact, trainees may currently welcome the blending of the ARCP and revalidation, such that two processes relying on the same evidence base are not run separately thereby creating an additional administrative burden. However, it is important to note that the two processes are different: the ARCP is a decision about a trainee's educational progress against curriculum standards, whereas revalidation is an annual review of a trainee's professional practice against the GMC's professional standards. As noted by Black ((72) p. 571),

> *It is possible to have no concerns educationally (an ARCP 1), but still have ongoing potential revalidation concerns. Equally, it is possible to be released from the programme (with an ARCP 4) but with no revalidation concerns.*

Thus, the revalidation focus of the ARCP panel should be a separate concern from the issue of progression. In practice it is possible that particular evidence presented in the course of the educational process may give rise to concerns about fitness to practice, and this is expressly the case for doctors in the first foundation ('F1') year. It is reasonable to suppose that adding the potential for a fitness to practice query to follow from the panel decision increases the stakes of the ARCP process in the minds of trainees.

In our view, the current arrangements for trainee revalidation, while light-touch and efficient are potentially confusing and are not well aligned with appraisal and revalidation as a consultant or specialty doctor. One suggestion for addressing this is for trainees to receive feedback in the ARCP that is aligned with the major domains of Good Medical Practice.

Suggestion: The guidance regarding the ARCP and revalidation to be gathered together into one section of the Gold Guide for clarity and ease of reference.

Suggestion: Clarification in the guidance as to how the processes for determining progression and revalidation relate to each other.

Suggestion: Trainees could receive feedback that is aligned with the major domains of Good Medical Practice, to help align revalidation during and after training.

5. Summary of findings and suggestions

The overarching suggestion from this report is that the ARCP panel is recognised as a high-stakes assessment that is likely to have a significant impact on patient care. The ARCP panel should therefore be subject to the same scrutiny and psychometric considerations as other high-stakes assessments, such as medical school finals and postgraduate examinations.

Throughout the report we have considered the ARCP to be an assessment with summative and formative purposes. The main summative purposes are to assess a trainee's progression and performance to date, and to determine whether a trainee can progress to the next stage of training. The main formative purpose is to provide feedback to guide trainees' learning (appraisal element).

The report's findings and suggestions are summarised under headings relating to the research questions. Major findings and suggestions are presented first, followed by a summary of the more detailed findings and suggestions, as appropriate.

5.1.1 How can the summative components of the ARCP process be standardised, and what measures can be taken to ensure that the process is robust, reliable and valid?

5.1.1.1 Major findings and suggestions

Major finding: The guidance lacks consistency and detail about how panels should make decisions. This hinders the standardisation, reliability, and validity of the ARCP.

Major Suggestion: Guidance to provide more detailed and practicable information about the information panels should use to make decisions, and how panels should go about making decisions in practice (including that panel decisions should be made without the trainee present). Decision-aids to be developed that are applicable nationally, and that are consistent in quality across specialties.

Major finding: There are concerns that ARCPs only identify very poorly performing trainees and fail to identify other trainees with specific or less serious performance issues. This may be partly result from the 'failure to fail' phenomenon.

Major Suggestion: Minimise failure to fail by having panels provide constructive feedback for all trainees post-decision; by providing time for panels to discuss trainee performance; by having panels consider multiple sources of evidence not just the Educational Supervisor's Report, by improving support for trainees who are failed, and by introducing lower-stakes formative pre-ARCP reviews.

Major finding: There is a lack of empirical research and published evaluations of the ARCP, and ARCP data provided for research has sometimes been of poor quality. This makes it difficult to assess the quality of the ARCP as an assessment.

Major Suggestion: Collect and provide good quality data for rigorous research and evaluation.

5.1.1.2 Summary of more detailed findings and suggestions

Finding: The guidance is unclear about the basis upon which panels should make decisions, leading to inconsistency and threatening validity.

