Annex A: Northern Ireland Pharmacy Needs Assessment Additional Needs Modelling Exercise

Contents

Introduction

The additional needs component of the Pharmacy Needs Assessment (PNA) sought to estimate the level of pharmaceutical need across Northern Ireland which was over and above that associated with population, age and gender.

Additional needs explain why populations of similar size and demographic structure can have differing levels of demand for pharmacy services. They encompass factors such as morbidity, deprivation and other socio-economic factors.

They cannot be measured directly but their impact can be estimated using statistical modelling. This modelling involves the analysis of known pharmacy demand against a range of factors which might in combination act as proxies for additional need.

The additional needs modelling work for the PNA was undertaken by independent statisticians seconded from the Northern Ireland Statistics and Research Agency (NISRA) to the Business Services Organisation (BSO).

Overview of modelling

The additional needs models were produced by using multiple linear regression modelling to identify a relationship between a pharmaceutical demand measure or *dependent variable* and a range of indicators across varying domains e.g. morbidity, deprivation, health status etc. (*independent variables*).

The demand measure for the modelling is based upon prescription item dispensing that is constructed to exclude the age and gender demand that will already have been captured in the age gender adjustment.

The additional need modelling was conducted at Super Output Area (SOA) level using weighted least squares regression. This approach was adopted to ensure that undue weight was not given to less populous SOAs[1] within the model because dispensing activity in such SOAs exhibits greater variability. The weighting applied relates to SOA population.

The regression also includes a number of supply variables relating to the availability and accessibility of health services in an area. These can have a distortive impact on dispensing volumes and can mask the actual level of additional need in an area. This affect is negated within the final model by substituting a Northern Ireland average for the individual SOA values.

[1] Individual SOAs contain between 400 and 6000 people

Finally, there are a handful of dummy variables, which relate to intangible characteristics which may influence dispensing activity within an area such as its Local Commissioning Group or geographical classification. In the current modelling exercise, these were negated in the final model through omission as they were considered to be masking real need.

It was important to ensure that the final additional needs model could be easily understood by all stakeholders. Statistical models tend to link the dependent variable with a large number of variables, some of which contribute little to the model and some of which have counterintuitive mathematical relationships with the dependent variable.

This was addressed through the simplification of the original models using an iterative approach, which involved the gradual removal of those factors which had a minimal or counterintuitive impact on the overall model.

Model Creation

Dependent variable

The first stage of the process involved the creation of a dependent variable which could be used as the basis of the subsequent linear regression.

For the purposes of additional needs modelling, this dependent variable had to be a measure of pharmaceutical demand at SOA level. The requirement for an SOA-based measure stemmed from the need to incorporate the final model into an overall composite measure of pharmaceutical need.

The main source of activity within community pharmacy is prescription dispensing and through further consideration and consultation it was decided that dispensing volumes would be a good proxy indicator of overall pharmacy service demand.

A database of dispensing activity for 2019/20 was produced using data from FPS Pharmaceutical Payments system matched to patient attributes (such as age, sex and SOA) using Health and Social Care number data.

It was not possible to match all prescriptions to patient attributes due to limitations of the automated prescription scanning processed by FPS Pharmaceutical Services, which means that HCN data is not always captured. It was possible though to achieve a match for around 90% of all items dispensed. On reviewing these rates and with this high match percentage, this was confirmed to be a representative sample of prescriptions dispensed in 2019/20.

