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

**REFERENCE MODEL ANALYSIS**

**1. Background and summary**

The MLC element of the formula reflects the effect of the socio-economic circumstances of the local population on their need for health care services, once the effects of age and sex have been addressed. There are two main aspects to the development of an MLC adjustment: first, identifying indicators of need (i.e. factors which are associated with variation in the level of need for services); second, quantifying the size of the relationship between the indicators of need and the need for services (i.e. by how much does the need for services increase for a unit increase in the value of the indicator of need).

The current indicators of need for the Acute MLC adjustment are: all-cause standardised mortality ratio (SMR) in ages 0-74, obtained from death records; and limiting long-term illness (LLTI) ratio (age-sex standardised) from the Census. The size of the MLC adjustment is derived from regression analysis of the relationship between these variables and the costed activity data (the proxy for need) across data zones. The current formula regression model which includes both the indicators of need and the supply variables is referred to as the ‘reference model’. The reference model provides the baseline against which we can assess the results from the new analysis.

This paper reports the updating of the reference model to the specification agreed by the sub-group: firstly, the LLTI variable is now derived from the 2011 census data; and secondly, the data zones have been updated to the version calculated for the 2011 census.

The following section displays the updated data in a set of scatter plots. Section 3 reports the results of the regression runs on both 2001 census data zones (‘DZ2001’) and 2011 census data zones (‘DZ2011’) in terms of model fit (R-squared), the estimated size of the relationship between indicators of need and the need for health care services, and the number and influence of outliers.

In summary the findings are:

* The regression results are broadly similar across both versions of the data zones;
* The MLC adjustment coefficients are slightly higher across all diagnostic groups with DZ2011;
* There are a number of outliers at both geographies, but no influential points were found.

**2. Indicators of need variables and data**

The current Acute MLC indicators of need (ION) are the all-cause Standardised Mortality Ratio (SMR) in ages 0-74, and the Limiting Long-Term Illness (LLTI) ratio. These were included in the NRAC regressions as an index calculated as the sum of the z-scores of these two variables.

The SMR is calculated using death records from five financial years (2009/10 to 2013/14), and the LLTI ratio is calculated using 2011 Census data. For LLTI, the AMLC sub-group previously agreed to use the total number of positive responses to the limiting long-term illness question (respondents stating either “Yes – a little” or “Yes – a lot”) in analysis with the reference model.

The utilisation of healthcare (the proxy for need) is represented by the ratio of the actual costs of healthcare (taking into account activity type and length of stay in that specific neighbourhood) to the expected costs (based on age/sex composition of the neighbourhood’s population and national age/sex average cost per head). The cost ratios are calculated using 2011/12, 2012/13 and 2013/14 cost and activity data, with 2011, 2012 and 2013 population data (a mean of the three mid-year estimates is used).

These cost and ION variables are displayed in the following scatter plots of cost ratios for all diagnostic groups at DZ2011 level. All scatter plots suggest the same overall impression that there is a positive relationship between the explanatory variables and the cost ratios. The scatter plots are shown below (Figure 1).









**

**

*Figure 1. Scatter plots of diagnostic groups’ cost ratios against SMR and against LLTI.*

The cost ratios are also plotted against the two variables combined in the form of the current Acute needs index (Figure 2). As we would expect, all scatter plots suggest the same overall impression that there is a positive relationship between the explanatory variables and the cost ratios.









*Figure 2. Scatter plots of diagnostic groups’ cost ratios against current Acute needs index.*

From both Figure 1 and Figure 2, we notice several outliers for the cost ratios which deviate substantially from the linear trend: these are discussed in section 3.3.

**3. Performance of reference model (adjusted R-squared values)**

The current Acute MLC adjustment is derived from regression of the cost ratios on the needs index. Health board ‘dummy’ variables and supply variables are also included in the regressions, to account for variations in supply influencing the estimate of the MLC adjustment; the supply variables (IPACX & OPACX) are statistical measures representing the size of health care facilities serving the data zone weighted by the (inverse of the) distance between the population grid centroid and the facilities. The health board dummies, supply variables and the needs index are together known as the ‘reference model’.

The following sections consider in turn: model fit; the size of the estimated relationship; and, outliers. In each of these, the reference model results at DZ2011 are compared with those previously obtained at DZ2001.

***3.1 Model Fit:***

Linear models have been fitted to the data, testing SMR and the LLTI ratio – separately, as well as combined in the current needs index (sum of the z-scores of SMR and LLTI) – as explanatory variables for the cost ratios. In addition, a multivariate model including SMR and the LLTI ratio as separate variables is fitted for further comparison. Adjusted R-squared values – the percentage of variance in the cost ratios that is explained by the model – are used as a goodness of fit measure. These are shown in Table 2 below for all diagnostic groups. The values for the supply model + current index (i.e. the reference model) are highlighted in red.

