National Selection to Trauma and Orthopaedics training in England, 2020

Dear Trainee,

The last few weeks have been difficult for us all, but particularly for those of you who were hoping to enter orthopaedic training during 2020. The frustrations that have accompanied this year's selection process have not been confined to those of you affected personally but have been shared by everyone involved with the delivery of training, albeit to a lesser extent.

Covid-19 has meant that we are all working in unusual and at times highly stressful environments and there was never any desire to add to those stresses by delivering a less than perfect, albeit well-intentioned selection process. It would not be unreasonable at this stage for us all to reflect that the process as a whole could have been better managed, and whilst much of the recent criticism that has come to our attention appears justified, to reflect constructively it is important to understand in detail not only the way in which this year's process has been modified, but also the way in which the selection process normally works. With this knowledge, which is provided below, we hope that those of you unsuccessful this year will be able to prepare effectively for national selection next year.

Orthopaedics is a popular specialty without any historical difficulty in attracting trainees although recent years have seen a slight decline in the number of applicants deemed appointable. Nevertheless, the selection process remains highly competitive and as with any similar process, there will inevitably be some applicants who feel that the system is unjust, particularly given the changes imposed by Covid-19. To expect otherwise would be unrealistic.

Traditional Selection Design

The responsibility for delivery of the selection process lies with Health Education England (HEE) but much of the process is devolved to the T&O Selection Design Group (SDG) who work closely with the dedicated administrative staff at HEE Yorkshire and the Humber. The SDG is made up in part by consultant trainers appointed through a competitive process led by the Chair. Many of the members also sit on the Specialty Advisory Committee (SAC) and all members have an impressive record of involvement with the organisation and delivery of training. The Group also includes four senior trainee members appointed through BOTA who play a role of equal importance. In addition to the time spent in Leeds, the Group meet on two or three further occasions in the summer and late autumn to plan the following year's process. Each member therefore dedicates up to nine days per year to selection and in many cases, this represents leave not recognised by their employer. Each member is highly aware of the serious nature of their role: the decisions they make will determine the careers of many young doctors and will determine the future direction of the specialty.

The annual selection process undertaken in Leeds around Easter has followed a similar format in recent years. It is rare for the process to alter significantly but where minor changes are deemed beneficial such changes are made, one recent example being the removal of the controversial "killer" station. We want to appoint trainees who will become the best consultants, but we recognise that we can only achieve this goal if the process can both fairly and effectively identify applicants with the maximal potential to benefit from training. For those unfamiliar with the process, the assessments are normally undertaken within five stations, where eleven separate qualities are assessed, these including for example, situational awareness and communication skills. Many of the qualities are assessed at multiple stations and there is a robust quality assurance process, not only during the assessments, but also at the design stage and at the outcome analysis stage. It is also important to understand that the process does not aim to replicate the MRCS exam: a far wider range of qualities is assessed with limited emphasis on the recall of knowledge which feeds into only 24% of the total marks available.

The maximum possible score in recent years has been 302. A threshold appointability score is determined in accordance with detailed marking descriptors, a process not dissimilar to that used in the FRCS (Orth) exam. Applicants are ranked according to their score and a pre-determined tie-break mechanism exists to ensure that no two applicants are ranked equally. This is necessary to ensure that programme allocation can be determined according to performance, and to ensure that there is a clear cut-off between successful and unsuccessful applicants.

Part of the annual process review involves the analysis of the results and a careful assessment of the value of each of the stations is undertaken. This has in the past led to the recognition that for some stations, or parts of stations, the results reflect more accurately, or less accurately, the overall outcomes and occasionally this has led to minor adjustments to the weighting of the component parts of a station. Furthermore, there is a careful analysis of the scores according to demographic group, those assessed ranging for example, from gender, to ethnicity, country of graduation and age. Whilst we are mindful of the need to meet the requirements of the Equality Act, 2010, we feel that a more important aim is to run a selection process that encourages diversity within T&O, and whilst in recent years the data reviewed suggest that this is being achieved, we acknowledge that more work can be done within the wider recruitment process, in accordance with the developing BOA Diversity and Inclusion Policy.

The revised 2020 selection process

It became clear during early February this year that the impact of Covid-19 would affect the selection process, primarily because it was felt that both the assessors, who often number over a hundred, and the applicants may well be required to support their local hospitals in providing enhanced services for patients affected by the pandemic. A modified process was necessary, and although each specialty was asked to provide its preference through JCST (Joint Committee on Surgical Training), it rapidly became clear that HEE was in favour of one that used the information contained within the application form without any face-to-face assessment. Our view was that given our confidence in the traditional process, we were reluctant to use an unvalidated, truncated process to appoint to National Training Numbers (NTN), and our expressed preference was to appoint to LATs (Locum Appointment to Training) using the self-assessment score component of the application but with validation through portfolio review. This had a number of attractions, not least the removal of the risk of sub-optimal appointments to NTNs through an untested process, but also the ability for successful applicants in 2021 to use their LAT experience to count towards CCT.

This final decision of HEE, as everyone is aware, was to proceed with National Selection to NTN using the unvalidated self-assessment score. It is important to reiterate that this was neither our preference nor our decision, although we were able to appreciate why the priority for HEE was to deliver a relatively uniform selection process across all medical specialties with minimal additional manpower input.

The self-assessment score

This now brings us onto the issue of the use of the self-assessment score and whilst never advocating this as a sole method of selection to NTN, we were able to use historical selection data to show that a good correlation exists between this and the overall score. This provided some reassurance that if used with an appropriate threshold appointability score, the process was unlikely to lead to the appointment this year, of applicants who would have scored less than the threshold appointability score last year. The maximum possible self-assessment score is 32, and the threshold score was set, after much deliberation, at 21. This was predicted to carry a reassuringly low risk of appointing an “unappointable” applicant, but with a threshold score of 20 the risk would have increased and at 19 would have increased considerably. It is important to understand that we did not take the view that all applicants who scored less than 21 were unappointable, the value of this score simply relating to the acceptably low risk of such an appointment.

