## NRAC Formula – Potential Improvements to Data Inputs

**Background**

The NRAC Formula uses a range of data sources to provide the input data required to run the Formula. As well as data on likely changes to the shape and size of the Scottish population, these sources include national data on recent patterns of patient activity and costs for a variety of services. These datasets are used by the Formula to predict how the need for services is likely to be distributed geographically in a particular financial year. This referred to in this paper as ‘relative need’.

When the NRAC Formula was developed it was acknowledged that there were either gaps or weaknesses in some of the data inputs available to run the Formula. However the affected elements in the Formula were retained on the basis that new national data would become available in time to either fill a data gap or to strengthen the input data used initially. Indeed the NRAC Review made a number of recommendations where it was felt that efforts to develop national data sources should focus on improving the evidence base used by the Formula. For example it was recommended that there should be development of national data sources relating to services provided in the community.

In the mean time the Formula was run by either using the best available national data or using ‘proxy’ data to fill data gaps i.e. data that was assessed to be broadly representative of the input data required by the Formula but not directly related to the input required. For example at the time the Formula was developed there was no comprehensive national dataset covering patient attendances at A&E services that that could be used to understand how such activity in Scotland is typically distributed with respect to the age and gender of patients - this is required within the age-sex adjustment part of the Formula when considering the relative need for Acute Services. Instead, as a ‘proxy’ solution, it was assumed that the age-gender distribution of A&E was the same as that for patients attending acute outpatient services (see next section for more detail).

Since the Formula was first developed, opportunities to improve particular data inputs have arisen. Each opportunity has been assessed on a case-by-case basis to understand

* whether the newly available data provides a better ‘evidence base’ and so better input data than currently used by the Formula
* the likely impact on future outputs from runs of the Formula if the newly available data is used.

It is then the responsibility of TAGRA to make an informed decision on whether to adopt the use of a new data source in future runs of the model based on this assessment.

This paper discusses two such opportunities that have arisen recently. The opportunities are described in detail and recommendations are made in respect of the next steps for taking these forward.

1. **Use of National Data on Attendances at A&E Services**

As mentioned earlier there has been a long-standing use of proxy data within the age-sex adjustment part of the Formula when considering the relative need for A&E services. This takes place within the part of the formula that predicts how the relative need for Acute services varies according to the age-gender distribution of a particular population (more detail on this follows).

At the time the Formula was developed robust data was not available on how A&E attendees are distributed with respect to age and gender. Instead, as a ‘proxy solution’, it was assumed that this distribution was the same as that for patients attending outpatient clinics. Since then subsequent Formula runs have used national data on outpatient attendances (SMR00) to measure and monitor changes to this distribution.

However, since the Formula was developed there has been significant improvement to national data on patient activity within A&E Services, the main driver being to monitor the 4-hour wait target. Data on patient attendances at A&E services are collected and maintained by ISD in the national data warehouse known as the A&E datamart. Data are collected at source on local clinical systems, then extracted and submitted to the datamart soon after the end of each month. NHS Boards are required to submit data on all attendances for emergency care, whether this is delivered in:

* an A&E Department,
* an Emergency Department,
* a Minor Injuries Unit,
* the trolleyed area of an Assessment Unit.

There are two types of data submitted to the A&E datamart: episodic and aggregate level data. All hospitals with Emergency Departments submit episodic data containing a detailed record for each patient attendance, including the patient’s gender and age Some smaller sites with minor injury units or community hospitals submit only aggregate level data as they do not have the information systems and support to enable collection of detailed patient based information. The aggregate level data consists of the total monthly attendance figures and the number of these seen within 4 hours and does not contain data on the gender and age of patients. However the national coverage of the episodic data is now high with 94% of new and unplanned attendances to sites that now supply this type of data.

As a consequence of this improvement to national data it is possible to examine the age-sex distribution of patients recorded on this source with the ‘proxy’ distribution that has been used by the Formula. Figure 1 shows this for the last published Formula run (calculating shares for 2021/22), which was based on new attendances at acute outpatient clinics in 2017/18. In comparison, also shown is the corresponding distribution for A&E attendances occurring in the same year and for which episodic data was recorded on the A&E datamart. Using an alternative measure of patient activity, an additional data series shows a distribution based on the time spent (minutes) in A&E by these attendees. The advantage of this approach is that time is more likely to be more correlated with complexity and so cost.

**Figure 1 - Age-gender Distribution of A&E Patient Activity in 2017/18**



Figure 1 suggests the age-gender distribution for A&E attendances is very different from the ‘proxy’ distribution used in the last Formula run. In particular younger patients constitute a much higher proportion of patients. However, when considering the time spent by patients in A&E, the distribution is less dissimilar with a higher proportion of activity in the older age groups. This reflects that, although younger patients frequently attend A&E departments, younger patients are more likely to be less complex than older patients, a greater proportion attending for minor injuries and ailments, for example. However the ‘time spent’ measure also suggests the ‘proxy’ data does not fully reflect that patient activity is greater and ‘peaks’ for the very older age groups.

