# What should be in an early safety assessment? Lessons from the experts and RTO experience

FINAL RESEARCH REPORT

ASSOC PROF JILL BENSON AM
DR AMELIA WOODS
DR TARYN ELLIOTT
MS MICHELLE PITOT

*In partnership with*

This research project is supported by The Royal Australian College of General Practitioners with funding from the Australian General Practice Training Program: An Australian Government initiative.

## Acknowledgements

GPEx would like to acknowledge that this research project has been supported by the Royal Australian College of General Practitioners with funding from the Australian General Practice Training Program: An Australian Government initiative.

GPEx would also like to acknowledge Steering Group members: Dr Graham Emblen (GPTQ), Dr Kristen Fitzgerald (GPTT), Prof Parker Magin (GP Synergy), Dr Helen Mullner (GPEx) and Prof Lambert Schuwirth (Prideaux Centre, Flinders University) for their guidance and input into this research.

# 1. Executive summary

**Aims and objectives**

Assessing the safety of GP registrars when they first start community practice is critical for the safety of patients, the registrar themselves and also the practice in which they are working. An Early Safety Assessment (ESA) is not only about assessing current competency, or saying the registrar is safe for independent practice, but is also about whether the registrar is able to self-assess their competency limits, and seek help appropriately when required.

This project aimed to answer the following questions about Early Safety Assessments.

1. What is currently included in each RTO's ESA and why were these assessments chosen?
2. What should be included in an ESA and how should this inform flagging of registrars at risk of safety breeches?
3. What criteria are used for flagging in each RTO, how many registrars are flagged in each domain, and is this similar or different across RTOs?
4. What is the feasibility, acceptability and cost-effectiveness of an ESA?

**Method**

There were three main streams in this research.

**Stream1:** Interviews with the Directors of Training (DoT) at four Regional Training Organisations (RTOs) about what their RTO is currently using in an ESA, what they thought should be included in an ESA and why. This then informed the questionnaire for the first round of a modified Delphi consensus. The DoTs continued on as Delphi participants along with other experts. Three rounds of Delphi questionnaires were needed to reach consensus. The DoTs were then re-interviewed about the feasibility, acceptability and cost-effectiveness of the recommendations.
**Stream 2a**: Documents from the three participating RTOs used in their ESAs were collected and analysed.
**Stream 2b**: Flagging data was analysed from participating RTOs (based on the first semester of 2021). Data showed when registrars were flagged, by whom, the reason for the flag, the seriousness of the flag and the outcome.

Information from the three streams was triangulated to inform the final recommendations.

**Summary of results**

The majority of participating RTOs began their ESA before the commencement of community placement, but all ESA programmes were different. However, in the final interviews there were concerns about the feasibility of beginning an ESA before training. Very early in the placement, the ESA should include a Multiple Choice Questionnaire (MCQ) (including prior to community placement), registrar self-assessment tool, an OSCE-style workshop (though the feasibility of this was questioned by the DoTs), but not a multi-source feedback before placement, nor indeed as part of the ESA at all.

Supervisors should be given templates and guides to assist them for example with orientation, building the relationship with the registrar, direct observations, random case analysis, case-based discussions, competency frameworks, high risk/call for help lists, flagging protocols and global assessments. There was consensus that there should be a period of between 1-4 weeks when the registrar is supernumerary so they can undergo orientation, and have shared consultations. In addition, there should be between 1-4 weeks, tailored to each registrar's needs, when the supervisor discusses every patient with the registrar, and the supervisor should directly observe the registrar consulting for the equivalent of at least one session before week 2.

Similarly, the Medical Educator (ME) and/or External Clinical Teacher should have guidance templates to assist with assessing a wider range of skills and issues. They should directly observe the registrar between weeks 4 and 12, of at least four patient consultations, undertake random case analysis, and have a discussion with the supervisor and practice manager.

A global assessment such as one using an EPA-style template, should triangulate information from a variety of assessments and sources to decide whether the registrar should be 'flagged' or not. A 'diagnosis' of the reason for the flag and what the next steps might be, and how the flag will be 'signed off', should be done in collaboration between the supervisor, medical educator, remediation committee and/or DoT depending on the severity and nature of the flag.

The ESA should end when the registrar is flagged and a plan developed, or when the triangulation of data reaches the conclusion that the registrar is 'safe to practise with the supervisor available the majority of the time'.

Domains for flagging in the participating RTOs were graded as minor (watching) or major (active) and were mostly identified by the supervisor or ME. Flags were used for concerns relating to clinical knowledge, communication, personal or family issues and professionalism. About a third of the flags were in the first 12 weeks of community placement. Opinion was that some problems will take time to reveal themselves, especially those regarding personal or professionalism issues, which can then further impact clinical or communication problems in the future.

Barriers to the implementation and ongoing conduct of an ESA include: supervisor engagement, lack of supervisor training, supervisor reluctance to make a judgement, lack of time, geography, IT issues and bureaucracy. Facilitators included: a strong relationship between the supervisor and registrar, availability of information about assessments, stable technology, adequate funding for supervisors, flexibility, training, and a longer period of time in which to assess the registrar.

**Discussion**

The balance between funding, time, and an adequate assessment of a registrar's safety in the early part of their training will always be a difficult one. However assessing whether a registrar is safe to see patients without direct supervision, when the context is often completely new, is essential. A suite of assessments, templates, guidance documents, training, funding, support and personnel should be embedded in any General Practice Training organisation. Flexibility to tailor the process to the needs of the registrar, the practice, the supervisor, the geography and the context should also be built into the model. Flagging processes should include the ability to identify the reason for the flag, the severity of the flag, what the next steps might be and how the flag will be 'signed off'. These processes should be transparent with an aim to support the registrar to self-reflect and improve, and not be seen as a pass/fail.

**Implications**

This research has highlighted the complexity of an ESA, but also the importance of having flexible and strategic processes in place that allow for all involved to be appropriately supported to assess and remedy early safety issues.

**Future Research**

There is a plethora of issues regarding early safety assessments that would benefit from more research. These include the optimal duration of the ESA in the first semester of GP training; which are the most useful assessments for identifying and flagging registrars at risk; what are the possible outcomes of an ESA; how long a

period of supernumerary practice or closer supervision should be; how should a 'diagnosis'of a flag be made and remedied; and does an ESA and suitable remediation mean the doctor is safer in the long-term.

## 2. Background

The RACGP Standards for General Practice require that registrars practise within the scope of their competence to minimise threats to the safety of patients, themselves and other practitioners, the practice, and the profession. Since registrars are in a training program, it follows that their scope of competence, and awareness of this, will be less than that of Fellowed GPs. Training programs, in collaboration with practices, therefore have the responsibility to assess registrars' competence and provide guidance on how to address any gaps identified in order to ensure the registrar is practising safely.

However, an Early Safety Assessment (ESA) is not only about assessing competency, or saying the registrar is safe for independent practice, but also about whether they're behaving in a safe manner and not practising unsafely. It's about whether the registrar is able to self-assess their competency limits and seek help appropriately when they're outside their competency or comfort zone. Safety mechanisms need to be in place to assist registrars to recognise their own limitations and have adequate supervision in place when it is required. Safety is about setting up a system and supporting supervisors to ensure that those safety mechanisms are in place, and that 'flagging' will happen if the registrar needs assistance beyond what would normally be needed. Flagging is about early recognition of registrars requiring additional training support for a variety of reasons.

An early safety assessment will also give the training organisation information about the bigger picture of current and emerging training gaps. It will assist with planning the ongoing training program and education and remediation interventions. It is important that the Training Organisation communicates this purpose of the safety assessment to the registrars. As a part of this communication, the clear message should be that this is not a pass/fail assessment. Instead it is a learning opportunity and feedback should not only be used by the RTO- but also by the registrar to identify strengths and opportunities for development.

Research and theory provide some guidance for how such safety assessments should be conducted and what form they should take. The optimal point for these assessments is at the beginning of registrars' training. From a theoretical point of view, this aligns with the concept of secondary prevention – the screening and detection of early signs of an illness, which, if left unmanaged, would become worse (WHO). In this context, early signs of potential problems can be identified before they worsen and threaten safety. 'Flagging' is the term generally used in medical education to describe the identification of at-risk learners and those in difficulty in order to provide them with the 'necessary supports or remediation in their development to becoming effective, safe, self-reflective and resilient doctors' (Prentice 2021). An additional benefit is that early flagging with appropriate intervention of at-risk registrars is associated with an increased likelihood of passing exams (Prentice 2021), presumably because this affords more time to address issues.

## 3. Literature review

As Morgan et al (Morgan 2015) have said, 'patient safety is the cornerstone of quality care, and monitoring patient safety is the key aspect of clinical supervision'. Assessing new registrars' safety in practice is particularly critical for general practice training, as there is an increasing disconnect between junior doctors' experiences in the hospital system, and the workstyle and clinical characteristics of general practice (Wearne 2018). Furthermore, the solitary nature of general practice training means that there are difficulties in registrars knowing the boundaries of their competence, which in turn, places them at increased risk of experiencing

'unconscious incompetence' (Byrne 2012). Indeed, there is cause for concern regarding the safety of patients who see registrars. Morgan et al (Morgan 2015) found that, of supervisors participating in a study using RCAs, 30% found that Random Case Analysis (RCA) had identified safety issues, of which half required patient contact or management changes. This suggests that there is a relatively high rate of 'unconscious incompetence' unnoticed within GP training. The lack of mandated one-on-one supervision for a period of time at the beginning of GP training is a notable difference for general practice in Australia compared to many other countries (Ingham 2019). Furthermore, Ingham et al (Ingham 2015) found most supervisors in their sample relied on opportunistic engagement to monitor the safety of registrars' patients rather than using observation and audit techniques, suggesting the quality of such monitoring is not necessarily able to be met by current mechanisms. Hence there may be many examples of unconscious incompetence not necessarily identified by the current supervision practices observed in this study. Yet, Australian General Practice Training (AGPT) Program Regional Training Organisations (RTOs) largely assign responsibility to monitor the safety of registrars to supervisors and practices. There is no standard requirement for specific follow-up by RTOs in order to assess the safety of patients seen by registrars (Ingham 2019).

