**TAGRA ACUTE MLC SUBGROUP Wednesday 20th January 2016**

**SUPPLY VARIABLES SELECTION**

**Background**

Exploration of the potential supply variables was presented in paper TAMLC36, which was circulated by email in November 2015. A suggestion was made in that paper to retain the existing supply model, consisting of IPACX, OPACX and the health board dummies, on the basis that it did the best job of explaining variation in the cost ratios.

The feedback from the Subgroup by email suggested that there was a need to present a much fuller rationale for the purpose of the supply variables, and for the particular mathematical forms of the variables tested, for the Subgroup to make a more informed decision. It was suggested that IPACX and OPACX be examined separately rather than treated as a pair; since they are highly correlated, it might be more appropriate to retain only one of them. A further suggestion made was to look at the effect of each supply variable option on the coefficient of the needs index.

**1. Summary**

This paper provides a fuller description of the supply variables and their purpose (section 2). The analysis of the supply model options from paper TAMLC36 is updated, based on the suggestions that followed from that paper (and with a correction applied to IPAC and OPAC), in section 3. Section 4 presents the options for the supply model along with an evaluation against the TAGRA core criteria; a new proposal is made to use IPACX plus the health board dummies for the Inpatient diagnostic groups, and OPACX plus health board dummies for Outpatients.

**2. Rationale for the supply variables**

The purpose of the MLC adjustment is to predict the geographic variation in healthcare costs that is due to variation in *need*, by analysing the relationship between cost ratios and the indicators of need. In order to be confident of the validity of the inferred relationship, any other factors influencing the cost ratios should be taken into account where possible. One such factor is the likelihood of higher healthcare activity rates in places where healthcare is more readily available – either because need is not being met adequately in areas with lower supply, or because need is being “over-met” in areas with high supply. In particular, there is a need to control for the effect of previous budget allocations on the pattern of supply of hospital services. For this reason, supply variables always have been included in the NRAC formula’s MLC regression model. The supply variables are control variables: that is, they are included in the regressions to ensure unbiased estimates of the coefficients representing the relationship between cost and the indicators of need, but are not used in the *prediction* of cost. To include them in the prediction of cost would disadvantage any areas in which the level of healthcare provision was lower than it needed to be – or, conversely, reward areas in which higher levels of activity resulted purely from greater supply. Including these variables in the *regression* assures us that the effect of supply is accounted for, and we can be more confident that the resulting model uses the indicators of need to predict only cost variations that are due to differences in need, and not due to variations in supply.

Table 1 shows the supply variables that were previously explored in the 2007 NRAC review and have now been re-examined in the current Acute MLC review. The current supply model, for all care programmes, consists of the health board dummies along with IPACX and OPACX.

*Table 1. Potential supply variables for investigation*

|  |  |
| --- | --- |
| **Variable**  | **Description** |
| GPCount | Number of whole time equivalent (WTE)[[1]](#footnote-1) GPs serving each data zone. This variable is the sum of the number of practices serving each DZ weighted by the whole time equivalent number of GPs in each practice. |
| GPSup, GPSup5, GPSup10 | A function of both the number of whole time equivalent (WTE)1 GPs serving each data zone and the distance to the practices. This can be expressed algebraically as:$$GP supply n km= \sum\_{all practices serving the DZ}^{}\frac{number of WTE GPs per practice}{\left(distance to practice+n\right)^{2}}$$where *n* (the “intrazonal cost”) is 1, 5 or 10 kilometres. The intrazonal cost is an assumption about the part of the travel *within* the small area; the distance to practice represents travel *between* the small area (i.e. its population centroid) and the practice. |
| IA1, OA1 | Size of nearest inpatient / outpatient facility, where size is measured as the number of inpatient or outpatient episodes in the last year. Nearest here means closest to the DZ population centroid. |
| IA2, OA2 | Size of nearest or second-nearest facility, whichever is the larger. |
| IPACX, OPACX | A function of both the size of the inpatient / outpatient facilities serving each data zone and the distance to the facility. This can be expressed algebraically as:$$IPACX/OPACX= \sum\_{all facilities serving the DZ}^{}\frac{number of episodes per facility}{\left(distance to facility+10\right)^{2}}$$ |
| IPAC, OPAC | Similar to IPACX and OPACX, but with an attempt to correct the size of the facility for the size of population that it serves. Expressed algebraically as:$$IPAC/OPAC = \sum\_{h}^{}\frac{{number of episodes per facility}/{\sum\_{d=1}^{N}\left(\frac{DZ population}{distance to facility^{2}}\right)}}{\left(distance to facility+10\right)^{2}}$$where the summation over ‘h’ is a summation over all hospitals serving the given data zone, and the summation over ‘d’ is over all data zones served by the particular hospital. |
| Health Board dummy variables | Represents the effect of the health board. There is one binary variable representing each Health Board; it takes the value ‘1’ for data zones within the Health Board and ‘0’ otherwise. |