Within the Formula this age-gender activity profile is used to estimate how A&E costs are distributed with respect to age and gender. These costs are then added to other acute service costs to produce a ‘cost curve’ showing rates for the cost per head of population for or acute services for each age-gender group. These rates are then used in the age-sex adjustment part of the Formula to calculate expected costs in a particular geographical area based on the age-gender distribution of the population resident in that area. This is used to measure the distribution of the relative need for acute services in this part of the Formula.

In the last Formula run data on national costs for 2017/18 (sourced from the Cost Book) were used for this calculation. Figure 2 shows how cost of the A&E services in this year, totalling £218M, would be distributed when using a ‘time-spent’ patient activity measure sourced from the national A&E datamart to distribute costs across the gender-age groups.

**Figure 2 - Estimated distribution of A&E service costs 2017/18 using a ‘time-spent’ patient activity by gender and age**



Figure 3 shows the consequential ‘cost curve’ when these costs are added to other acute service costs and comparing to the rates used in the last Formula run. Note that the negligible change to the rates shown. This is because A&E services accounted for only 4.7 per cent acute service costs in 2017/18 and so the changes to how A&E costs are distributed with respect to gender and age have very little impact on the overall rates for acute services. This suggests using the ‘time spent’ A&E activity data would have little impact on Formula outputs.

**Figure 3 – Estimates of Costs per Head of Population in 2017/18 using NRAC Formula**



***Discussion: It would appear appropriate for the NRAC formula to use a measure of patient activity within A&E services that is based on national data now available and relating directly to these services, as opposed to the current ‘proxy’ solution. If TAGRA agrees, then ISD/ASD recommend that the patient activity measure used is that relating to the time spent by attendees at A&E services, given this is more likely to be reflective of the cost of attendance. Note also the minimal impact this was have on the stability of outputs from this part of the Formula.***

1. **Use of National data on District Nursing**

Another part of the Formula, where there have been gaps or weaknesses in respect of the national data available to provide the required input data, is that relating to prediction of the relative need for Community Health Services. At the time the Formula was developed the was very little national data relating to patient activity within these services and the NRAC Review made a number of recommendations to address this. As a consequence a new national data collection system, Community Health Activity Data (CHAD), was initiated to collect episodic data for patients utilising these services. CHAD is still in development with initial work focused on collecting data on District Nursing (DN), Health Visiting and Community Mental Health services. The data being developed for District Nursing appears to be the most advanced and so may provide a more immediate opportunity to improve data used by the Formula ahead of the next run.

In the same way as for A&E services, the Formula uses patient activity data for Community Services to estimate how costs are distributed with respect to gender and age. Since the Formula was developed, data from Practice Team Information (PTI) has been used to do this for DN services. The PTI data collection no longer exists but at the time it collected data from the clinical systems of a sample of approximately 60 participating GP practices which were intended to be broadly representative of the Scottish population in terms of age, gender, deprivation and urban/rural mix. The DN Nursing patient data activity used by the Formula is currently based on DN patient contacts recorded through PTI in 2005-06. Due to the difficulties of recording and collecting nursing activity data on GP systems, as well as the time taken for CHAD to develop, up to now it has not be possible to source more recent data that could utilised by the Formula.

Currently 4 NHS Health Boards, Forth Valley, Greater Glasgow & Clyde, Lanarkshire and Lothian are submitting CHAD DN activity data and work is ongoing to improve the completeness and quality of data submitted from these Boards. In addition efforts are ongoing to bring other NHS Boards on stream. The four submitting Boards submitted episodic data on 1.9M DN patient contacts that took place in 17/18. Figure 4 shows how this activity was distributed with respect to gender and age and in comparison to the PTI data currently being used by the Formula. The age-gender profile of the patient activity recorded on CHAD is very similar to the PTI data, despite the differing sources and the time lag between when the activity occurred.

Figure 5 also shows a comparison of the contact rates per head of population by gender and age for the populations that were covered by either data source. Again the rates of contact are remarkably similar.

**Figure 4 Gender/Age Distribution of DN Patient Contacts**



**Figure 5 DN contacts per Head of Population**



**Discussion: The gender-age profile of the patient activity within the CHAD data submitted by 4 Boards may not necessarily be representative of Scotland as whole. However it shows great similarities to that derived from PTI data and that has been relied upon by Formula for a number of years. Further work is ongoing to understand the quality and completeness of this data and to explore if can be used to extrapolate a profile than can be claimed to representative of DN activity across Scotland. For example, by understanding any variation between urban and rural areas or between areas of high and low deprivation. ISD/ASD intend to report back to TAGRA on these findings. However it is very likely that it will be recommended that CHAD DN data be used in future Formula runs, particularly as data will likely improve in respect of completeness and quality. This ensures the Formula is using more up-to-date evidence for assessing relative need for these services. In addition, given the similarity to the PTI data currently used by the Formula it unlikely this will have major impact on Formula outputs, although this will be tested and reported back on to TAGRA.**