An ESA involves assessments undertaken by all registrars early in their training journey (usually toward the start of GPT1) that are collated to identify whether they are having difficulties, or where there are gaps in knowledge and practice. An ESA can guide early flagging of those registrars who need additional assistance before unsafe or unhelpful gaps or behaviours have become embedded or have potentially caused harm. Moreover, early detection of issues has been shown to be important not only for safety of all involved, but also when accompanied by tailored remediation to be associated with an increased likelihood of passing Fellowship exams on the first attempt (Prentice 2020, GPEx 2019). This in turn can offset costs associated with increased training time and late remediation.

The idea for this project emerged from the work of the Workplace Based Assessment (WBA) Education Research Grant (ERG) of 2018-2019 (GPEx 2019). In that study, five of the nine RTOs who were part of the WBA ERG were using some type of ESA. The WBA ERG demonstrated that, by comparing different flagging and remediation features across RTOs, a broader understanding of each could be ascertained. In light of recommendations from the WBA ERG, other RTOs have now also developed an ESA to be undertaken by all GP registrars early in their training. Each RTO however, has developed their ESA in isolation, and there is no data currently available describing the content, structure, outcomes, feasibility or acceptability of ESAs within RTOs.

Moreover, there is very little evidence around when an ESA should be conducted, what can be expected from a new registrar, and how this is best assessed. This information is needed to inform the ongoing review and development of ESA content and processes across RTOs, leading to an evidence-based, feasible and acceptable system.

When assessing how safely a registrar is practising early in their training time it is important to consider the safety of the patient, the practice and the registrar themselves, such as their personal well-being and future retention in the work-force. There are various suggestions for exactly what areas in each of these domains should be assessed as a part of an ESA. One aspect of early safety that has been well researched in the literature is 'call for help' lists. Bowie et al (Bowie 2012) produced a 47-item checklist of 14 safety-critical domains that supervisors should address with their registrars within their first 12 weeks of training. Ingham et al (Ingham 2019) identified that most RTOs already have a similar list that they provide to supervisors as a resource, although the degree of overlap between the various lists is unknown. Yet, such lists in isolation are unlikely to translate into an effective assessment of safety. As raised by participants in the study by Ingham et al (Ingham 2019), prescribing such a list may suggest to registrars that items not on this list are not as critical, potentially

meaning registrar practice becomes less safe. Additionally, the provision of a list as a resource does not necessarily make it an assessment of trainee's safety, and the lack of formal monitoring of registrars' safety and how they address the supervision requirements for the items on the list, remains uncontrolled in many RTOs.

Moreover, in designing an ESA it is important to consider how safety should be assessed. While Wearne et al (Wearne 2018) proposed that, as part of the selection of GP registrars, a clinical knowledge assessment may be included so as to highlight the gaps that need addressing, this is unlikely to be adequate. Magin et al (Magin 2017) found in their study that colleague feedback and Direct Observation Of Procedural Skills (DOPS) visits were significantly associated with formal remediation. They also proposed that interpersonal skills questions, a Multiple Choice Question (MCQ) exam and personal reports by the registrar of any personal issues (e.g. psychiatric illness) were important considerations (Magin 2017). However, they found that selection scores were not accurate predictors of the need for remediation. Murphy et al (Murphy 2009) found that Multisource Feedback assessments (MSFs) were particularly useful when conducting WBAs, though this finding was not specific to ESAs. Ingham et al (Ingham 2019) found that most RTOs already have such assessments, but that these are perceived to be more useful for identifying registrars' learning needs than assessing registrars' competence. Viewed through the lens of Miller's pyramid (Miller 1990), the inadequacy of such assessments for assessing registrars' competence is that they can only assess up to the point of 'shows how', whereas what is attempting to be assessed is what a registrar 'does' (the pinnacle of the pyramid). Accordingly, for ESAs to be most effective, they likely need to comprise WBAs. Indeed, this aligns with the suggestions of Byrnes, Crawford (Byrnes 2012) that monitoring registrars' safety is best achieved through regular monitoring of trainees' actual performance. Furthermore, by focusing on what the registrar has 'done', particularly in cases where there were near misses, this may provide a strong impetus for, particularly transformative, learning moments (Branch 2005), and so be more able to effect changes in the registrar.

## 4. Objectives

The following objectives were explored through this ERG.

1. What is currently included in each RTO's ESA and why were these assessments chosen?
2. What should be included in an ESA and how should this inform flagging of registrars at risk of safety breeches?
3. What criteria are used for flagging in each RTO, how many registrars are flagged in each domain, and is this similar or different across RTOs?
4. What elements of the early safety assessment are most useful for flagging, feedback and design of tailored remediation?
5. What is the feasibility, acceptability and cost-effectiveness of an ESA?

## 5. Methodology

Ethics approval for this research was obtained from Flinders University HREC (Project No. 4167).

A summary of the data collection methodology is in Figure 1.

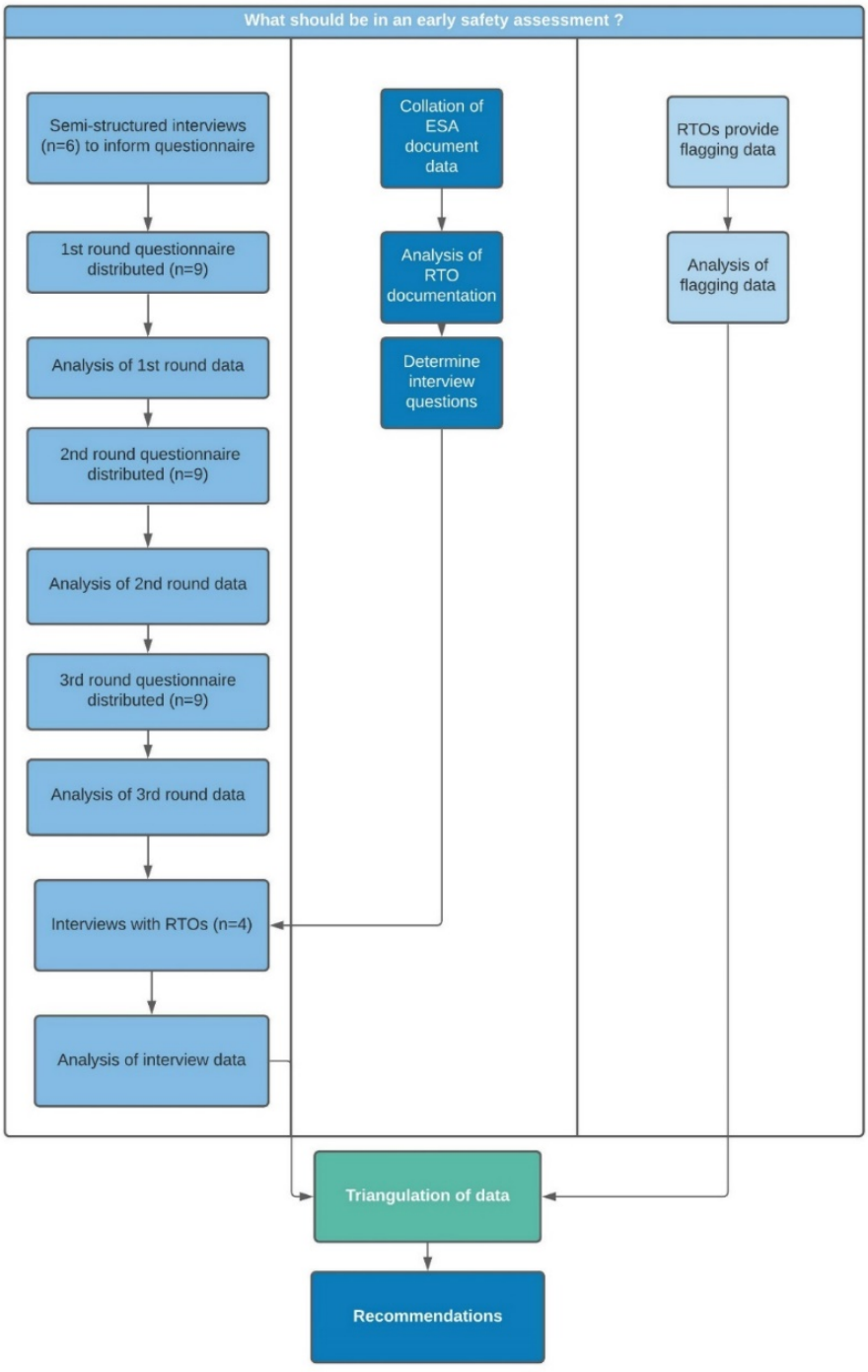


*Figure 1: Data collection for the research project*

### 5.1 Stream 1

Because minimal literature was available to inform recommendations about the content and structure of an ESA, and even less available in an environmental scan, Stream 1 involved a modified Delphi consensus. Guidelines for the development and implementation of a Delphi consensus (Hasson 2000) were followed, with the aim that the experts reached agreement on what the features of a safe GPT1 registrar are, how this can be assessed and what criteria would be useful for flagging registrars in difficulty. The Delphi consensus is a group facilitation technique that seeks to obtain consensus on the opinions of experts through a series of structured questionnaires (referred to as rounds). Questionnaires are completed anonymously and the responses from each questionnaire are fed back in summarised form to the participants for each of about three rounds, or until consensus is reached (Hasson 2000). The expert panel is de-identified from each other.
Our project conducted a modified Delphi, whereby a group of experts from RTOs were interviewed about their perspectives on a range of features of an ESA to understand what should be included, when in the registrar's training timeline it should occur, and how the ESA should be conducted. This methodological approach has been performed previously (Humphrey-Murto 2017), and was especially useful as there was very little published reports or research on this subject on which to base an initial questionnaire. This semi-structured interview consisted of initial open-ended questions with follow up and clarifying questions as needed. They were inductively coded and the research team agreed on the key points which informed the draft of the Round 1 questionnaire. The draft was then reviewed and extensively discussed by the Steering Committee to develop the Round 1 Delphi questionnaire. This first questionnaire was initially piloted by a GPEx ME, and then again reviewed with the assistance of Professor Schuwirth to ensure the questionnaire was clear, simple and the Delphi technique was being implemented correctly.