Suggestion: The following needs to be clarified in the guidance to enable decision-aids to be developed that are applicable nationally and comparable across specialties, and which help panels make valid and robust decisions:

- Expectations around trainee progression and performance.
- How panels should weight different pieces of information (research needed to discover what panels are doing currently, and the impact weighting can have on outcomes).
- How panels should take into account any contextual and environmental factors that affected a trainee's performance (whether positively or negatively).
- The purpose of the additional information sought and the processes to be followed when a trainee is anticipated to receive an unsatisfactory outcome; and how the additional information should be used by panels to improve the validity of their decision-making.
- Expectation that assessments submitted in the portfolio are sampled across the curriculum and assessors.
- Requirement that Educational Supervisors be absent when panels are making decisions. If a trainee is present at panel, this will not influence panel decision-making (i.e. decisions to be made in absentia). Consider having all trainees submit written evidence stating whether they support or disagree with the Educational Supervisor's report to lessen the conflict between the Educational Supervisor's dual roles as mentor and assessor.

Suggestion: Decision-aids to be developed that are applicable nationally, and that are consistent in quality across specialties. To achieve this, it may be helpful to map all UK local and specialty-specific ARCP guidance to draw out the similarities and differences in the competencies required and the information provided about how panels should make judgements. Following this, a Delphi process (or similar) could be performed to achieve consensus about how panels should make decisions in practice. This could be done by specialty or by families of specialities.

Suggestion: Newly-developed decision aids to be compared to ensure consistency across specialties and locations.

Finding: There is a lack of evidence about whether the ARCP is able to reliably distinguish between satisfactory and unsatisfactory performance, or between different levels of unsatisfactory performance. Trainees and trainers are sceptical about the ARCP's ability to identify anything other than extreme poor performance, feeling it is not able to identify patient safety concerns but may be overly harsh to trainees with protected characteristics. Medical education research from outside the UK demonstrates reliable judgements about a trainee's competence can be made on a basis of a several workplace based assessments sampled across curriculum areas and assessors.

Suggestion: Collect and provide good quality data for rigorous research and evaluation regarding the validity and reliability of the ARCP. Ensure research is published.

Suggestion: Put in place training and support for supervisors and panels to fail trainees if necessary. This includes: normalising 'remedial' training by providing constructive feedback for all trainees (including stretching feedback for trainees performing well); better information for supervisors, panels, and trainees about support provided to trainees receiving unsatisfactory outcomes, especially for those receiving an outcome 4; providing sufficient time for panels to discuss each trainee's performance; ensuring panels consider multiple sources of evidence not just the

Educational Supervisor's Report; having formative relatively low-stakes pre-ARCPs for all trainees and/or using other tools to help the early identification of problems.

Finding: Panels make more reliable judgements than individuals, but steps need to be taken to guard against problems that can arise from group decision-making.

Suggestion: Panels to agree on their roles and their expectations of trainees at the start (potentially including clearer role expectations and Terms of Reference for panel members); to have proformas to aid structured decision-making; sufficient time for panel members to discuss each trainee, with panel chairs regularly summarising and ensuring panel members contribute and share as much information as possible. More panel members are recommended (with the proviso that all should be fully engaged) with three being a minimum.

Finding: On average IMGs, male doctors, older doctors, and doctors from black and minority ethnic (BME) backgrounds are more likely to have an unsatisfactory ARCP outcome. It is unclear whether this reflects panel decision-making or other factors. Unconscious bias training may not be sufficient to combat the potential for unfair bias.

Suggestion: Panels to explicitly state their commitment to ensuring decision making is fair during the ARCP. Panel membership to be monitored, efforts made to ensure diversity, and monitoring undertaken to check for any unfair bias. Panels to have greater consideration of the impact of training environments or other external factors on trainee progression and performance.

Finding: Many trainees feel demotivated by the fact that it is impossible to achieve an ARCP outcome higher than 'Satisfactory' and see it as discouraging excellence.

Suggestion: Consideration of how the ARCP process can recognise excellence. Constructive feedback for all trainees, including 'stretching' feedback for those performing well.

5.1.2 How can the appraisal elements of the ARCP process be standardised, ensuring that appropriate formative feedback is provided to all trainees?

5.1.2.1 Major findings and suggestions

Major finding: At present, only trainees expected to receive an unsatisfactory outcome are required to talk to the panel to discuss their performance, although the limited literature suggests in practice most trainees receive some kind of feedback from the panel, and many receive this in person. There is little guidance about the format in which feedback should be provided or what feedback should be about (e.g. past performance or future performance). There appears to be considerable variation between specialties and geographic locations, which is likely to hinder the validity of the ARCP.