The supply variables form three conceptual groups: GP supply, hospital supply, and health board dummy variables. The hospital and health board variables relate clearly to supply of secondary care; however, the role of the GP supply variables is more ambiguous. GP supply could be seen as either a substitute for secondary care or as a complement. GPs are substitutes when they treat patients who would otherwise have gone to A&E or would have needed more serious and higher cost treatment without the early intervention provided by more readily available primary care. In contrast a higher supply of GPs might generate additional episodes of hospital care through a higher level of referrals.

The health board dummies aim to capture broad differences between the 14 health boards. The GP and hospital supply variables are all intended to represent the availability of healthcare at a finer-grained level, in various ways. The simplest of these (GPCount, IA1, OA1, IA2, OA2) look only at the ‘size’ of the facilities serving each data zone – in terms of numbers of GPs, or episode counts for nearby hospitals – with the rationale being that bigger facilities represent a greater availability of healthcare (or, perhaps, a perception thereof on the part of patients). More complex variables (GPSup, GPSup5, GPSup10, IPACX, OPACX) include a term for the distance between the data zone and the facility, since facilities that are nearby are more accessible than facilities some distance away. Finally, the most complex variables, IPAC and OPAC, apply a further correction that accounts for the size of the population served by each facility. This can be thought of as an adjustment for “demand”. A large hospital can treat many patients, but if a very large number of potential patients live close by then this might not amount to a high availability of healthcare (or a perception thereof) for an individual. IPAC and OPAC try to account for this.

The equation for IPAC / OPAC in Table 1 differs slightly from that given in previous papers. This is a correction. Technical Report D of the 2007 NRAC review cites the book ‘Urban Modelling’ by Michael Batty[[2]](#footnote-2) in relation to IPAC and OPAC; the equation in Table 1 above is now consistent with what is given in the book.

**3. Analysis of supply variables**

In this section, the analysis of the different supply variables presented in paper TAMLC36 is updated, in three ways: IPAC and OPAC are corrected as noted in section 2 above; the inpatient and outpatient hospital variables are examined separately; and the effect on the coefficient of the needs index (i.e. the slope of the fitted line) is also examined.

**3.1 Correlations among supply variables, cost ratios and current acute index**

The correlations between the supply variables and the cost ratios for each diagnostic group are shown in Table 2. The correlations with the current Acute needs index are also shown.

IPACX, OPACX and IPAC are correlated most strongly with the cost ratios and with the needs index. OA1, OA2, and the GPSup variables all have some (small) negative correlations with the cost ratios. This is counter-intuitive for a supply variable: activity (and hence cost) would be expected to be higher where availability of healthcare is higher. For this relationship to go the opposite way implies that the supply variables in question are not a good representation of supply, or that GPs are substituting for hospital care