Subsequently, three rounds of questionnaires were disseminated to a panel of nine national and international experts, including the original DoTs who were originally interviewed. The Delphi participants were quasi-anonymous, which is to say some of the participants were aware of who some of the other participants were, but at no stage could identify answers within the rounds of questions. The questionnaires consisted of a series of statements and the participant was asked to rate aspects on a Likert scale (eg how strongly they agreed with a statement, how important they felt various elements were). The survey was built on and the results collected by Qualtrics in a de-identified manner. Any non-responders to the initial questionnaire were excluded from subsequent rounds. The second round of questionnaires was not piloted but was again reviewed with Professor Schuwirth. For those questions that did not reach consensus, anonymous feedback on the answers (including any comments) were provided in subsequent rounds.

How consensus would be defined in this project was agreed prior to the questionnaires being analysed, using a threshold percentage agreement, as has been recognised as one of the valid and more common strategies in the literature (Boulkedid 2011, Sinha 2011, Diamond 2014, Humphrey-Murto 2017). The percentage threshold was agreed to be when either 1) >70% of respondents' answers were within 2 adjacent Likert rating or 2) >90% of respondents answers were within 3 adjacent Likert ratings at one end of the scale. This was decided based on what the researchers felt was feasible given the number of respondents involved (meaning 7 out of 9 respondents would need to agree) and a literature review of existing Delphi consensus papers (Boulkedid . 2011, Sinha 2011) . Qualitative data from the comments associated with each question were discussed within the research team, and were used to refine questions or statements so that they were clearer, or to develop new questions. The first round consisted of 51 questions, the second round 30 questions and the third round 6 questions. The Delphi consensus process was ceased after 3 rounds when the researchers felt that the relevant questions where clearly answered with consensus or it seemed that consensus could not be reached. The Directors of Training (DoTs) were re-interviewed after dissemination of the recommendations from the Delphi consensus to assess feasibility, acceptability and resource investment of the recommendations.

### 5.2 Stream 2A

As part of Stream 2A, the DoTs or senior MEs in each RTO were asked to identify the ESA structure, content, and flagging procedures in their RTO. This occurred during the Stream 1 interviews.
RTO provided documents were examined to extract descriptions of the content and structure of ESAs and flagging criteria in each participating RTO. The documents presented were very diverse, with no consistency between the RTOs eg structure of ESA used for training supervisors, forms for direct observation, flagging templates etc. Questions that would assist with developing an overall view of each RTO's ESA framework and filling in gaps from the document review were followed up in the final DoT interview (Stream 1). Interview transcripts were analysed using NVivo. This was then synthesized with the gathered documents to produce rich descriptions of ESA structure, content, timing, who delivered the assessments, rating scales, flagging protocols, various templates and associated feasibility, acceptability and resource investment. The results were compared between RTOs to identify similarities and differences.

### 5.3 Stream 2B

In Stream 2B, the current prevalence of ESA flags for GPT1 registrars was collated from each participating RTO. Data collected included who flagged the registrar, when they were flagged, the severity of the flag, how many times they were flagged and the nature of/reason (domain) for the flag (clinical knowledge, communication/consulting skills, personal/family issues, professionalism, compliance, practice problems). The severity of the flags and each domain's prevalence was compared between the RTOs to determine any differences between RTOs in the prevalence of flags per domain.

### 5.4 Triangulation

Information from Streams 1 and 2 was triangulated to develop recommendations for an improved ESA and flagging system. The following questions were developed by the researchers and the steering group as the research progressed :

1. Which ESA assessments currently used in RTO ESAs achieved consensus by the Delphi participants?
2. Are there ESAs that tend to be associated with different types of flagging (e.g. professionalism, knowledge) and are there certain domains that are more difficult to flag?
3. When are registrars being flagged as part of an ESA, by whom, and using what criteria?
4. Have the recommendations from the Delphi consensus identified the best process for an ESA?
5. Which elements of the recommended ESA are deemed to be effective, feasible and acceptable and why?

Using these findings, the researchers drafted final recommendations for an improved ESA system which has been included in the results section of the paper.

### 5.5 Research participants

The experts for the Delphi consensus were chosen in consultation with the steering group, to mitigate the risk of bias, by purposive sampling based on their experience, publications and interest in the early safety of GP registrars. We set out to include 15 experts, though 9 were ultimately included due to non-response to recruitment emails or unavailability for the duration of the various Delphi rounds. The expert group included DoTs or senior MEs in each RTO, and other experts in medical education and assessment (as identified by the Steering Group). An honorarium was budgeted to acknowledge experts' time invested in the Delphi consensus.

DoTs of those RTOS currently conducting ESAs participated in interviews (Stream 2B) or delegated to a senior ME with a particular expertise and in-depth knowledge of the ESA.

### 5.6 Data collection

The semi-structured interviews were recorded and transcripts transferred securely to a professional transcription service. De-identified transcripts are held in secure, password-protected computers at GPEx and Flinders University with access only by the project team.

Data for Stream 2B on the prevalence of each type of flag (e.g. knowledge, professionalism, personal) was collected with the assistance of participating RTOs. All data was de-identified by the RTO before collection.

All project information was de-identified both within an RTO (for individuals and regions) and in the analysis of the data.

All data obtained will be held for 7 years on a secure, password-protected computer system at GPEx and Flinders University.

## 6. Research Findings

The outcomes of this research are based on the initial interviews with the DoTs, the three rounds of the Delphi consensus questionnaire, the comments collected from the Delphi experts, the documents collected from the participating RTOs, the flagging data collected from the participating RTOs, and the final feasibility interviews with the DoTs.

It was realised from the Delphi consensus that a definition of 'safety' was not as straightforward as originally thought. There was consensus that an Early Safety Assessment is important in improving patient safety, and registrar safety, and to a lesser extent practice safety. An ESA is not a one-off measurement, but more of a 'narrative', built with various assessments, resources, and expertise to encompass the ability of the registrar to use the skills they have developed in a manner that is safe for all concerned. It is a process of learning to be safe in the face of uncertainty, to self-reflect on their limitations, and to make accommodations in order to improve the way they practice and safety net in the complex environment of general practice. This is a life-long skill that an ESA will assist in developing.

Safety also needs to be contextualised to the registrar's personal circumstances, practice and the patient population. "So if a registrar develops a mental health problem or an anxiety problem or family circumstances, a registrar moves to a different practice where there are systems issues that weren't evident before. Or they move to a different cultural group, so they move to an AMS, for example, and they struggle there. So, competency is not just in the registrar, it's also in the environment the registrar is functioning." The supervision also needs to be appropriate and effective for the registrar's situation.

Registrars are often anxious about how safe they are with patients and find the concept of safety more suitable than "poor performance" or "under-performance". Some DoTs talked about "supporting and protecting the registrar" as well as protecting the patient and the supervisor and the assessments should be designed to ascertain how best to do that. "Being overt about the early safety assessment making patient safety a priority and that this is what is being evaluated. It's about setting that culture right at the beginning, and hopefully this would open up a conversation every single day between the registrar and the supervisor about patient safety…that they can carry with them for the rest of their careers and definitely for the rest of their training".

Table 1, below, outlines the mapping of the research methodology against the research questions.

*Table 1: Research metholodogy mapped against the research questions*

| Research Question | Element(s) of the methodology answering research question |
|---|---|
| What is currently included in the partner RTOs' ESAs and why were these assessments chosen? | • Stream 1- Initial semi-structured interviews and Delphi questionnaires<br>• Stream 2A - ESA Processes documentation from RTOs |
| What should be included in an ESA and how should this inform registrar flagging? | • Stream 1- Delphi questionnaires<br>• Final interviews with DoT |
| What criteria are used for flagging in each RTO, how many registrars are flagged in each domain, and is this similar or different across RTOs? | • Stream 1- Initial semi-structured interviews<br>• Stream 2B- Flagging prevalence documentation from RTOs |
| What elements of the ESA are more appropriate for flagging, feedback and design of tailored remediation? | • Stream 1- Initial semi-structured interviews and questionnaires<br>• Stream 2A- ESA Processes documentation from RTOs<br>• Stream 2B- Flagging prevalence documentation<br>• Advice from Steering group |
| What is the feasibility, acceptability and resource investment of early flagging? | • Stream 1- Final interviews with DoTs and comments from the Delphi consensus |

### 6.1 What is currently included in the partner RTOs' ESAs and why were these assessments chosen?

This question was answered by collating what is being used in the ESA in each RTO and why, from the interviews with the DoTs, the Delphi consensus, and from the documents provided by participating RTOs.

As part of the 'pre-Delphi' round, the researchers interviewed DoTs of the partner RTOs or their delegates, about what was being used at their RTO, when it was being used, how it was being implemented and why. Documents were also collected from each RTO associated with the early safety assessment of their registrars. Each RTO is different, and the ESA and flagging procedures have been developed based on the historical and demographic context of their region. This 'pre-Delphi' round also informed the development of the first formal round of the Delphi consensus. This was done in order to ensure that the whole consensus was relevant to the AGPT environment. The aim of the Delphi consensus was to determine a core portfolio of processes and assessments that could inform the future of ESAs and flagging across Australia. Hence, within this document, an attempt has been made to gather the similarities and differences in processes, assessments and outcomes in a de-identified manner.