Major suggestion: Ensure all trainees receive constructive feedback to improve their learning and performance, including trainees who are progressing satisfactorily. Provide guidance to standardise the way in which trainees receive feedback across specialties and locations. Ensure trainees and panels know panel decision-making will not be influenced by a trainee's attendance.

Major finding: There is no national guidance relating to preparing trainees for ARCPs. In practice, it seems different LETB's/Deaneries provide different types of pre-ARCP meetings for different groups of trainees. This lack of standardisation is likely to hinder the validity of the ARCP.

Major suggestion: Ensure all trainees have a pre-ARCP meeting with their Educational Supervisor and another person, possibly an ARCP panel member, to check progress and provide feedback. Provide guidance to standardise the pre-ARCP meeting process.

5.1.2.2 Summary of detailed findings and suggestions

Finding: All trainees are likely to benefit from feedback. Research on appraisal suggests the benefits depend on appraiser and appraisee factors, but it is uncertain how relevant this research is to panels providing feedback. ARCP prioritises summative assessment in the form of the Educational Supervisor Report and assessments of trainees by seniors, but peer- and self-assessment is important for learning.

Suggestion: All trainees to receive constructive feedback post-ARCP panel, including stretching feedback for those performing well. Standardised guidance on feedback to be developed and evaluated. Panel members to be trained to provide feedback. Panels to consider peer and self-assessment as well as assessor by seniors.

Finding: Neither the Gold Guide nor the Foundation Programme Guidance describe in any detail how the trainee should be prepared for the ARCP. The literature provides various ways to prepare trainees for the panel, such as ensuring trainees portfolios meet up-to-date curriculum requirements, use feedforward technique, or interim/pre-ARCP panels.

Suggestion: Formative interim/pre-ARCPs for all trainees, which focus on providing constructive feedback and feedforward, and on identifying any problems a trainee is having including any contextual or environmental factors affecting their progression. A degree of externality, possibly from a panel representative, is likely to be beneficial.

Finding: Lack of standardised support for trainees receiving an Outcome 4, despite many of these doctors continuing to practice medicine. Potentially risky for patient safety. Also likely to increase failure to fail.

Suggestion: Greater educational and career support for trainees receiving an Outcome 4. Trainees, panels, and supervisors to have more information about the support for trainees who are failed in order to combat failure to fail.

Finding: Limited literature supports the usefulness of tools to track trainee progression. These can help trainees ensure they achieve all curricular requirements, and identify early and problems in the training environment or trainee approaches to learning.

Suggestion: Explore the potential for developing a standardised tool (adaptable to local contexts), to help trainees and supervisors track achievement over the course of the year and map these to curricular requirements.

5.1.3 How can ARCP processes be aligned to GMC revalidation requirements?

5.1.3.1 Major findings and suggestions

Major finding: Lack of clarity about how revalidation decisions and progression decisions relate to one another. Lack of consistency between revalidation for trainees and non-trainees (e.g. consultants, staff grade doctors). The former is based on a summative ARCP and progression against

a curriculum, the latter is based on multiple formative appraisals and considered against Good Medical Practice.

Major suggestion: Clarification in the guidance as to how the processes for determining progression and revalidation relate to each other. Trainees could receive feedback from the ARCP panel that is aligned with the major domains of Good Medical Practice, to help align revalidation during and after training.

5.1.4 What understanding do trainees and supervisors have of the educational principles that underpin the summative and formative elements of ARCPs?

5.1.4.1 Major findings and suggestions

Major finding: The limited literature suggests many trainees feel the an annual review is valuable in principle, but have serious criticisms of the ARCP, perceiving that it does not provide meaningful feedback, that it can discourage excellence, and that it is not sensitive enough to pick up anything but very poor performance.

Major suggestion: Communicate to trainees how the ARCP review aims to:

- **ensure all trainees receive constructive feedback** in preparation for the ARCP and after the ARCP, including feedback to stretch trainees who are already progressing satisfactorily to encourage excellence, and by finding ways for process to recognise excellence; and by providing interim/pre-ARCPs and developing tools to help trainees and supervisors track a trainee's progression and identify problems early.
- **improve the reliability and validity of the ARCP** (and thereby contribute to patient safety) by standardising panel decision-making across specialties, grades, and locations; by increasing the panel's consideration of environmental and contextual impacts on a trainee's progression; by reducing the impact of a single Educational Supervisor's report on outcomes; by ensuring panels make decisions without the Educational Supervisor or the trainee present; by ensuring rigorous and transparent evaluation of the ARCP.

