

For decision

UK Medical Education Data Warehouse

Issue

- 1 To work with medical schools and other key interests to pilot development of a database of undergraduate and postgraduate medical education data. This database would facilitate evaluation of our education policymaking, and improve our ability to quality assure the organisations responsible for medical education.

Options

- 2 This paper summarises the following options for consideration:
 - a Option A: Create a database within the GMC by collecting undergraduate data and linking it to postgraduate education, registration and fitness to practise data already held.
 - b Option B: Contribute data to a database held outside of the GMC.

Recommendations

- 3 The Strategy and Policy Board is asked to:
 - a Endorse engagement with key interests on the basis of counsel's advice on the two models above, preferring Option A (Ultimately key interests must agree the terms on which they will support the project. The proposed pilot will serve as a proof of concept that could apply to either model).
 - b Agree that, however the data is housed, the governance arrangements to oversee this database must be collaborative, ensuring key interests have an equal say in developing the database and commissioning research.
 - c Agree that we should set up a cross directorate project group to undertake a pilot in 2014 using a limited set of undergraduate, foundation data and GMC data in order to provide 'proof of concept'.

UK Medical Education Data Warehouse

Issue

- 4** In both the Education Quality Assurance (QA) Review in 2012 and during the data strategy work undertaken by KPMG in 2013 there is a strong call from medical schools, royal colleges, deans, NHS Education Scotland and Health Education England (HEE) for us to maximise the use of our data to make it more accessible to those working to improve medical education. At its meeting on 5 December 2012, Council committed to improving our use of data and to be more transparent about the data we hold. Furthermore, the reports of Sir John Tooke (2008) and Lord Naren Patel (2010) recognised the critical importance of addressing this issue as a benefit of the merger of the Postgraduate Medical Education and Training Board (PMETB) with the GMC. Action is now necessary, indeed arguably it is overdue.
- 5** So far we have largely focused on improving and providing better visibility of postgraduate outcome data. We collect and report on the following information:

 - a** Annual Review of Competence Progression (ARCP) for every doctor in a GMC-approved specialty training programme. We use these data to provide a set of aggregated reports benchmarking the outcomes of local programmes against UK-wide specialty group figures.
 - b** Postgraduate recruitment information for England and Wales, in collaboration with HEE. By mapping recruitment data to National Training Survey census data we can develop a fuller picture of doctors' career pathways. For example the number of doctors who don't apply to further training on completing the Foundation Programme.
- 6** In 2014 we are working on two areas which will greatly increase our understanding of doctors' pathways and the progression of different groups of doctors through training:

 - a** National exam pass/fail data from royal colleges. We will work with colleges to produce benchmarking reports along the lines of those we produce for ARCP data.
 - b** Analysis of postgraduate data by medical school. We are working with medical schools on reports identifying variation in the outcomes for graduates from different medical schools in recruitment, ARCPs and college exams.
- 7** We take advice on each project from our Information Governance Team and from leading counsel to ensure we are fully compliant with data protection legislation.

- 8 The data we hold is all postgraduate, starting with provisional registration. This means reports cannot be adjusted for achievements at medical school or in higher education when comparing outcomes across postgraduate programmes.

The data warehouse

- 9 Last year we received a proposal from the Medical Schools Council (MSC) to create a data warehouse linking undergraduate and postgraduate data. The driving factor in creating this database is that no single organisation can currently link all data on their own. Understanding individuals' performance at different points during their study and medical career is helpful to understand the factors that make doctors more or less likely to progress through the training pathways.
- 10 Currently, undergraduate data is held by the medical schools (e.g. on university exams) and their partner organisations that provide selection tests. 26 medical schools use the UK Clinical Aptitude Test (UKCAT) in their selection processes (albeit in various ways) and data for these schools are held centrally. The MSC is investigating the availability of the data for other remaining seven medical schools that use different selection tests. Linking medical school selection data to data we already hold would allow our routine reports, such as benchmarking of achievement in ARCP, to take into account prior achievement. Linking the data would also allow researchers to measure the predictive validity of selection tools such as the situational judgement test (SJT), which is now used for the Foundation Programme.
- 11 The warehouse would therefore improve access to GMC data and help medical schools to evaluate their assessment and selection methods, for example, by tracking what happens to the top 25% of their graduates versus the bottom 25% after graduation. This process of evaluation is not an end in its own right: it would in turn provide a basis for medical schools to improve their selection and assessment methods, ultimately benefitting patients. It would have key benefits for other key interests: colleges would be able to assess training programmes' performance in their exams adjusting for the prior attainment of candidates, and funders would have better information about the progression of doctors from undergraduate to postgraduate. Further detailed information about the benefits, scope and proposed governance approach is at Annex A, in papers prepared for a meeting chaired by the MSC on 21 October 2013.
- 12 From our perspective there are three important benefits:
 - a Mapping individuals' performance at secondary school and while an undergraduate is also necessary to understand the effect of policy changes for selection tools and assessment tools.