The structure of this section is based on the questions asked in the initial questionnaire and then the subsequent consensus from the three Delphi rounds. Quotes throughout this section are from the initial DoT interviews and comments from the Delphi consensus.

### 6.1.1 Prior to commencing community placement

The Delphi consensus and the interviews revealed that the majority of RTOs started the ESA before registrars even began their community placements in GPT1 of AGPT training. Some of this was overt, for example the MCQ, OSCE-style workshop, self-evaluation questionnaire and pre-selection interview. But the more informal "watching for signs of registrars in difficulty" also begins before they start in the community.

This can include reports about competency from practice observation as an intern or a Registered Medical Officer (RMO). Some DoTs discussed the previous Prevocational General Pracrtice Placement Program (PGPPP) system where Interns and RMOs could elect to undertake a supervised period in general practice under full supervision in a supernumerary capacity. This would give insight into those doctors who are unable to learn from experience, self-awareness, or reflective practice; were unaware of the context of the consultation; had interpersonal skills deficiencies, learning difficulties etc.

Registrars can be flagged based on their pre-selection or commencement interviews, how they interact on a personal basis with the programme coordinators, whether they have problems setting up their initial paperwork, or have difficulties with the relationship with their employer before they even begin. If they don't attend or are disengaged at orientation, they will also be flagged.

For many participants, having information about a registrar before they commenced seeing patients in general practice, facilitated assessment of competence and safety. However the use of hospital-based assessments continued to be a controversial issue with others, saying that it's only possible to assess GP competency and safety once registrars have started seeing patients. As one DoT remarked:

> *"It's very hard to judge from their hospital years how they're going to go in general practice".*

While assessments prior to the commencement of community placement were suggested for inclusion during the Delphi Consensus, there were concerns about their feasibility from the DoT interviews, such as: the registrars were in a context where there was no access to internet; the registrars had no previous experience of general practice at all and so would be unable to realistically answer questions about general practice; the registrar was going straight from one job to another and so would not have time to do an assessment; an OSCE-style workshop before commencing community placement would not be possible as registrars might be in other states or in the country and this would take a whole day; an OSCE-style workshop at orientation would interrupt the flow of the workshop and the 'positive vibe' that the training organisation is aiming to generate.

### 6.1.2 MCQ (including prior to community placement)

A knowledge-oriented MCQ assessment, including one before the registrar begins their GP training, was endorsed by the Delphi consensus participants as an objective measure of gaps in a registrar's clinical knowledge to help guide learning. It is viewed as a clear-cut identifier, benchmarked against other registrars, and any specific topics where there are gaps that can then be discussed with the ME and/or supervisor to address learning needs. Those registrars who for instance score two or more standard deviations below the mean for their cohort, can be flagged as a potential concern.

Two DoTs had noticed that there is a very strong link between whether registrars choose to complete the MCQ in the timeframe or not, and whether they pass their exams. Those who don't complete the MCQ have a much

higher failure rate. Some RTOs monitor registrars throughout their training with further MCQs, including an additional assessment in GPT1 as a comparison to their pre-training MCQ.

### 6.1.3 OSCE-style workshop

The OSCE-style workshop achieved consensus from the Delphi experts, but not all RTOs were familiar with OSCE-style workshops of common GP problems. For those who use them either at orientation or very early on in training, it is a potent assessment for identifying registrars who might struggle with general practice or who have specific knowledge gaps. Registrars are given qualitative feedback about their performance on the day. The OSCE stations are run by MEs, and involve simulation of common general practice scenarios such as hypertension, diabetes, early pregnancy bleeding, drivers licence assessment, immunisation, mental health, non-cardiac chest pain, paediatrics etc. Assessment involves how much supervision they might need for a similar case in real life, as well as communication style, consultation skills, cultural awareness, safety netting, professionalism etc. The development of a differential diagnosis, dealing with uncertainty and when to ask for help can also be assessed. During the workshop day a significantly struggling registrar is likely to be flagged by multiple MEs at different stations. At the discussion meeting at the end of the day, when the MEs reconvene to discuss the registrars' performance, they may see certain registrars have been flagged several times.

Along with the self-assessment tool, their supervisor in GPT1 will see the feedback from this workshop and work together with the registrar and possibly the ME to develop a learning plan.

> *"I will maintain that the early OSCE-style workshop is extremely helpful as part of an ESA. It gives further assessment information from medical educators who do not necessarily know the registrar beforehand, and are objective assessors. It allows standardisation of cases and situations, and benchmarking of registrars to their peers. In my experience the OSCE-style workshop reinforces feedback received in other direct observation assessments, and leads to some registrars being flagged who otherwise would not have been."*

It can also give the registrar more confidence as they have been 'exposed' to general practice issues that they may not have previously seen in a hospital environment.

> *"I think the OSCE workshop gives registrars lots of confidence for their first day in general practice as well. Like, for registrars that haven't spent time in general practice, I think at the end of that workshop, they feel much more like they know what they're going to expect to see, what their day is going to look like on their first day".*

### 6.1.4 Self-assessment tool

The other assessment that achieved consensus from the Delphi experts for use before GPT1 begins was an early or pre-placement self-assessment tool. Several RTOs have developed such tools, some mapped across the RACGP curriculum and some mapped to common general practice issues. This self-assessment is based on the level of supervision the registrar thinks they need, and their ability to practice independently. For example, it might be marked on a scale from 'I would need direct supervision and assistance with this' through to 'I'm really confident in this and I could teach it', or in a more 'EPA-style' format as 'I need frequent direct in-room review by the supervisor' to 'I am safe to practise unsupervised'.

Some RTOs match an EPA style self-assessment tool with a supervisor assessment of the level of supervision they think is needed, usually starting in the first four weeks and then continuing as an ongoing assessment

throughout training. The aim of this is to generate discussions with the registrar about learning needs, which can not only identify gaps in knowledge or practice, but can also assist in identifying the over-confident or 'unconsciously incompetent' registrar.

### 6.1.5 Multisource Feedback (MSF)

The use of MSF in the assessment of safety early in GP training did not reach consensus in the Delphi rounds. Some RTOs use MSF as part of their ESA, as well as at other times in training. An external provider was used in some MSFs to collect information from patients, adminstrative staff, allied health professionals, and other specialists. Others use an internally developed and analysed MSF completed by internal staff in the practice only, based on: initiative and enthusiasm; punctuality and reliability; feedback from patients; teamwork, cooperation and attitude; communication skills; and time management skills.

Despite the fact that the literature seems to support the use of MSF as a way of exploring professionalism issues in trainees more broadly, the Delphi participants did not agree that the MSF would be a valid or feasible part of an ESA. Indeed, when reviewing the RTO documentation, it was found that the MSF did not trigger a flag for any registrar in any RTO. As one DoT interview participant remarked: "We haven't yet had anyone who would be flagged purely based on their MSF when everything else is perfectly fine. But then, I think, that probably also reflects that if there is a concern raised on their MSF, there's very likely going to be a major concern raised in their other assessments as well."

Whilst exploring whether the use of an MSF was valid and feasible, we asked the Delphi participants when it should be done and by whom. There was no consensus on when, with one participant's comments echoing the sentiments of others:

> *"I think an early MSF places unnecessary additional stress on a registrar and is difficult to interpret. This is a more useful tool once the registrar has had time to develop their consulting and has had some regular interactions with patients, colleagues and staff. Good feedback requires trust which takes time to develop".*

It was agreed that if an MSF is used then it should include information from patients, practice staff and supervisors only.

### 6.1.6 Supervisor assessment

There was agreement by both the DoTs and the Delphi experts that supervisors are the backbone of the ESA as they will have more 'exposure' to the registrar than anyone else. One participant expressed it succinctly as:

> *"real-life and the practice information back from the supervisors I think is the most important from a patient safety point of view."*

Supervisors are usually busy clinicians and few will have formal medical education qualifications. Templates and guidance documents that are clear, practical and user-friendly are essential, and there are a range of these developed by the RTOs. These include orientation checklists; relationship-building guides and evaluations; direct observation assessments; random case analysis (RCA) and case-based discussion (CBD) templates; competency frameworks; high risk/call for help lists; formal teaching guides; professionalism parameters; and flagging protocols.

The Delphi participants agreed that there should be a period of time at the beginning of the registrar's first community placement which should be completely supernumerary, when the income for the registrar is paid for by the training organisation. This would ensure that there is enough time for a full orientation, including with the nurse, the practice admin team, IT, local pharmacist and other allied health professionals etc. During this time the registrar can 'sit in' with the supervisor, the relationship between the supervisor and the registrar can begin to develop, the supervisor can directly observe the registrar consulting, areas of 'unconscious incompetence' can be revealed etc.

As stated by one Delphi participant

> *"I think the supernumerary practice stuff is essential to me. Also, it sometimes takes quite a while to figure out that your registrar is not asking you for help appropriately. I think having focus on that relatively early is important and that's going to be more likely to happen if you're reviewing every patient".*

During this supernumerary period, the high risk/call for help list can be discussed and a process agreed upon about how the registrar will ask for assistance with those patients. Such a call for help list also empowers the registrar to call on their supervisor, not just because they are unsure what to do, but because they want to be safe, they want their patients to be safe and their practice to be safe.

As one DoT remarked in the initial interviews:

> *"I think the high-risk, that call for help list is - has been good for us in orientation because it gives the registrar some power in that relationship to say here's a list that we need to go through, rather than nagging someone and finding - it gives some authority to it, which is good. Then make sure that they do ask for help".*

The Delphi experts and the DoTs agreed that discussing every patient with the registrar for the first one to four weeks either at the time the patient is seen, or at the end of the day, will assist the supervisor in deciding when the registrar is able to practise without such close supervision in various different domains e.g. prescribing, referrals, paediatrics, women's health, emergencies, mental health and chronic disease management. Some RTOs have developed a short Entrustable Professional Activity (EPA)-style list of questions that can be used by the supervisor to help with this process. Other RTOs have developed competency assessments, and checklists that clearly outline what a registrar should be flagged for and the areas where there should be direct supervision until the registrar is deemed by the supervisor to be competent. Registrars who do not appropriately ask for help or are not responding to feedback should also be identified in this process.