6. Conclusions

We found relatively little published research on the ARCP, and much of the evidence was small-scale. This reflects the fact that, despite its high-stakes nature, the ARCP panel is not officially an assessment and does not receive the same scrutiny as other high-stakes assessments in medical education. Our primary suggestion therefore that is the ARCP panel is officially recognised as a high-stakes assessment likely to have a significant impact on patient care and subject to the same scrutiny as other high-stakes assessments.

The lack of standardisation in ARCP processes is concerning, and we have suggested much greater clarity in the guidance in a number of areas to address this. We recognise prescriptive rules are not always helpful, however the current lack of clarity has led to a proliferation of local and specialty-specific guidance, which is likely to undermine the reliability and validity of the assessment components of the ARCP. We recommend new guidance is piloted to determine its feasibility in practice. We have also suggested that the relationship between the ARCP and revalidation, and the appraisal elements of the ARCP are more clearly defined.

The ARCP is generally perceived negatively by trainees, who feel it does not provide meaningful feedback, that excellence is not rewarded, and only the poorest performance is identified. We have suggested a number of changes that are designed to improve the validity of the ARCP by making it more meaningful to trainees. In addition, communicating with trainees about the changes being made to the ARCP and the rationale for those changes is likely to be crucial to restore confidence in the assessment.

Finally, we firmly believe in the principle that 'assessment is not easy to develop and is only as good as the time and energy put into it' (1) (p.707). As such, throughout this report we have made several suggestions that will require effort and resourcing. We recognise that investment is particularly challenging in the current circumstances, but we believe it is worthwhile. As Eva *et al* (2) point out: 'It seems antithetical to the very reasoning behind assessment (the protection of patients) to suggest that we should not think about how to improve current assessment practices, not only in terms of their role in gatekeeping but also in terms of their opportunities for shaping further learning' (p.907). We suggest investment in undertaking proper and continual evaluation of the process and outcomes of the ARCP, including any changes made, is essential to ensure the validity, reliability, robustness, and defensibility of the ARCP and its role in postgraduate training.

7. References

1. van der Vleuten CPM, Schuwirth LWT, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. Best Practice & Research Clinical Obstetrics & Gynaecology. 2010;24(6):703-19.

2. Eva KW, Bordage G, Campbell C, Galbraith R, Ginsburg S, Holmboe E, *et al*. Towards a program of assessment for health professionals: from training into practice. Advances in Health Sciences Education. 2016;21(4):897-913.

3. Goodyear H, Wall D, Bindal T. Annual review of competence: Trainees' perspective. Clinical Teacher. 2013;10(6):394-8.

4. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. Medical Education. 2005;39(3):309-17.

5. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist. 1995;50(9):741-9.

6. Cook DA, Lineberry M. Consequences Validity Evidence: Evaluating the Impact of Educational Assessments. Academic Medicine. 2016;91(6):785-95.

7. Gardner J. Assessment and Learning. London: Sage; 2012.

8. Arksey H, O'Malley L. Scoping studies: towards a methodological framework, . International Journal of Social Research Methodology. 2005;8(1):3.

9. A Reference Guide for Postgraduate Specialty Training in the UK: The Gold Guide. 6th Edition ed: Conference of Postgraduate Medical Deans of the United Kingdom; 2016.

10. The Foundation Programme Reference Guide. Foundation Programme; 2016.

11. Guide to the Foundation Annual Review of Competence Progression (ARCP) Process – 2017. Foundation Programme; 2017.

12. Generic professional capabilities: guidance on implementation for colleges and faculties. Manchester: Academy of Medical Royal Colleges and the General Medical Council; 2017.

13. The GMC protocol for making revalidation recommendations: Guidance for Responsible Officers and Suitable Persons. Manchester: General Medical Council; 2015.

14. COPMeD guidance on making revalidation recommendations for doctors in postgraduate training. Conference of Postgraduate Medical Deans of the United Kingdom; 2017.