We predicted that a threshold score of 21 should lead to around 105 successful applicants. Although this represented a probable shortfall in the number of posts that were likely to be declared, we felt that this was justified for two main reasons, firstly that with an unvalidated process we should err on the side of caution in terms of the number of appointments, and secondly, that there was a strong likelihood that a considerably greater than normal number of ST8s would wish to remain in the training system, either because of the cancellation or postponement of post-CCT fellowships, or because of the need for training extensions (what we now know as Outcome 10.2).
Having satisfied ourselves that the process was safe, we wanted to have confidence that the process was fair. As outlined above, historical data exists to allow comparison of scores according to demographic group, and requests were made for their release. Unfortunately, HEE was unable to provide these data, but we were eventually able to retrieve evidence that satisfied us that there is unlikely to be any gender bias.

Academic output

A particular issue that has emerged over the last few weeks appears to relate to the way in which research output is incorporated into the self-assessment score. It is important to again reiterate that the self-assessment score was never intended to be used as the sole selection method, and that it was not our preferred modified selection method.

As originally intended, the self-assessment score fed into the Portfolio Station and represented roughly 10% of the overall assessment score. As outlined above, when designing a selection process, a wide range of qualities need to be considered, only one of which is academic achievement and of the five stations normally used it is only at the Portfolio Station that academic output is assessed. It is inevitable therefore that when using the self-assessment score as a selection tool in isolation, a greater emphasis than originally intended is placed on academic output. Being aware of this, and being permitted to have some, albeit limited, input into the process, we were able to minimise the influence of academic achievement by prioritising clinical questions when ranking the questions for tie-breaker purposes.

The use of the denominator

A further issue that has re-emerged this year is the way in which the denominator “N” is used to calculate the score relating to academic output, but it is important to understand why it is used. If we use absolute numbers, bias clearly swings towards the trainee with many years of experience, and away from more junior trainees. We feel that credit should be given to the trainee who can demonstrate a greater rate of accumulation of output and this is provided by the use of a denominator. It is difficult to defend the view that the principle of the denominator is unfair, but it is possible to be critical of the way in which it has been implemented. A banding system is used to calculate “N” and it is inevitable therefore that there is an element of bias, both favourable and unfavourable, for trainees at the boundaries of the bands. The influence of this bias is minimal if the self-assessment score is used as intended but becomes amplified roughly ten-fold if used alone. We were able to mitigate this to a degree this year by down-ranking for tie-breaker purposes the questions where the denominator was used as referred to above, but that represents the limit of our influence.

Length of experience

Criticism has also been applied to the use of the score relating to the length of experience in T&O. Again, it is important to understand why this is used. When both the overall scores and the individual station scores are correlated with the length of T&O experience, there tends to be an initial rise, a short plateau, followed by a steady fall. This shouldn’t be particularly surprising as the chance of success for an applicant would be expected to fall with increasing number of applications. Thus, by using a length of experience score, weighting is effectively given to a positive predictive factor. Furthermore, there are some groups, for example military trainees on deployment, who spend long periods of time away from the educational environment and are thus penalised as result of being unable to accumulate any academic output. We want to avoid negative selection bias

relating to this type of trainee and this is achieved to some extent by giving credit for having optimal exposure to T&O.

Results announcement

We now turn to events of Wednesday 22nd April. HEE released the results of the selection process at 12.00h and it rapidly became apparent that the number of successful applicants was surprisingly low. In fact, a total of 51 applicants had been told they were being offered a post, and large numbers of seemingly good applicants were told that they had been unsuccessful. It was brought to the attention of HEE that there may have been a score calculation error, and this led to a rapid review of the way in which the scores had been processed. As a result, it was realised that there had been a data transcribing error onto Oriel during the programming stage several months earlier. Trainees with between 10-30 months experience of T&O were given 6 points rather than 8, and for many this represented an effective score deduction of 2, bringing them from just above to just below the threshold. This was a single, simple, unintended error made by someone who has provided invaluable assistance and advice to the Group over a number of years, but nevertheless one with profound consequences. All applicants were promptly advised of the error and were told that the revised results would be released as soon as possible. Those applicants initially informed that they had been successful were reassured that their score would be unaffected. The revised scores were released the following day, with offers being made to 108 applicants.

We have a great deal of sympathy not only for everyone affected by this year's selection process, but for all trainees who have been affected by Covid-19, the overwhelming majority of whom have provided invaluable clinical support in a novel and challenging environment. We recognise that we have a reciprocal duty to support you all as trainees and would point out that when providing advice to HEE (predominantly through JCST) on a wide range of issues influenced by Covid-19, we have constantly been aware of the need to take into account the unusual and often difficult situations that you currently face. We hope that you understand however, that we have no accountable role in determining definitive policy.

The future

So, where do we go next? We hope that in 2021 we can resume our normal selection process, and if this proves to be the case, you can be assured that the process will have been reviewed in the light of this year's events. The SDG will inevitably suggest changes to ensure that no applicant is penalised as a result of Covid-19 and the areas where issues have emerged with a perception of the process being unfair will be considered in depth. We want an effective and fair selection process as without this we can't expect our specialty to improve, but we also want a selection process with which all stakeholders, particularly you as trainees, have complete confidence. That is our challenge and we will do our utmost to meet it, but if you currently feel frustrated and disappointed please try not to despair, don't give up, keep building your portfolio and please apply again next year.

Rob Gregory
Chair, SAC T&O
Trustee, BOA

Mark Crowther
Chair, SDG T&O