As well as giving ad hoc clinical advice or discussing the patient reviews with the registrar, the Delphi participants reached consensus that it is essential that the supervisor directly observe the registrar whilst consulting at least once before week two. This formal observation assessment should focus on history, examination, diagnosis, therapeutic alliance, investigations, medications, referrals, safety-netting and follow-up, and also include RCA and CBD of a previous, unsupervised consult that the registrar had conducted. The frequency of further direct observations could then be tailored to the needs of the registrar. In one RTO, the final questions for the supervisor in the competency assessment was: 'Do you have any concerns regarding GP registrar safety?' and 'Do you have any concerns regarding patient safety?'. Such a question would minimise the risk of professionalism, communication and cultural issues; over- or under-investigation or over- or under-referral; inadequate safety-netting; inadequate note-taking; serious time-management problems etc. going unnoticed. As one participant said:

*"Probably the most important thing for me was to actually observe them consult and see them with patients. Then looking at their actual consulting and using random case analysis as well to probe areas that might not be discovered just by patients that come through the door."*

Both Delphi participants and DoTs commented that the supervisor should act as a role model and mentor for the registrar, as well as advise and give feedback about clinical issues. The registrar should observe the supervisor during the first two weeks and then at other times as appropriate. A strong educational alliance with mutual respect between the registrar and supervisor will ensure that feedback, advice and the 'hidden curriculum' are heeded and useful.

6.1.7 Medical Educator (ME) and/or External Clinical Teacher (ECT) assessments

The tasks of the MEs/ECTs are different in each RTO. Consistently they are viewed as "*another pair of eyes*" on the registrar, their advantages being that they are not there every day and so bring a fresh perspective; can be more objective; and deliver more targeted feedback. They can ask about what's going on at home and in the practice, and about the registrar's training and career path, without the registrar feeling that their employment may be compromised.

The guidance documents for MEs/ECTs will thus be different than those for supervisors. The expectation from the Delphi consensus was that they will see a minimum of four patients as part of their observation assessment, and include RCA and CBD, in order to ensure a broad range of consultations and skills are assessed. The number of RCAs or CBDs was not discussed in the Delphi consensus.

This will include looking at communication skills; physical examination skills; procedural skills; professionalism; motivation; understanding or use of the roles or role conflict - any of the skills that are needed to be an effective GP. But they can also ask about the registrar's personal health; family health issues; systems issues; the practice; the environment and community; cultural issues etc. According to the Delphi consensus, the template for these visits should include:

- Asking about the registrar's personal safety
- Assessing patient safety
- Unsafe practice system issues
- Registrar supervisor relationship
- Registrar wellbeing

The Delphi participants also agreed that the ME or ECT should have a formal documented conversation with the supervisor and the practice manager about the registrar, either at the time of the Direct Observation Visit, or at another time, to inform the ESA.

They also reached consensus that the ME/ECT should undertake RCA as well, as they cannot possibly assess all the domains of general practice, and registrars are likely to be on their "best behaviour when they are being observed". It is also an opportunity to discuss issues such as professional and ethical attributes, population health and communication skills, and the template should reflect this.

### 6.1.8 Global assessment

The Delphi experts agreed that a global assessment of some sort, either formally or informally pooling information from a variety of different sources as a programmatic assessment, was important as part of an ESA. In most RTOs, a global assessment is undertaken either by the supervisor, the ME/ECT, or both. In some RTOs the documentation for this is in the style of EPAs, particularly around patient safety, and in others it is set out as broad competencies with flagging suggestions. In both models, the expectation is that the assessor will look more broadly at the registrar's progression and assess through observation and the experience of others in the practice, whether the registrar is safe to move on to less supervision. This may be more quantifiable but may start with a sense of unease. As one Delphi participant expressed it:

> *"For the supervisor, they're probably dealing with their day-to-day interactions with the registrar and some of it is that sort of gut feeling".*

In some RTOs, the registrar will also undertake an assessment, using the same EPA form as the supervisor, looking at the level of supervision they think they need which will generate a discussion between the two. Three of the participating RTOs use the EPAs developed by Valentine et al (Valentine 2019) and two RTOs use a shorter EPA-style safety questionnaire, specifically designed for use as part of an early safety assessment.

MEs will develop an even wider, more evidence-based picture of the learning and other needs of the registrar, taking into account the OSCE-style workshop, the MCQ, the supervisor reports, and then their own direct observation. One DoT commented:

> *"It's not just a gut feeling. I think it's a combination of all those different points of information and one may outweigh another if it's more important. The beauty of lots of different pixels of information allows you to have a combinational picture".*

The ME can then begin to 'diagnose' what the problem areas are for the registrar and flag them appropriately. For instance: clinical knowledge skills, communication skills, working in a team, professional and ethical issues, practice system skills, personal traits, health and family, the practice environment, the learning environment. They can then discuss with the supervisor, the training coordinator, the RTO remediation committee or the DoT, depending on the issue and the severity of the flag.

### 6.2 What should be included in an ESA and how should this inform registrar flagging?

This research question was answered from the outcome of the RTO document analysis and Delphi consensus.

*Table 2: Inclusions for an Early Safety Assessment*

| Early Safety Assessment Inclusions | |
|---|---|
| Prior to registrar community placement or at Orientation | • MCQ<br>• OSCE-style workshop<br>• Self-evaluation questionnaire about the level of supervision required in various topic areas<br>• Training of supervisors in early safety assessment |
| First 1-4 weeks of placement | • 1-2 weeks of supernumerary practice when the registrar is paid separately from the practice or their Medicare billings. This time will be used for activities such as orientation, relationship building, the registrar sitting in with the supervisor and the supervisor sitting in with the registrar<br>• Between 1 and 4 weeks of GPT1 where the supervisor reviews every patient seen by the registrar at the time or at the end of the day. The period of time and level of supervision should be tailored to the registrar based on their level of competency, safety and gaps.<br>• There should be a high risk/call for help checklist for registrars at the beginning of community placement.<br>• It is important to find out whether the registrar appropriately asks for help (not exclusively as part of the call for help list). |
| Direct Observation | • The following activities should be performed as part of an ESA:<br>  ○ A formal direct observation assessment by the supervisor<br>  ○ A formal direct observation assessment by the dedicated ME and/or ECT<br>  ○ All direct observation assessments should include Random Case Analysis and case-based discussion.<br>• The number of times a supervisor directly observes individual consultations should be tailored to each registrar (rather than being a generic number). This should include a broad range of consultations (ie paediatrics, mental health etc)<br>  ○ At least the equivalent of one session before week 2. |

| Early Safety Assessment Inclusions | |
|---|---|
| | • A minimum of 4 patients needs to be observed during the ME/external clinical teacher observation assessment but it's important to note the content and complexity of the consultations.<br>  ○ This should be between weeks 4 and 12.<br>• As well as assessing the registrar's clinical competence, the ME/external clinical teacher observation assessment should also include:<br>  ○ Asking about the registrar's personal safety<br>  ○ Assessing patient safety<br>  ○ Unsafe practice system issues<br>  ○ Registrar supervisor relationship<br>  ○ Registrar wellbeing<br>• It is important for the ME or external clinical teacher to have a formal documented conversation with the supervisor and the practice manager about the registrar, either at the time of the Direct Observation Visit, or at another time, to inform the ESA. |
| Programmatic assessment | • When a training organisation decides to flag a registrar as 'struggling' this decision would ideally be supported by 'evidence' or a rationale from observing this registrar performing certain activities.<br>• The ESA processes should continue to be in place for the whole of the first semester of community training<br>• The time when sufficient data is available from an ESA to triangulate and make a decision about whether a registrar is safe for practice is dependent on the registrar and their context.<br>• Using a programmatic assessment method across different assessment instruments, the ESA should end when the GPT1 registrar is flagged and intervention implemented, or clearly deemed as 'safe to practise with the supervisor available the majority of the time', whichever comes first. |
| Flagging | • The 'diagnosis' of the flag should be tailored to the individual but a template would be helpful as guidance to make it less subjective.<br>• Flagging should be sorted into categories of severity e.g. mild, moderate and severe.<br>• There should be a central location/database for flagging. The people who should have access to viewing the contents of the flagging data include: |

| Early Safety Assessment Inclusions | |
|---|---|
| | ○ Registrar's dedicated medical educator<br>○ Training co-ordinator<br>○ Flagging and remediation committee<br>• Registrars should remain flagged until any safety issue is completely resolved.<br>• The following people should be included in the sign off of the ESA and declare that the registrar is 'safe to practise' with the supervisor available most of the time:<br>○ No flag – ME and current supervisor<br>○ Mild flag – ME and current supervisor<br>○ Moderate flag – Regional senior ME and remediation ME (the current supervisor could be considered for this level also- however consensus not achieved on this)<br>○ Severe flag – Committee and director of clinical education and training (the current supervisor could be considered for this level also- however consensus not achieved on this) |
| Useful documents currently used by RTOs | • Self-evaluation questionnaires<br>• High risk/Call for help lists<br>• Different styles of guidance and templates for use in the global assessment and in a programmatic assessment<br>○ Entrustable professional activities either as an ESA or overall<br>○ Competency assessments<br>○ Direct observation global assessment for MEs and ECTs with flagging criteria spelled out<br>• Direct observation and RCA templates<br>• Diagnostic frameworks and templates |

### 6.3 What criteria are used for flagging in each RTO, how many registrars are flagged in each domain, and is this similar or different across RTOs?