*Table 2. Correlations between supply variables and (i) cost ratios, (ii) needs index.*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Correlations**  | **Whole Acute cost ratios** | **Cancercost ratios** | **Heartcost ratios** | **Digestivecost ratios** | **Injury cost ratios** | **Othercost ratios** | **Respiratorycost ratios** | **Outpatientscost ratios** | **Current Acute needs Index** |
| **IPACX** | 0.360 | 0.187 | 0.140 | 0.292 | 0.129 | 0.269 | 0.299 | 0.353 | 0.290 |
| **OPACX** | 0.322 | 0.167 | 0.118 | 0.221 | 0.121 | 0.212 | 0.277 | 0.401 | 0.247 |
| **IPAC** | 0.281 | 0.136 | 0.127 | 0.237 | 0.094 | 0.254 | 0.206 | 0.207 | 0.245 |
| **OPAC** | 0.170 | 0.062 | 0.069 | 0.085 | 0.047 | 0.131 | 0.131 | 0.247 | 0.159 |
| **IA1** | 0.140 | 0.076 | 0.071 | 0.170 | 0.033 | 0.104 | 0.126 | 0.090 | 0.144 |
| **OA1** | -0.002 | 0.019 | -0.026 | -0.021 | -0.013 | -0.009 | 0.009 | 0.033 | -0.008 |
| **IA2** | 0.148 | 0.087 | 0.075 | 0.161 | 0.046 | 0.088 | 0.137 | 0.134 | 0.130 |
| **OA2** | 0.020 | 0.029 | -0.017 | 0.007 | -0.009 | 0.014 | 0.019 | 0.048 | -0.024 |
| **GPCount** | 0.055 | 0.127 | 0.109 | 0.137 | 0.086 | 0.114 | 0.193 | 0.305 | 0.148 |
| **GPSup** | 0.007 | 0.034 | 0.064 | -0.015 | 0.087 | 0.019 | 0.057 | 0.043 | 0.084 |
| **GPsup5** | -0.021 | 0.015 | 0.043 | -0.041 | 0.056 | -0.030 | 0.030 | 0.011 | 0.029 |
| **GPSup10** | 0.219 | 0.004 | 0.030 | -0.055 | 0.038 | -0.054 | 0.012 | -0.011 | -0.003 |

Tables 3 and 4 show the inter-correlations among the hospital supply variables and the GP supply variables, respectively. As we would expect, there are strong associations among some of these variables since they are conceptually similar and are based on common data sets.

*Table 3. Correlations among hospital supply variables.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Correlations** | **IPACX** | **OPACX** | **IPAC** | **OPAC** | **IA1** | **OA1** | **IA2** | **OA2** |
| **IPACX** | 1.000 | 0.961 | 0.670 | 0.423 | 0.500 | 0.245 | 0.460 | 0.343 |
| **OPACX** | 0.961 | 1.000 | 0.649 | 0.515 | 0.462 | 0.269 | 0.421 | 0.364 |
| **IPAC** | 0.670 | 0.649 | 1.000 | 0.859 | 0.471 | 0.263 | 0.426 | 0.385 |
| **OPAC** | 0.423 | 0.515 | 0.859 | 1.000 | 0.316 | 0.245 | 0.310 | 0.338 |
| **IA1** | 0.500 | 0.462 | 0.471 | 0.316 | 1.000 | 0.342 | 0.552 | 0.406 |
| **OA1** | 0.245 | 0.269 | 0.263 | 0.245 | 0.342 | 1.000 | 0.181 | 0.593 |
| **IA2** | 0.460 | 0.421 | 0.426 | 0.310 | 0.552 | 0.181 | 1.000 | 0.310 |
| **OA2** | 0.343 | 0.364 | 0.385 | 0.338 | 0.406 | 0.593 | 0.310 | 1.000 |

*Table 4. Correlations among General Practice supply variables.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Correlations** | **GPCount** | **GPSup** | **GPSup5** | **GPSup10** |
| **GPCount** | 1.000 | 0.096 | 0.142 | 0.154 |
| **GPSup** | 0.096 | 1.000 | 0.890 | 0.813 |
| **GPSup5** | 0.142 | 0.890 | 1.000 | 0.985 |
| **GPSup10** | 0.154 | 0.813 | 0.985 | 1.000 |

Table 5 shows the correlations between the GP supply variables and the hospital supply variables. GPCount is highly correlated with IPACX and OPACX, and GPSup10 is moderately correlated with IPAC and OPAC. Most other correlations are quite low.

*Table 5. Correlations between GP and hospital supply variables.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Correlations** | **IPACX** | **OPACX** | **IPAC** | **OPAC** | **IA1** | **OA1** | **IA2** | **OA2** |
| **GPCount** | 0.741 | 0.782 | 0.036 | 0.089 | 0.301 | 0.173 | 0.348 | 0.251 |
| **GPSup** | 0.043 | 0.111 | 0.077 | 0.140 | -0.003 | 0.014 | -0.028 | 0.004 |
| **GPSup5** | 0.046 | 0.119 | 0.079 | 0.142 | 0.018 | 0.033 | 0.015 | 0.034 |
| **GPSup10** | 0.032 | 0.104 | 0.585 | 0.488 | 0.020 | 0.035 | 0.025 | 0.041 |

**3.2 Performance of supply variables in regression analysis**

Adjusted R-squared values are used as a goodness of fit measure, indicating the amount of variation in the observations that is explained by a model. The performance of the different supply variables (or pairs of variables), when entered into the regression model along with the current needs index and health board dummies, has been examined by looking at R-squared. Table 6 shows the results, for the best-performing supply variable options.