- b Mapping individuals' prior achievements to their future performance will help us understand the 'value-added' by each stage of training (medical school, foundation school and postgraduate training programmes).
 - c Having a full set of data about each individual's attainment at different points in the training pathway will improve our ability to investigate issues of fairness at different stages of training and understand the key factors associated with doctors' progression.
- 13** There are likely to be two key products: a set of benchmarked reports on outcomes, adjusted for prior achievement which update periodically according to the data source, and a programme of research to support evaluation of policy and development of educational tools/methods.
- 14** The proposal for the data warehouse has been shared with the Education and Training Advisory Board which unanimously advised the GMC to make this work a priority. At a recent meeting hosted by the MSC all participants, including the Academy of Medical Royal Colleges, postgraduate deans, NHS Education Scotland, HEE, the Department of Health, the BMA students and Academy trainee doctors supported the proposal.

Options

- 15** Essentially, the proposal is to develop a shared repository of data linked by GMC reference number. Details of the nature of the data and sources are at Annex A. There are two potential approaches, both of which we think are achievable based on advice from counsel (at Annex B).

Option A: Create a database within the GMC

- 16** We would receive identifiable undergraduate data from other organisations and host the database in an integrated unit within the GMC, with appropriate firewalls, for example so identifiable data is not available for fitness to practise.

Option B: Contribute data to a database held outside of the GMC

- 17** We would provide data to a third party, which is held separately from the GMC. Either an independent entity entering into a subcontract with a data management centre; or a body including representation from all participants, contracting out governance and data management to a project board and a data management centre. The data we provide would need to be identifiable for the third party to link it to data from medical schools.

Analysis of options

- 18** In data protection terms, under Option A we would likely be the data controller, although introducing collaborative governance could result in sharing the data controller responsibilities. Under Option B, the data controller would likely be

the newly created independent entity or, if a representative body was created, we with other participating organisations would likely be either joint data controllers or data controllers in common.

- 19** Based on advice from Counsel, we prefer Option A as it would allow us to retain control over the data and could therefore link fitness to practise and National Training Survey results to create a comprehensive de-identified dataset to then be shared with researchers applying for access to the governance group. We have given public assurances about how identifiable survey responses and some identifiable fitness to practise data are held and it is difficult to see how these data could be shared with a third party to create a dataset without breaking these assurances¹.
- 20** A key consequence of Option A is that we would be responsible for the database development. This may impact on our ability to deliver other IS priorities, including those requested by key interests during the Data Strategy work.
- 21** Option B places less responsibility for delivery on the GMC however the trade-off is that the data would be less comprehensive. We are unlikely, for example, to compare SJT scores at entry to medical schools with unpublished FtP data to identify whether there was a correlation between low scores and the number or nature of complaints made against a doctor. Option B is currently preferred by some key interests; feedback obtained through the QA Review and Data Strategy has been that we are not making enough use of the data we already hold. Key interests also see Option B as more independent. It may be that we could retain key interests' confidence in the shared venture through collaborative and independent project governance.

Governance

- 22** Regardless of where the data warehouse is situated, the governance arrangements for the development of the database, the production of reports and the commissioning of research would need to be collaborative, involving representatives from medical schools, the UK Foundation Programme, postgraduate deaneries and royal colleges. It would need to demonstrate to students and doctors in training that the project focuses on improving the quality of medical education in a spirit of genuine inquiry and openness, not reputation management.