In order to answer this question, the criteria for flagging were collected from each participating RTO from the interviews with the DoTs and from the GPT1 flagging data submitted by the RTOs. The data collected showed: how many registrars were flagged, when they were flagged, how they were flagged, by whom, and why they were flagged. Given the small sample size of flagged registrars, no inferential statistical analyses were conducted.

Flagging is generally defined as a process for identifying those registrars who need more input and support in order to complete their training as safe, effective, independent, self-reflective GPs (Prentice 2021). The formal criteria that each RTO uses for flagging obviously depends on the assessments that are part of the ESA, and what is documented, and by whom, differs throughout Australia (Ingham 2019). While there are some core similarities between RTOs; flagging criteria, thresholds, processes, pathways, and signing off varies widely nationally. In addition, in the interviews the DoTs discussed the informal processes for flagging as well as those assessments that are formally part of their ESA. These are more difficult to quantify but are equally important, especially for personal, professional and communication flags. Depending on the manner in which the RTO records flags, these issues may not have appeared in the formal data collection that is part of this research.

#### 6.3.1 Formal flagging

During the interviews it became clear that DoTs would like to have more information about the registrar's knowledge in relation to practising General Practice before they start training. This would be in the form of multiple choice and short answer extended knowledge questions online that are GP relevant. In one RTO for example, all registrars who score two or more standard deviations below the mean for their group are flagged as a potential concern. MCQs are often done at various other stages of training in order to track progression and quantify gaps in knowledge.

Some RTOs begin their safety assessment with the selection process. Flagging those who score '3 or below' on any of their Multiple Mini Interviews (MMIs) is one formal method of flagging registrars of concern who need 'watching' on future assessments. Some RTOs also conduct hour-long induction interviews where communication issues, for instance, may be identified.

The self-assessment questionnaire used by some RTOs is helpful for both the registrar and the supervisor to identify gaps and areas of learning need. Registrars might 'flag' themselves as needing more assistance in a particular domain, but while this will not be used as part of the RTO formal flagging process, the registrar can identify this as a learning goal

For those RTOs who run an OSCE-style workshop at orientation or early in community training, the feedback from the 10 to 12 MEs is collated at the end of the day. If concerns or significant gaps are flagged, a summary is put together, fed back to the registrar, and sent to the supervisor and dedicated ME.

Supervisors are the mainstay of ESAs and it is most appropriate that they 'sign off' on the registrar's safety (Ingham 2019). Most supervisors have not had formal medical education training and need training in how to give effective feedback and enhance learning opportunities, and about the importance of flagging as part of

formative and low stakes assessments, especially when assessing learning needs and safety. For busy supervisors, a process for thinking in a more global manner about the elements of safety that are important, and templates for assessing these, can be very helpful. Supervisors will need to change their usual way of thinking about assessment if they are using EPAs, which rely on the level of supervision that is recommended in a particular domain rather than on reaching a pre-determined standard of competency.

The use of an EPA-style assessment (such as the 'Safe Seven' used by two RTOs) moves from this pass/fail mentality to one tailored to the supervision needs of individual registrars. The ME will also be involved, either in discussion with the supervisor about the supervisor's assessment, or when the ME does the assessment themselves and discusses flags or concerns with the supervisor. One RTO uses formal competency assessments, with clearly defined unacceptable behaviours, and guidance for giving a 'green', 'amber', or 'red' flag (Prentice 2021). A minor, watching or amber flag will mean that the supervisor and/or ME undertake other assessments earlier than they otherwise would in order to ascertain if there is an issue that needs support, intervention or remediation. A major, active or red flag means there is a serious concern about patient safety, a knowledge gap, a behavioural issue, professionalism, a personal concern or a communication problem.

The templates for directly observing the registrar consulting will usually include questions about registrar and patient safety that will trigger a flag. Some have more guidance, particularly for MEs, with a broader range of flagging criteria such as: consultation skills, knowledge, clinical reasoning, procedural skills, physical examination skills, professionalism, personal issues practice, family, personal and family health, cultural issues, communication skills, environment, personal safety, diversity of training, career, exam failure and role conflicts.

Global assessments undertaken by supervisors, MEs, and often registrars as well, will assist in quantifying the 'gut' reaction that a registrar is in difficulty. In three RTOs, this is in the form of 12-13 EPAs questions answered at mid- and end-term by supervisors and registrars (Valentine 2019). This can flag specific issues or domains where the registrar needs more supervision and/or assistance, and replaces the more generic 'Levels of supervision' (Ingham 2019).

Programmatic 'triangulation' of assessments, either by the ME or a committee, is usually done at a particular time, such as week 8 or week 12, in order to flag registrars in difficulty. However if significant issues or red flags appear earlier, particularly from the supervisor, the ME or committee will step in.

At the point when all of the assessments have been completed, a bigger picture of a series of minor flags may emerge. An Assessment Committee will deal with major flags and develop an action or remediation plan for the registrar, monitored by an ME. The documentation about the flag will assist in the 'diagnosis' and 'management' or remediation plan for a registrar in difficulty.

### 6.3.2 Informal flagging

As the 'hidden curriculum' is to professionalism, the informal flagging processes are just as important as the formal assessments. Two DoTs commented for instance that it wasn't actually the score in the MCQ that was the more serious flag, but whether the registrar attempted or completed the MCQ in the timeframe, as this predicted a higher failure rate in the exams.

Lack of engagement and compliance with educational activities is often one of the most important pointers to deeper issues that will lead to the flagging of a registrar. This could be by the administration staff at the RTO (training coordinators) and involves a formal escalation pathway to the supervisor, ME, flagging committee or DoT. All RTOs had a process for training coordinators to flag registrars because of compliance or behavioural issues, though this is not usually a formal part of an ESA.

Professionalism is another example of an important flag, but is much more difficult to formally assess. One DoT stated that:

> *"Probably the commonest one is professionalism. The knowledge one does come up, that's one of the easier ones to deal with, although often … the knowledge one is a result of a professionalism issue."*

Another thought that the RTO had a

> *"pretty good visibility of clinical performance",*

but

> *"conduct and behaviour are much harder. I don't think we've got a very good formal way of collecting that… we kind of hear a bit of that informally. So I think that's potentially a bit of a gap".*

Similarly, serious health issues that can impact training and practice can often be missed.

> *"We've certainly got registrars with significant health problems that we're aware of and they're very open with us about that and hopefully, we support them. I suspect we've got some with health issues that we don't know about."*

Another DoT thought that the relationship between the registrar and the supervisor was one of the most important flags:

> *"Most of the ones that I deal with are problems, or difficulties in the employer-employee relationship or some problem at a practice level and universally or a large portion of those have a personal basis to them."*

Flagging an issue about the relationship with the supervisor can also be difficult, often identified by the ME. 'Failure to fail' and supervisors 'giving the registrar the benefit of the doubt' means that even what looks like robust formal Early Safety Assessments can miss registrars in difficulty. For many supervisors, giving an assessment that finds that the registrar is 'incompetent' or 'underperforming' is difficult, as they say such things as "It's early days, let's give them a chance" and are likely to excuse the problem (Gingerich 2020). This of course also depends on the personalities and the relationship between the registrar and the supervisor, a good relationship being a predictor of more effective feedback and support –

> *"if we could continue to encourage a non-judgemental interaction between supervisor and registrar that allows for transparency and accepts that everybody brings their own uniqueness to whatever the situation is, and that it might influence their performance - so, what can we do to help?"*

All RTOs acknowledge that supervisors can also flag registrars separately from the specific assessments. Training coordinators and practice managers can bring struggling registrars to the RTO's attention as well. This might be for issues such as professional issues, behavioural issues, extended leave, or punctuality. Those RTOs that used formal Multi-Source Feedback (MSF) as part of their ESA, had not seen any registrars flagged purely via this formal process, but did have informal flagging occur from other sources.

Programmatic assessment goes some way towards addressing the issue of informal flagging:

> *"But when you add up all the elements of, oh, they've been given a chance here and given a chance here and given a chance here, then you're like, well, you told them that feedback four weeks ago. I saw them do the same thing again today. Then that paints a picture."*

### 6.3.3 Flagging data

As expected from the above discussion, the data from each RTO differed markedly from the others. The percentage of registrars who were flagged varied from 9.7% to 13.6%. A total of 74 of the 650 registrars who were part of this data collection (11.4%) were flagged - 46 (62%) as 'watch' or amber flags, and 28 (38%) as active flags. How many flags were 'watch' and how many 'active' differed across the data, with one RTO having 70% of the flags with some specific activity associated with the registering of the flag, and another having only 9% as active flags.

In all participating RTOs, only 30-35% of the flagged GPT1 registrars were identified in the first 12 weeks of community placement and 45% overall were identified in week 19-26. The majority of flags (90%) were by Supervisors, MEs and ME/ECT visits, with other flags identified prior to placement, at workshops, and by office staff. It is unclear from most of the data from the RTOs whether flags were triggered by formal ESA assessments.

Overall, 33 (45%) of the flags were concerning clinical knowledge, 28 (38%) concerned communication/consulting skills, 15 (20%) concerned personal or family issues and 9 (12%) were flagged due to professionalism issues. Other flags were for compliance or practice problems. One RTO had no registrars formally flagged due to personal/family issues as compared to another where 40% of the flags involved personal or family issues.

Given the importance of early flagging for the registrar's future practice, personal health and exam success (Prentice 2021), 'early' safety could be expanded to include the whole of GPT1, and not just the first 12 weeks. The identification of personal and family issues that might be the 'cause' of clinical or communication problems, and the 'diagnosis' and active management of flags should be implemented as part of any future early safety assessment.