While GPSup did not perform well, it was the best of the GP supply variables and since it is not highly correlated with the hospital supply variables and is conceptually different, a combined GPSup-IPACX-OPACX option is analysed.

The methodology of the 2007 NRAC review treated the hospital supply variables as pairs, with the inpatient and outpatient components for each variant taken together. However, since most hospitals are both inpatient and outpatient facilities, these variables are highly correlated (see Table 3), and so it may be more reasonable to use only one of the pair. Table 6 therefore includes results for IPACX alone and OPACX alone; for face validity reasons, OPACX is applied to the Outpatients group while IPACX is applied to all inpatient diagnostic groups.

*Table 6: Adjusted R2 values*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Option** | **Whole Acute**  | **Cancer** | **Heart** | **Digestive** | **Injury** | **Other** | **Respiratory** | **Outpatients** |
| IPAC + OPAC | 58.8% | 10.7% | 20.5% | 38.3% | 25.8% | 44.9% | 37.1% | 48.5% |
| IPACX + OPACX | 58.9% | 10.8% | 20.5% | 38.4% | 25.8% | 44.8% | 37.5% | 48.9% |
| IPACX | 58.8% | 10.8% | 20.5% | 38.3% | 25.7% | 44.7% | 37.5% | - |
| OPACX | 58.8% | - | - | - | - | - | - | 48.9% |
| IPACX + OPACX + GPSup | 59.6% | 10.8% | 20.5% | 38.4% | 25.8% | 44.8% | 37.5% | 49.0% |

There is little to choose between any of these options in terms of the R-squared values. The values for IPACX + OPACX are slightly higher than those for IPAC + OPAC. Taking IPACX or OPACX on their own, there is almost no difference in R-squared compared to taking both together. Including GPSup in the supply model does not increase the model’s explanatory power.

**3.3 Effect of supply variables on the Acute needs index coefficient**

Including the supply model in the linear regression has an effect on the coefficient of the Acute needs index, since some of the variation in costs is now explained by the supply variables. Table 7 shows the effect, for each of the considered supply models.

*Table 7: Acute needs index coefficient values obtained from regression using different models*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Whole Acute**  | **Cancer** | **Heart** | **Digestive** | **Injury** | **Other** | **Respiratory** | **Outpatients** |
| needs index only | 0.097 | 0.063 | 0.113 | 0.131 | 0.113 | 0.104 | 0.197 | 0.047 |
| health board dummies + needs index  | 0.091 | 0.055 | 0.117 | 0.117 | 0.121 | 0.098 | 0.187 | 0.037 |
| IPAC + OPAC + health board dummies + needs index | 0.090 | 0.054 | 0.115 | 0.116 | 0.120 | 0.096 | 0.185 | 0.037 |
| IPACX + OPACX + health board dummies + needs index | 0.089 | 0.054 | 0.115 | 0.116 | 0.120 | 0.097 | 0.184 | 0.037 |
| IPACX + health board dummies + needs index | 0.090 | 0.054 | 0.115 | 0.117 | 0.121 | 0.097 | 0.184 | - |
| OPACX + health board dummies + needs index | 0.090 | - | - | - | - | - | - | 0.036 |
| IPACX + OPACX + GPSup + health board dummies + needs index | 0.090 | 0.054 | 0.115 | 0.116 | 0.120 | 0.097 | 0.185 | 0.037 |

Generally the effect of including the supply models is to lower the needs index coefficient, although this is not the case for the diagnostic group Injury. There is almost no difference between the considered supply models in terms of the resulting needs index coefficients.

**4. Selecting the supply model**

Following the above analysis, there are now three decisions involved in selecting the supply model:

* Whether to include GPSup;
* Whether to treat hospital supply variables as a pair or use them singly;
* Whether to use IPACX / OPACX (variables with no population correction) or IPAC / OPAC (variables with population correction).

In Section 3, it was found that GPSup did not add anything to the supply model. Additionally, there is less of a clear relevance to GP supply for modelling Acute hospital activity.

Using only a single hospital supply variable produced results that were just as good as using the pair. The collinearity of each pair is an argument for using just a single variable.