15. Andolsek K, Padmore J, Hauser KE, Holmboe ES. Clinical Competency Committees A Guidebook for Programs: Accreditation Council for Graduate Medical Education (ACGME); 2015.

16. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. Medical Teacher. 2010;32(8):676-82.

17. Woolf K, Rich A, Viney R, Rigby M, Needleman S, Griffin A. Fair Training Pathways for All: Understanding Experiences of Progression. University College London; 2016.

18. McManus I, Woolf K, Dacre J, Paice E, Dewberry C. The Academic Backbone: longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the specialist register in UK medical students and doctors. BMC Medicine. 2013;11(1):242.

19. McManus IC, Dewberry C, Nicholson S, Dowell JS, Woolf K, Potts HWW. Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: meta-regression of six UK longitudinal studies. BMC Medicine. 2013;11(1):243.

20. Woolf K, Rich A, Viney R, Needleman S, Griffin A. Perceived causes of differential attainment in UK postgraduate medical training: a national qualitative study. BMJ Open. 2016;6(11).

21. Aronson J, Fried CB, Good C. Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence. Journal of Experimental Social Psychology. 2002;38(2):113-25.

22. Cohen GL, Garcia J, Apfel N, Master A. Reducing the Racial Achievement Gap: A Social-Psychological Intervention. Science. 2006;313(5791):1307-10.

23. Viney R, Rich A, Needleman S, Griffin A, Woolf K. The validity of the Annual Review of Competence Progression: a qualitative interview study of the perceptions of junior doctors and their trainers. Journal of the Royal Society of Medicine. 2017;110(3):110-7.

24. van der Vleuten CPM, Schuwirth LWT, Driessen EW, Dijkstra J, Tigelaar D, Baartman LKJ, *et al*. A model for programmatic assessment fit for purpose. Medical Teacher. 2012;34(3):205-14.

25. Wrigley W, Van Der Vleuten CPM, Freeman A, Muijtjens A. A systemic framework for the progress test: Strengths, constraints and issues: AMEE Guide No. 71. Medical Teacher. 2012;34(9):683-97.

26. Rangan A, Pitchford J, Williams P, Wood B, Robson S. Standardisation of delivery and assessment of research training for specialty trainees based on curriculum requirements: Recommendations based on a scoping review. BMJ Open. 2017;7(2).

27. Rothwell C. A study to identify the factors that either facilitate or hinder medical specialty trainees in their Annual Review of Competence Progression (ARCP), with a focus on adverse ARCP outcomes Durham University; 2017.

28. Ntatsaki E, Tugnet N, Nadesalingam K, Lawson T. The pre-Annual Review of Competence Progression Checklist: Demystefying the Annual Review of Competence Preparation for Rheumatology. Rheumatology. 2015;54(Suppl 1):1.

29. McKee RF. The Intercollegiate Surgical Curriculum Programme (ISCP). Surgery. 2008;26(10):411-6.

30. Eynon-Lewis A, Price M. Reviewing the ARCP process: experiences of users in one English deanery. BMJ Careers. 2012.

31. Peiris L, Cresswell B. How to succeed in the Annual Reviews of Competency Progression (ARCP). Surgery (United Kingdom). 2012;30(9):455-8.

32. Dormandy L, Laycock K. Triumph of process over practice: changes to assessment of physicians. BMJ Careers. 2015.

33. Bedward J, Davison I, Burke S, Thomas H. Evaluation of the RCGP GP Training Curriculum. University of Birmingham, University of Warwick; 2011.

34. Bodgener S, Denney M, Howard J. Consistency and reliability of judgements by assessors of case based discussions in general practice specialty training programmes in the United Kingdom. Education for Primary Care. 2017;28(1):45-9.

35. Edwards J, Petra H. The effects of external quality management on workplace-based assessment. Education for Primary Care. 2013;24(2):105-10.

36. van der Vleuten CPM. A Programmatic Approach to Assessment. Medical Science Educator. 2016;26(1):9-10.

37. ten Cate O. The false dichotomy of quality and quantity in the discourse around assessment in competency-based education. Advances in Health Sciences Education. 2015;20(3):835-8.

38. van der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJB, Heeneman S. Twelve Tips for programmatic assessment. Medical Teacher. 2015;37(7):641-6.