## 6.4 What elements of the ESA are more appropriate for flagging, feedback and design of tailored remediation?

While we had originally hoped to develop an answer to this question from the data, the Delphi consensus and the interviews, it has ultimately proven difficult. This is due to the small number of registrars who were flagged; the diversity of elements used in an ESA across the stakeholders; the range of reasons for flagging and how registrars are flagged; and the decision by the steering group that remediation was not part of our remit. Given

all of these factors, we ultimately decided that we could not answer this question. The decision was discussed with the Steering Group and it was agreed that this could not be included in the final report. The Delphi consensus agreed on which elements should be included in an ESA (for instance that MSF should not be included), and the data identified that most flags were by the supervisors and MEs, but a link could not be made with giving feedback or remediation.

### 6.5 What is the feasibility, acceptability and resource investment of early flagging?

The answer to this question was mostly from the final DoT interviews, as well as comments from Delphi participants. A series of questions were developed based on the TELOS model of assessing feasibility (https://www.mindtools.com/pages/article/telos.htm). DoTs were asked about their thoughts about the recommendations proposed after the completion of the Delphi rounds and whether the ESA would serve its purpose. TELOS is an acronym for:

- **T**echnological.
- **E**conomic.
- **L**egal.
- **O**rganizational.
- **S**cheduling

The following points were raised during the interviews.

*6.5.1 Technological*

Technology can be a facilitator or a barrier, and its use is dependent on the context of the placement and the skills of the supervisor, ME and registrars. Most DoTs commented that some MEs and many more supervisors needed assistance with online forms and uploading of documents into the system.

Similarly, with the advent of virtual observation assessments, there are pros and cons. They allow flexibility and save travelling time and money. However as one DoT mentioned “There is still a small group of MEs and ECT visitors who are resistant…and feel like it’s not quite good enough that you’re not in the room…It’s not perfect but it should be good enough”.

An online system also needs to be kept secure, and often this means password changes and updates that can be a barrier for supervisors who are not regular users of the system. “Need to find a happy medium between keeping registrars’ information secure and making a system accessible to supervisors who only log in every 3 months”.

During the interviews, there was no discussion of the use of technology from an RTO perspective to access and collate data or to aid in the capturing and examination of flags. The absence of these discussons was noted by the researchers.

*6.5.2 Economic/time*

With time poor supervisors and MEs, ensuring prompt completion of the ESA can be difficult. This is particularly the case for supervisors and registrars in areas of work-force need, who are likely to be overwhelmed with clinical work, and struggle to find the time to spend with the registrars.

> *"I think supervisors working in areas of need don't have two hours a day to be talking to registrars. So, there are supervisors that do have that time, you know, very established doctors probably in well-serviced private billing settings who've got the time and resources to put into teaching".*

One of the points of difficulty raised by some of the DoTs was the relationship between the supervisor and the ME, particularly around who 'signs off' the ESA. As one said

> *"If sign-off depends on waiting for an ME to triangulate all the data, then this may not happen in a timely manner. It depends on the availability of the ME".*

Payment for supervisor time will always be a contentious issue as 'good' supervisors will often spend more time with a registrar than they are remunerated for. They will think about ways of improving not only their teaching, but also their relationship with the registrar.

> *"We know there's lots of variability in the amount of time spent on orientation within a practice. Some practices would spend a whole day and some practices would spend half an hour. We know that some supervisors do direct observation for 30 minutes and others do it for three hours in that first week. I think it's quite difficult to make a time estimate because you've got such a spectrum of performance of your supervisors anyway, but if you're genuinely going to do it properly, to me it's a significant increase in the amount of time. At the moment it's three hours a week, isn't it, that you get for first term registrar supervisor teaching time and I think that's woefully inadequate in the first few weeks at least".*

Administration and support time was not discussed in the feasibility interviews although it would be expected that this should also be considered.

There needs to be flexibility in the arrangements so that good supervisors are able to tailor their assessments to their own practice and the needs of the registrar.

> *"When you look at [it] from an education perspective, for me as a supervisor, if I sit in with a registrar for one patient, provide some suggestions and guidance, go away, come back a week later, sit in with another patient, look at what they've used and then build on it, if I have six patients over a period of six weeks and then I've provided progressive training over six weeks, that registrar will have received far more than if I sit in for one session of six patients with the registrar's capacity to learn only one or two things from those six patients"*

It was important to the DoTs that if an ESA becomes mandatory, then appropriate payment for the time requirement, particularly for the supernumerary time at the beginning of community placement, must be included.

*"That takes quarantine time when both the registrar and the supervisor aren't earning income for the practice."*

*"But wherever the payment comes from, I think it needs to be adequate, it needs to recognise the amount of time that's involved in doing that work. It's not just filling in the form because if you're putting important things in those forms, then hopefully supervisors end up having conversations about those with the registrars, not just writing it down on paper for the registrars to see in their portfolio."*

*6.5.3 Legal*

DoTs were concerned that designating a registrar as 'safe' with an ESA will have legal implications. They were especially concerned that some supervisors may not adhere to the level of supervision suggested in the ESA, and that if a critical incident occurs then they would be liable. This added to the discussion about what the ESA should actually be called and how 'safety' is described.

*"From a legal perspective, something we have to be careful with is that if we dictate or declare a certain approach to supervision, (unless we have the medical educator signing off, that helps) but if we've said that these are the processes that need to be followed to move a registrar to a lower level of supervision and a practice doesn't follow that process and there is a - for example, a patient-adverse outcome, that could have legal implications for the practice. Because if they're deemed to have not followed a process that the college has instigated, then there is a legal liability that we might be exposing practices to".*

Contracting and funding for supernumerary time were also discussed as components that may not be feasible and practice compliance would need to be monitored. There is the risk of the training organisation if the ME flags a registrar, but that this is not agreed to by the supervisor/practice, or maybe the registrar. A policy and process would need to be developed to navigate the possibility of disagreements and resolution.

Although not mentioned within the interviews, there were other factors which the research team identified may need to be considered under legal feasibility of the proposed ESA. Contracting and compliance frameworks already exist between training oganisations and practices to ensure that the relevant standards are met. Aspects of the proposed early safety assessment would need to be built into these frameworks. For example, if funding is provided to a practice to enable a registrar to have additional time spent during the orientation before commencing consulting, there would need to be specific requirements built into the College standards associated with what must happen within this time, and an in-built reporting and compliance monitoring.

*6.5.4 Organisational*

DoTs were also concerned that there would be an increased burden of paperwork for supervisors, MEs and training coordinators with an ESA. This might then add to the legal and ethical risk if the paperwork is not processed in a timely manner.

They also mentioned that some supervisors would like to have a registrar in their practice but did not want to do the administrative tasks necessary to be a supervisor and may not have the IT skills to do this easily.

*"If there are a large volume of reports and the ME or the admin person has to read every one of them, then manpower is just not going to be sufficient. But if you're collecting data that no-one looks at, then that's a legal and ethical risk. There needs to be a process of either filtering or flagging such that important issues are actually captured".*

*6.5.5 Scheduling*

The personnel and infrastructure needed to ensure that some of the assessments run in a streamlined manner may be difficult on a large scale. This would apply to the OSCE-style workshop for instance as it would be difficult to extrapolate to a larger cohort, such as when the Colleges take over training.

Several DoTs liked the idea that the ESA was over the whole first term or equivalent of community placement.

*"I think it often takes sufficient clinical exposure for some things to be uncovered about lack of safety or other concerns" and "You just have to accumulate enough time and experience and exposure to see what's happening."*

## 6.6 Barriers and Enablers

### 6.6.1 Barriers

As part of the Delphi consensus, a list of possible barriers to an ESA was developed. These will need to be addressed in order for the successful implementation of an ESA.

- Supervisor engagement
- Funding
- Lack of resources (people)
- Lack of supervisor time
- Geographical location
- Patient 'Did not attends' during a ME/external clinical teacher observation assessment visit
- Supervisor's reluctance to make a judgement
- Bureaucracy/paperwork

Additional barriers were discussed by the DoTs both in the initial interviews and the final interviews.

*Fluctuation of registrar competency*
There can be a fluctuation on different days with different patients about the level of supervision a registrar will need. This means that any direct observation may not represent the actual safety of the registrar one way or another. It also depends on the complexity of the consultations and the length of time the registrar has in which to conduct the consultation. Early in their time in a new practice, registrars will often have easier 'walk-ins' and more time to see them.

*Supervisors*

RTOs need to select practices that have 'good' supervisors and processes, and then ensure that the supervisor feels that their assessment is respected. The supervisor is expected to sign off on an ESA as they are actually the ones who know the registrar the best. However in many busy practices, the supervisor may not have been able to spend enough time with the registrar, and their assessment may not be a realistic assessment of the registrar's safety. Such 'failure to fail' can also occur for a myriad of other reasons such as the supervisor wanting to be 'kind' or thinking they will 'get better with time'.

Supervisors will also need time and training (preferably in small groups) if assessments or processes are going to change. For example in one RTO where assessments changed from a grid with expected levels of competency to EPAs:

> *"with the EPAs, was the idea of assessing someone against fellowship standard rather than against their stage of training. So, the idea of saying to your GP1 registrar that they were below the standard expected of a fellow, they found that really difficult. In reality, that's what GPT1s are and we need to help them get them up to that standard. So, unless we say they're not at the standard, we're going to struggle to find what it is we need to do for them".*

*IT and Processes*

One of the difficulties of making a check-list, is that you have to exclude things that might also be important and

> *"invariably, the things that are excluded are the things that are hard to measure"*

Supervisors are all different. For instance, some just want a tick-box, and some like to use a global assessment. Some find that IT requirements can be difficult and would prefer to write their assessments by hand, or ring someone if they want to discuss a registrar in difficulty. Many will only use the IT system when it is absolutely necessary and do not develop the skills to use it effectively.

Any ESA should have a certain amount of flexibility built into it to take this variability into account, but without compromising best practice, accountability or safety.