The choice between IPACX / OPACX and IPAC / OPAC is more difficult as these were found to perform equally well in terms of the amount of cost variation that they explained.

Technical Report D expressed some concerns around IPAC and OPAC, noting that some remote data zones in the Island boards had similar values of IPAC / OPAC to some very urban data zones. We find that this difference between IPAC / OPAC and IPACX / OPACX still holds: for example, for data zones in NHS Greater Glasgow & Clyde, OPACX values range from 227 to 2,134, giving no overlap with the range of 4 to 44 for NHS Orkney; OPAC by contrast ranges from 0.11 to 0.54 for NHS Greater Glasgow & Clyde and from 0.02 to 0.47 for NHS Orkney, so there is a significant overlap in values between these two boards.

What this shows is that when using IPAC and OPAC, distant hospitals that serve only very small populations (as in Orkney) have a similar weight in terms of “supply” to hospitals close by that serve much larger populations (as in Glasgow). By contrast, IPACX and OPACX, which do not consider the population size, regard urban areas – where the distances to hospitals are much shorter – as having much greater levels of supply.

IPACX and OPACX have higher correlations with the cost ratios than IPAC and OPAC, and result in slightly higher R-squared in a few diagnostic groups. They also have a simpler, more easily explainable formulation, and are currently in use within the NRAC formula.

However, it makes no practical difference to the result which of these options is used, as seen in Section 3. This perhaps simply reflects that the supply variables (whichever ones are used) only explain a very small proportion of the variation in cost. Indeed, the increase in R-squared obtained when including the supply variables – as compared to a regression on just the health board dummies and needs index – is less than 1 percentage point for all diagnostic groups and typically 0.1—0.2 percentage points[[3]](#footnote-3).

Table 8 evaluates each of the three decisions against the TAGRA core criteria where possible.

**Q1: AST propose to retain the existing supply variables (IPACX or OPACX, plus health board dummies) for the Acute MLC model, with OPACX used for Outpatients and IPACX used for all inpatient diagnostic groups. Is the Subgroup happy to endorse this?**

*Table 8. TAGRA core criteria evaluation for supply variable options.*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Decision:** | **Equity** | **Practicality** | **Transparency** | **Objectivity** | **Avoiding Perverse Incentives** | **Relevance** | **Stability** | **Responsive-ness** | **Face Validity** |
| **Whether to include GPSup** |  | Including GPSup would make Acute supply model more complex |  | GPSup did not increase R2 when added to supply model |  | GP supply not clearly relevant to Acute need |  |  | GP supply not clearly relevant to Acute need |
| **Whether to treat hospital supply variables as a pair or use them singly** |  | Using a single variable simplifies the supply model |  | Using both variables together did not increase R2. The two variables are collinear |  | OPAC / OPACX not clearly relevant to inpatient activity and vice versa |  |  | Favours using only IPAC / IPACX for inpatient groups and OPAC / OPACX for Outpatients |
| **Whether to use IPACX / OPACX or IPAC / OPAC** |  | IPACX / OPACX are easier to calculate and are currently in use within the NRAC formula | IPACX / OPACX have a simpler, more easily explainable formulation | IPACX and OPACX have higher correlations with the cost ratios; slightly higher R-squared in a few diagnostic groups |  |  |  |  | Similar values of supply across diverse urban-rural settings is hard to justify -> favours IPACX / OPACXOn the other hand, correcting for population size (i.e. demand) could be seen as important |

1. After extensive enquiry, it seems that whole time equivalent GP numbers are no longer available by GP practice. We have therefore used GP headcount instead, although this may be an inferior indicator of GP supply. [↑](#footnote-ref-1)
2. Available online at <http://www.casa.ucl.ac.uk/urbanmodelling/UrbanModelling.pdf> - the relevant section is ‘The Allocation of Urban Activities’, p.35 onwards. [↑](#footnote-ref-2)
3. Other possibilities were briefly explored to see if the contribution of the supply variables could be improved. Firstly, the “buffer distance” of 10km for IPAC / OPAC and IPACX / OPACX (see equations in Table 1) was varied, as it had been for the GPSup variables, and an option to make the buffer distance dependent on the geographical area of the data zone was also tested. Secondly, while the inverse *square* of the distance is used in the equations in Table 1, this is somewhat arbitrary and so an inverse cubic function was also tried. None of these variants showed any real difference in the results. [↑](#footnote-ref-3)