39. Cho SP, Parry D, Wade W. Lessons learnt from a pilot of assessment for learning. Clinical Medicine. 2014;14(6):577-84.

40. Yepes-Rios M, Dudek N, Duboyce R, Curtis J, Allard RJ, Varpio L. The failure to fail underperforming trainees in health professions education: A BEME systematic review: BEME Guide No. 42. Medical Teacher. 2016;38(11):1092-9.

41. Ludka-Stempien K. Predictive validity of the examination for the Membership of the Royal Colleges of Physicians of the United Kingdom. London: University College London; 2015.

42. Tiffin PA, Illing J, Kasim AS, McLachlan JC. Annual Review of Competence Progression (ARCP) performance of doctors who passed Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study. BMJ (Clinical research ed). 2014;348:g2622.

43. Gale TCE, Roberts MJ, Sice PJ, Langton JA, Patterson FC, Carr AS, *et al*. Predictive validity of a selection centre testing non-technical skills for recruitment to training in anaesthesia. BJA: British Journal of Anaesthesia. 2010;105(5):603-9.

44. Pashayan N, Gray S, Duff C, Parkes J, Williams D, Patterson F, *et al*. Evaluation of recruitment and selection for specialty training in public health: interim results of a prospective cohort study to measure the predictive validity of the selection process. Journal of public health (Oxford, England). 2016;38(2):e194-200.

45. Davison I, McManus I, Taylor C. Evaluation of GP Specialty Selection. University of Birmingham, University College London, University of Warwick; 2016.

46. Barrett A, Galvin R, Steinert Y, Scherpbier A, O'Shaughnessy A, Horgan M, *et al*. A BEME (Best Evidence in Medical Education) systematic review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees. Systematic Reviews. 2015;4(1):65.

47. Wakeford R, Ludka K, I; M, K W. Poor performance on postgraduate medical examinations predicts increased likelihood of regulatory sanctions in UK-registered doctors Annual European Conference on Assessment in Medical Education; Egmond Ann Zee, Netherlands: European Board of Medical Assessors; 2017.

48. Hauer KE, Chesluk B, Iobst W, Holmboe E, Baron RB, Boscardin CK, *et al*. Reviewing Residents' Competence: A Qualitative Study of the Role of Clinical Competency Committees in Performance Assessment. Academic Medicine. 2015;90(8):1084-92.

49. Hauer KE, ten Cate O, Boscardin CK, Iobst W, Holmboe ES, Chesluk B, *et al*. Ensuring Resident Competence: A Narrative Review of the Literature on Group Decision Making to Inform the Work of Clinical Competency Committees. Journal of Graduate Medical Education. 2016;8(2):8.

50. Chahine S, Cristancho S, Padgett J, Lingard L. How do small groups make decisions? Perspectives on Medical Education. 2017;6(3):192-8.

51. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. JAMA. 2009;302(12):1316-26.

52. Moonen-van Loon JMW, Overeem K, Donkers HHLM, van der Vleuten CPM, Driessen EW. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. Advances in Health Sciences Education. 2013;18(5):1087-102.

53. Driessen E, Scheele F. What is wrong with assessment in postgraduate training? Lessons from clinical practice and educational research. Medical Teacher. 2013;35(7):569-74.

54. Pyne Y, Ben-Shlomo Y. Older doctors and progression through specialty training in the UK: a cohort analysis of General Medical Council data. BMJ Open. 2015;5(2):e005658.

55. Rich A, Viney R, Needleman S, Griffin A, Woolf K. 'You can't be a person and a doctor': the work–life balance of doctors in training—a qualitative study. BMJ Open. 2016;6(12):e013897.

56. Ahmed S. 'You end up doing the document rather than doing the doing': Diversity, race equality and the politics of documentation. Ethnic and Racial Studies. 2007;30(4):590-609.

57. Denney ML, Freeman A, Wakeford R. MRCGP CSA: are the examiners biased, favouring their own by sex, ethnicity, and degree source? British Journal of General Practice. 2013;63(616):e718-e25.

58. McManus IC, Elder AT, Dacre J. Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCP(UK) PACES and nPACES examinations. BMC Medical Education. 2013;13(1):103.