*Bureaucracy*

Comments from the DoTs were that good supervisors and good practices who nurture good relationships with their registrars, do not need templates and assessments to guide what they are already doing. However some supervisors, such as those extremely busy ones or those who struggle to be engaged with the supervision process, may need more guidance and documents. However, one DoT mentioned that by standardising the paperwork requirements this will feel like the RTO is penalising the good supervisors and the good practices, by making them do more paperwork to document what they're already doing.

> *"You need a process that really encourages supervisors to watch their registrars and to be very early on, being responsible for that. So, something like this, I think achieves that. I'm a little bit concerned about - like I said, external judgments and bureaucracy running it without proper data. I'd rather be supporting that relationship with a supervisor and the registrar to sort a lot of that out and then flagging if they need more help. I don't want to over-burden registrars and supervisors with things if it's going well".*

6.6.2 Enablers

Supervisors

The crucial role of the supervisor in any assessment of a registrar's safety was reiterated many times. This included the importance of the relationship between the supervisor and the registrar.

> *"The most central component is the supervisor/registrar relationship and the ability for the supervisor to directly interact and see what the registrar's doing".*

The DoTs also emphasised that major changes take time for busy supervisors to learn. An example given was the change from competency standards to EPAs as a formative, low-stakes assessment. The tools and processes provided should guide the supervisor to reflect on the registrar's safety and progress.

> *"The EPAs took us probably two years to really embed. It took the supervisors quite a long time to grasp that concept. But they all love it now and it's really well accepted. I think they find it really useful and we get so much fantastic feedback from a lot of the supervisors. Really insightful, diagnostic, really good feedback. So, I think that absolutely they're capable of observing and assessing and interrogating problems and articulating them. I think it's more practical things like finding the time to do it and uploading it that they struggle with."*

For many supervisors, flexibility within the assessment guidelines, is also a facilitator. So when discussing the recommendation that supervisors directly observe a registrar for at least one session before week two, one DoT suggested that this should be 'the equivalent' of one session as:

> *"When we're making assessment judgments, if I sit in with a registrar and I start formulating some ideas, I go away, think about it; I sit in with them again tomorrow, I've processed some of my thoughts and concerns from yesterday that I then use the next session to either validate or next cases to validate or extend my concerns".*

Although not discussed during the interviews, it would be useful if the ESA process is built using technology that is acceptable and simple to use by supervisors, MEs and RTO staff. The technology should be able to communicate the ESA requirements, provide a platform for completing and viewing assessment feedback, assist to collate feedback across assessments and identify and monitor flagging.

## 7. Final recommendations

These final recommendations are based on the pre- and post-DoT interviews, the Delphi consensus, the RTO documentation and the flagging data collected.

1. Training
   It is essential that supervisors, MEs and registrars have training on the processes and importance of an ESA as a safety assessment with consequent support. This will include the use of the assessments, particularly if it involves a new assessment such as EPAs or a competency assessment. Training will

also include such issues as: the relationship between the supervisor and registrar; support so that supervisors do not 'fail to fail'; ensuring there is adequate time for direct observations, teaching and mentoring to take place; how and why registrars are flagged etc. The hallmark of a successful ESA is where communication between all parties aids the registrar in understanding and communicating their gaps, that there is adequate support and processes in place to assist them to address those gaps, and that they develop skills in assessing and facilitating patient safety into the future.

2. Prior to commencement
   Before the registrar begins their community placement, an MCQ and self-assessment will help guide the registrar, supervisor and ME about where their gaps are. A call for help list and education plan can be developed based on these parameters and a standardised template.
3. The first 4 weeks
   This time will be tailored to the needs of the registrar, their gaps and competency, and the context of the practice. There should be 1-2 weeks of supernumerary practice when the registrar and supervisor are paid separately from their practice or Medicare billing. This will be for orientation, relationship building, shared consultations between the registrar and supervisor, discussion about the call for help list etc. Communication strategies should be established during this time.
   In the first 2-4 weeks the supervisor should review each patient seen by the registrar – initially before the patient leaves the practice, and then at the end of the day.
   The supervisor should pay particular attention to whether the registrar is asking for help appropriately.
4. Assessments
   It is recommended that the following assessments are undertaken:
   1. Knowledge oriented MCQ and self-assessment prior to commencing community practice.
   2. Supervisor direct observation before week 2, at least the equivalent of one session.
   3. ME/ECT direct observation between weeks 4 and 12 with a minimum of 4 patient consults.
   4. Global assessment triangulating information from a variety of sources.
5. Supervisors
   The supervisor should have easy access to the relevant process documentation and templates including: the high risk/call for help list, parameters for flagging, the diagnostic frameworks for flagged registrars, the processes for direct observation, random case analysis and case-based discussions, and how a global assessment can be made. It is important that IT support is provided, if necessary, and that there is as little bureaucracy as possible. There should be adequate payment for supervision.
6. MEs
   In addition to a clinical assessment, during the direct observation of the registrar, the ME needs to assess the registrar's personal safety, patient safety, unsafe practice system issues, the relationship with the supervisor, professionalism, communication skills and registrar wellbeing. They should also have a documented conversation with the supervisor and practice manager.
7. Flagging
   The process for flagging should be transparent and supported by 'evidence'. A 'diagnosis' should be made using a programmatic assessment method to aid the registrar. Any support or additional

assessments should be individualised to the diagnosis and the particular registrar and their context. Templates should be in place to guide supervisors and MEs in this process. Flags should be graded (for example: mild, moderate or severe) and a process should be developed clearly articulating who is responsible for addressing the flag, what needs to be done and how it will be signed off.
There should be a central data-base where flags are stored that is accessible to the registrar's ME, the training co-ordinator and the remediation committee.
Registrars should remain flagged until the safety issue is resolved.

8. Time-line for an ESA
   An assessment of safety for the patients of a new registrar should be in place for the whole of their first term. This acknowledges the change when doctors move from hospitals to working in the community, with the increased complexity of patients, uncertainty, and decreased onsite support in general practice. It will also take into account the changing circumstances and patient load during the first term, and the need to learn the life-long 'skill' of self-assessment of safety.
9. Guidance documents and templates
   Templates and process documents should be developed to ensure that the above assessments and outcomes are strategic, evidence-based and practical. There should, however, be flexibility within these according to the registrar's needs and the context of the practice. Many current RTOs have documents in place that would be useful in this regard.

## 8. Limitations

Remediation was not part of this project, but the 'diagnosis' of the flag is the beginning of that process. The threshold for remediation, who does it, how it is done, and the outcomes expected also differ widely between RTOs.

The number of registrars flagged was too small to ascertain which assessments might be most useful for flagging, feedback and/or remediation.

While DoTs were interviewed they may not have considered all aspects of feasibility, as they may not be responsible for the technology and administration for instance.

The Delphi consensus group did not have as many participants as expected as this project was run during the Covid-19 pandemic, which may have affected the uptake of the offer to be part of the Delphi consensus.

The project was also run at a time when the RTOs were transitioning their training to the RACGP and ACRRM, which may have affected their enthusiasm for discussing early safety assessments. Having said this, the DoTs were all very passionate about their own way of running ESAs but acknowledged that the new way of doing things would be different.

## 9. References

Boulkedid, R., H. Abdoul, M. Loustau, O. Sibony and C. Alberti (2011). "Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review." PloS one **6**(6): e20476-e20476.

Byrnes PD, Crawford M, Wong B. Are they safe in there?: Patient safety and trainees in the practice. Australian family physician. 2012;41(1/2):26

Diamond, I. R., R. C. Grant, B. M. Feldman, P. B. Pencharz, S. C. Ling, A. M. Moore and P. W. Wales (2014). "Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies." J Clin Epidemiol **67**(4): 401-409.

Gingerich A, Sebok-Syer SS, Larstone R, Watling CJ, Lingard L. Seeing but not believing: Insights into the intractability of failure to fail. Med Educ. 2020 Dec;54(12):1148-1158. doi: 10.1111/medu.14271. Epub 2020 Aug 5. PMID: 32562288.

Humphrey-Murto, S., L. Varpio, T. J. Wood, C. Gonsalves, L. A. Ufholz, K. Mascioli, C. Wang and T. Foth (2017). "The Use of the Delphi and Other Consensus Group Methods in Medical Education Research: A Review." Acad Med **92**(10): 1491-1498.

Ingham G, Morgan S, Kinsman L, Fry J. Are GP supervisors confident they can assess registrar competence and safety, and what methods do they use? Australian family physician 2015;44(4):236.

Ingham G, Plastow K, Kippen R, White N. Tell me if there is a problem: safety in early general practice training. Education for Primary Care. 2019;30(4):212-9.

Miller GE. The assessment of clinical skills/competence/performance. Academic Medicine. 1990;65(9):S63-7.

Morgan S, Ingham G, Kinsman L, Fry J. Clinical supervision using random case analysis in general practice training. Educ Prim Care. 2015;26(1):40-6.

Prentice S, Kirkpatrick E, Schuwirth L, Benson J. Identifying the at-risk General Practice trainee: a mixed-methods meta-analysis of General Practice registrar flagging. 2020; Adv Health Health Sci Educ Theory Pract. 2021 Aug;26(3):1001-1025

Sinha, I. P., R. L. Smyth and P. R. Williamson (2011). "Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies." PLoS Med **8**(1): e1000393.

Valentine N, Wignes J, Benson J, Clota S, Schuwirth LW. Entrustable professional activities for workplace assessment of general practice trainees. Med J Aust. 2019 May;210(8):354-359. doi: 10.5694/mja2.50130. Epub 2019 Apr 11. PMID: 30977150.

Wearne SM, Magin PJ, Spike NA. Preparation for general practice vocational training: time for a rethink. Medical Journal of Australia. 2018;209(2):52-4.

World Health Organisation. Health promotion and disease prevention through population- based interventions, including action to address social determinants and health inequity: World Health Organisation; [Available from: http://www.emro.who.int/about-who/public- health-functions/health-promotion-disease-prevention.html].