¹ Assurances relates to non-disclosure of information linked to identifiable doctors. We have shared aggregated and anonymised sets including these data.

- 23** It is likely that a representative group with a nominee from each of the organisations contributing data, combined with those with expertise in research and analysis and an independent chair would provide the balance of partnership working and independence. It is important to note that this group must have veto over the development of the project and the management of the database to function effectively and maintain key interests' confidence in the GMC.

Costs and piloting

- 24** The Education and Standards Directorate is resourced to manage the collection and reporting on most of the postgraduate data sources envisaged for the database. We have also created datasets linking registration, Professional and Linguistic Assessments Board test and fitness to practise data to postgraduate education data for researchers. However we do not create linked datasets in a systematic way, covering all data.
- 25** UKCAT, which collects and manages data for 26 of the 33 medical schools, has cost in excess of £100,000 and additional project management support to set up, and that was thought to be under-priced. The cost of developing and maintain the database could be significant as would the cost and resource implications for a fully developed database generating routine reports and supporting research work, regardless of the implementation model. For the database to fully recognise the potential benefits it would require a long term commitment from all organisations contributing data.
- 26** We are therefore proposing a pilot for 2014 to link a limited set of undergraduate data to foundation data and to GMC data. The aim would be to answer a research/policy evaluation question, and develop an approach to benchmarking reports that allow us to measure and compare the added value of postgraduate programmes. Piloting would help us better understand the long term costs and resource requirements and to develop the systems required to safely and legally share data systematically. This pilot falls squarely within the aims of the Data Strategy work on doctors' career pathways, and could be run collaboratively across directorates under the wider umbrella of our Data Strategy work programme.
- 27** A pilot linking undergraduate, foundation, ARCP and FtP data to evaluate the performance of the situational judgement selection tool used for foundation recruitment would be a good test as it would require linking across a range of datasets. It would be of value to medical schools, deans, colleges, students and doctors, who all have an interest in the validity of this type of selection tool.
- 28** The pilot would need to be scoped in detail with key interests to confirm the full resource requirement, however as an approximation the group would need support of a 50% FTE education lead, 30% FTE data strategy project manager, and data expertise and business analyst support from our IS team.

Supporting information

How this issue relates to the corporate strategy and business plan

- 29** Strategic aim 3: 'to provide an integrated approach to the regulation of medical education and training through all stages of a doctor's career', and Strategic aim 8: 'to deliver evidence-based policies that demonstrate 'better regulation' principles, and promote and support equality and diversity'.
- 30** In our Business Plan for 2014 we have identified 'developing a stronger data analysis and insight capability to better understand the profession and the effectiveness of our activities' as a key priority.

How the issues support the principles of better regulation

- 31** Linking undergraduate and postgraduate data sets will assist medical schools and others evaluate the impact of educational tools and key processes such as selection and identify good practice. The development of a database that can support medical education research will improve the accessibility of training data and increase the transparency of training pathways.

How the issues differ across the four UK countries

- 32** Most data sets are consistent across the UK, for example, the ARCP and Training Survey data. However not all recruitment processes are UK-wide and further work would need to be done to develop a data set that showed the full picture of postgraduate progression and to ensure all countries could benefit.

What equality and diversity considerations relate to this issue

- 33** Our standards require that training is fair and based on principles of equality. We have begun a programme of work to analyse the outcomes of training for different groups of doctors. We have already reported on differences in ARCP outcomes by UK and non-UK primary medical qualifications which are published here: <http://gmc-onlineeducationreports.org/ReportType.aspx?year>. We have begun to explore whether there are any patterns relating to ethnicity and gender and shared data to support related research. The linking of undergraduate and postgraduate data sets would enable further exploration of the characteristics such as prior achievement in secondary education, and in selection assessments such as the situational judgement tests.

If you have any questions about this paper please contact: Kirsty White, Head of Planning, Research and Development, kwhite@gmc-uk.org, 020 7189 5308.