*Table 2. Adjusted R-squared values of models comparing LLTI and SMR performance.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Adjusted** **R-squared** | Predictors | Cancer | Heart | Digestive | Injury | Other | Respiratory | Outpatients |
|  | supply model (IPACX & OPACX) | 6.9% | 3.2% | 18.2% | 4.0% | 15.9% | 10.9% | 40.3% |
| **3 years Data Zones 2011** | supply model (IPACX & OPACX) + SMR | 11.1% | 15.9% | 32.9% | 19.8% | 36.4% | 31.2% | 45.5% |
| supply model (IPACX & OPACX) + LLTI | 9.5% | 20.2% | 38.2% | 25.6% | 45.1% | 36.1% | 50.2% |
| supply model (IPACX & OPACX) + SMR + LLTI | 11.1% | 21.1% | 39.1% | 26.6% | 46.1% | 38.0% | 50.2% |
| supply model (IPACX & OPACX) + current index | 10.8% | 20.5% | 38.4% | 25.8% | 44.8% | 37.5% | 48.9% |
| **3 years Data Zones 2001** | supply model (IPACX & OPACX) | 4.7% | 1.2% | 10.3% | 1.6% | 8.2% | 5.8% | 39.8% |
| supply model (IPACX & OPACX) + SMR  | 12.3% | 13.3% | 31.5% | 16.7% | 32.8% | 28.8% | 53.9% |
| supply model (IPACX & OPACX) + LLTI | 12.3% | 20.0% | 39.3% | 25.4% | 44.4% | 37.7% | 57.6% |
| supply model (IPACX & OPACX) + SMR + LLTI | 12.8% | 20.3% | 39.6% | 25.8% | 44.9% | 38.6% | 57.6% |
| supply model (IPACX & OPACX) + current index | 12.8% | 18.9% | 38.1% | 24.0% | 42.5% | 37.0% | 56.3% |

The regression results suggest that, at both geographies, SMR is mostly a weaker predictor of cost ratios compared to the LLTI ratio. SMR and LLTI are however highly correlated, which explains why the adjusted R-squared values are not substantially higher when combining both variables in an index.

Across all diagnostic groups, broadly similar R-squared values are obtained between the DZ2001 and DZ2011 analysis, except Outpatients, which generally has lower values at DZ2011.

As we would expect, entering the ION variables separately – rather than as a combined z-scores index – increases the R-squared slightly: on average across all diagnostic groups by 0.8 percentage points or 2.5%.

***3.2 Size of the relationship:***

The current Acute needs index coefficients (‘*B*’) obtained from fitting the reference model at DZ2011 are shown in Table 3, along with the coefficients obtained at DZ2001 for comparison. We can interpret a coefficient value *B* as meaning that a unit increase in the needs index will be associated with an increase of *B* units to the cost ratio. Thus it gives an indication of the cost ratios’ sensitivity to the needs index and the strength of the relationship.

All coefficients are significantly different from 0, but are not particularly high. A 95% confidence interval for the coefficients is given; it appears that these intervals are quite narrow. The highest coefficient values are for Respiratory, Injury, Digestive and Heart; the lowest are for Outpatients and Cancer.

The coefficients for the reference model at DZ2011 are higher than at DZ2001 for all diagnostic groups (although this increase is not significant for Digestive and Respiratory). This suggests a more sensitive response: on average across all diagnostic groups the increase in the coefficient is around 8%.

*Table 3. Current Acute needs index coefficients for the reference model at DZ2011 and DZ2001.*

|  |  |  |
| --- | --- | --- |
| **Current Acute Index**  | **3 years Data Zones 2011** | **3 years Data Zones 2001** |
| B | 95.0% Confidence Interval for B | B | 95.0% Confidence Interval for B |
| Lower Bound | Upper Bound | Lower Bound | Upper Bound |
| **Cancer** | 0.054 | 0.048 | 0.060 | 0.047 | 0.041 | 0.053 |
| **Heart** | 0.115 | 0.109 | 0.121 | 0.107 | 0.101 | 0.113 |
| **Digestive** | 0.116 | 0.111 | 0.121 | 0.112 | 0.107 | 0.117 |
| **Injury** | 0.120 | 0.115 | 0.125 | 0.114 | 0.108 | 0.119 |
| **Other** | 0.097 | 0.094 | 0.100 | 0.091 | 0.088 | 0.094 |
| **Respiratory** | 0.184 | 0.177 | 0.190 | 0.180 | 0.173 | 0.187 |
| **Outpatients** | 0.037 | 0.034 | 0.039 | 0.032 | 0.030 | 0.034 |

***3.3 Outliers:***

In Figures 1 and 2, several outliers for the cost ratios were apparent, which deviate substantially from the linear trend. We have investigated the outliers, and have checked whether any outliers will change the slope of the fitted line substantially (influential points), using the same methods employed in paper TAMLC22.

A key finding for the analysis is that there are no such influential points, for all diagnostic groups. In general, there are slightly more cost ratio outliers at DZ2011 than at DZ2001. However, the percentage of *residual* outliers at DZ2011 is the same or slightly *lower* than that at DZ2001, and is much lower than the percentage of cost ratio outliers.

We looked into the two largest cost ratio outliers for Outpatients and discovered that the two data zones – ‘S01010174’ and ‘S01008481’ – both contain prisons (Barlinnie and Saughton, respectively). For both data zones, males between the ages of 20 and 49 accounted for the largest part of expenditure, which suggests the prisoner Outpatient activity is indeed the reason for these extreme values. The 2001 data zones containing these two prisons were also outliers.

AST propose that the few data zones containing prisons should be removed from the regression analysis. Despite the fact that these are not influential points, they represent concentrated populations with healthcare needs that are likely to have a very different relationship with the indicators of need as compared to the general population. Additionally, the funding of prisoner healthcare is not covered by the NRAC formula but handled separately.

**Q1. The Subgroup is asked to decide whether data zones that are dominated by prison populations should be removed from the data sets in future analysis.**