59. Rees CE, Cleland JA, Dennis A, Kelly N, Mattick K, Monrouxe LV. Supervised learning events in the Foundation Programme: a UK-wide narrative interview study. BMJ Open. 2014;4(10).

60. Bindal T, Wall D, Goodyear H. Annual planning meetings: views and perceptions. The clinical teacher. 2014;11(7):524-30.

61. Vance G, Williamson A, Frearson R, O'Connor N, Davison J, Steele C, *et al*. Evaluation of an established learning portfolio. The clinical teacher. 2013;10(1):21-6.

62. Vasudev A, Vasudev K, Thakkar P. Trainees' perception of the Annual Review Of Competence Progression: 2-Year survey. Psychiatrist. 2010;34(9):396-9.

63. Vasudev A, Thakkar P, K V. The 1st Annual Review of Competence Progression, a new way of assessing trainee doctors: Trainees' perception. Medical Teacher. 2010.

64. Brennan N, Bryce M, Pearson M, Wong G, Cooper C, Archer J. Towards an understanding of how appraisal of doctors produces its effects: a realist review. Medical Education. 2017;51(10):1002-13.

65. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Studies in Higher Education. 2006;31(2):199-218.

66. Speyer R, Pilz W, Van Der Kruis J, Brunings JW. Reliability and validity of student peer assessment in medical education: A systematic review. Medical Teacher. 2011;33(11):e572-e85.

67. Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, *et al*. Programmatic assessment of competency-based workplace learning: when theory meets practice. BMC Medical Education. 2013;13(1):123.

68. Ap Dafydd D, Williamson R, Blunt P, Blunt DM. Development of training-related health care software by a team of clinical educators: their experience, from conception to piloting. Advances in medical education and practice. 2016;7:635-40.

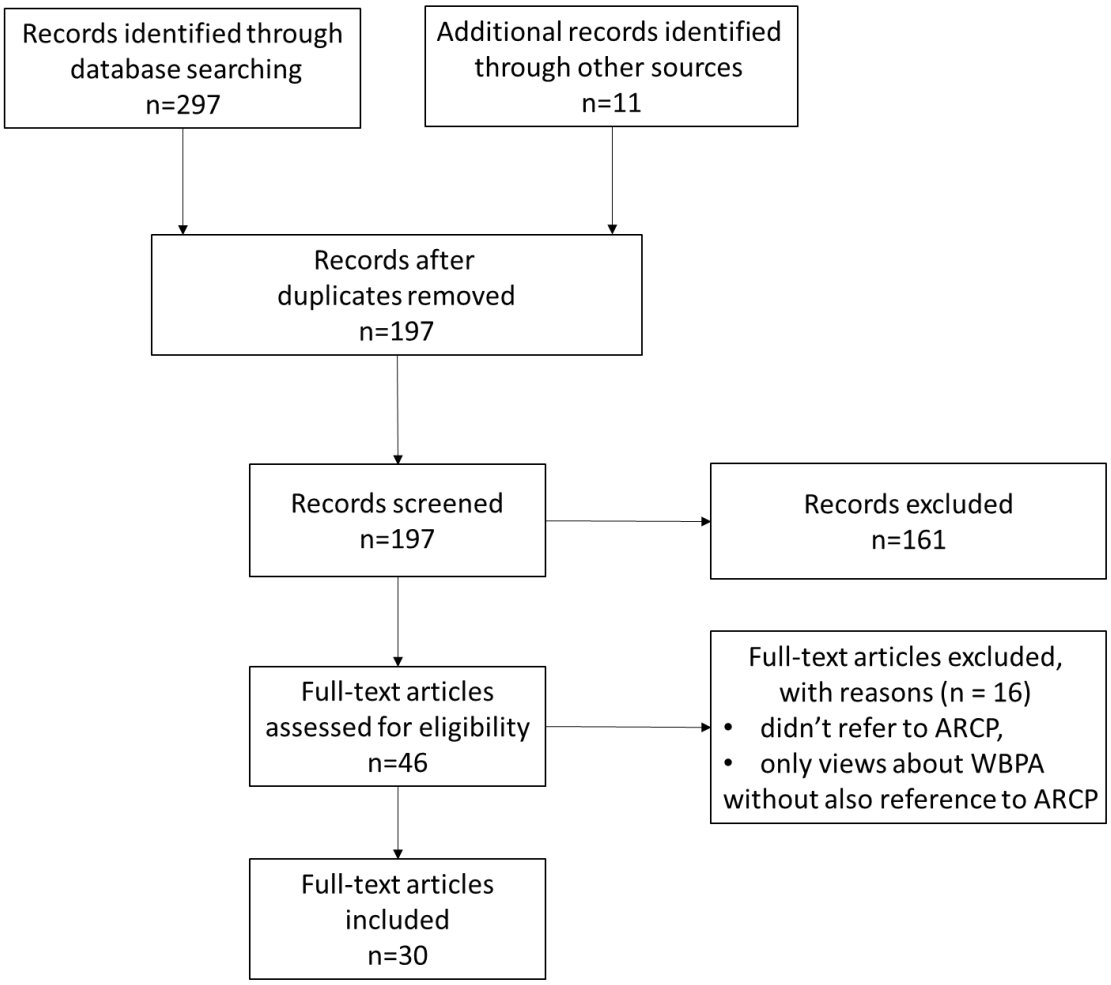
69. Wentworth L, Wardle K, Ruddlesdin J, Baht S, Baker P. Members' Presentations Abstract: Competency Mapping in Quality Management of Geriatric Medicine training – A Survey of Traineesin the North-Western Deanery. Medical Education. 2011;45:1-85.

70. Gale AC, Gilbert J, Watson A. Life coaching to manage trainee underperformance. Medical Education. 2014;48(5):539-.

71. Neville E. Modernising medical careers. Clinical Medicine. 2003;3(6):529-31.

72. Black D. Revalidation for trainees and the annual review of competency progression (ARCP). Clinical medicine (London, England). 2013;13(6):570-2.

8. Appendix



Supplementary Figure 1: PRISMA Flow diagram for the ARCP literature search

Supplementary Table 1: Summary of final full-text studies and reports. Report themes: 1=Statistical relationships with other variables; 2= Demographic differences in outcomes; 3= Stakeholder perceptions; 4=Evaluation or description of process; 5= Preparation and feedback.

Authors	Year	Report theme	Source	Report type
Tiffin et a;	2014	1	BMJ	Research paper
Pashayan *et al*.	2016	1	Journal of Public Health	Research paper
Ludka-Stempien	2015	1	University College London	PhD thesis
Goodyear *et al*	2013	1	Clinical Teacher	Research paper
Gale *et al*	2010	1	British Journal of Anaesthesia	Research paper
Davison *et al*	2016	1	Health Education England	Research report
Burnand *et al*	2014	1	R C S Bull	Research paper
Bodgener *et al*	2017	1	Education for Primary Care	Research paper
Bedward *et al*	2011	1,2	University of Birmingham	Research report
Rangan *et al*	2017	2	BMJ Open	Research paper
Pyne, Ben-Shlomo	2015	2	BMJ Open	Research paper
McKee	2008	2	Surgery	Description
General Medical Council	2016	2	General Medical Council	Research paper
Rothwell	2017	2,3	University of Durham	PhD thesis
Peiris, Cresswell	2012	2,3	Surgery	Opinion
Viney *et al*.	2017	3	J R Soc Med	Research paper
Vasudev *et al*	201	3	Psychiatrist	Research paper
Vasudev *et al*	2010	3	Medical Teacher	Research paper
Vance *et al*	2013	3	Clinical Teacher	Research paper
Goodyear *et al*	2012	3	Medical Teacher	Research paper
Eynon-Lewis, Price	2012	3	BMJ Careers	Careers article
Dormandy, Laycock	2016	3	BMJ Careers	Careers article
Cho *et al*	2014	3	Clinical Medicine	Research paper
Edwards, Petra	2013	4	Education for Primary Care	Research report
Black	2013	4	Clinical Medicine	Opinion
Wentworth *et al*	2011	5	Medical Education	Abstract
Ntatsaki *et al*	2015	5	Rheumatology	Poster
Gale *et al*	2014	5	Medical Education	Brief report
Bindal *et al*	2014	5	The Clinical Teacher	Research paper
Ap Dafydd *et al*	2016	5	Adv Med Ed and Prac	Description