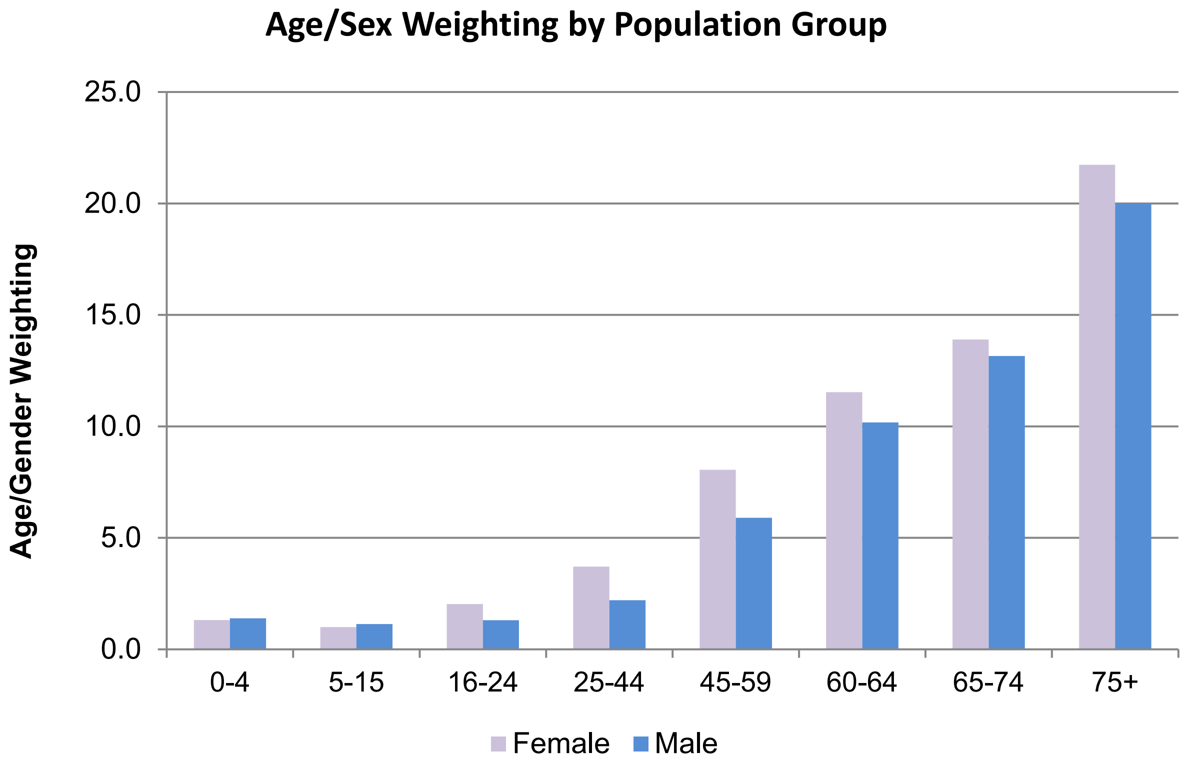
A demographic profile of each SOA was then built up using the latest available mid-year population estimate (June 2019) and GP registration information from the National Health Application and Infrastructure Services system.

Age/Sex adjustment

This profile formed the basis of the age/sex adjustment. The process sought to level the playing field for all SOAs by creating a standardised population which reflected differences in dispensing relative to age and sex.

The process began with the calculation of item weights for 16 age and sex groupings covering the entire population of Northern Ireland. These weights were calculated with reference to a benchmark comprising the age/sex group with the fewest items dispensed per head of population.

In this case, the benchmark group was females aged between five and 15. Their level of dispensing was assigned a weighting of 1.0. All other age/gender groupings were then expressed in relation to this benchmark e.g. a group whose dispensing per head was twice that of the benchmark would have a weight of 2.0.



	0-4	5-15	16-24	25-44	45-59	60-64	65-74	75+
Female	1.3	1.0	2.0	3.7	8.1	11.5	13.9	21.7
Male	1.4	1.1	1.3	2.2	5.9	10.2	13.2	20.0

Figure 1: Age/Sex weighting used in creation of dependent variable

As can be seen from the information above, the weighting for both sexes tended to increase with age. The highest weight (21.7) related to females aged 75 and above.

The weighting was then applied to SOA population data so that a female in the 5-15 group counted as a single person while, for example, a female aged 75 and above would be counted as 21.7 people. This was done for all age/sex groups within an SOA and summed. This allowed SOA dispensing rates per person to be calculated (by dividing the age/sex weighted population by the crude population) which could be expressed on a basis which was independent of age and sex.

The final stage of the process involved indexing the individual SOA dispensing rates per person in relation to the Northern Ireland average.

Additional Need Indicators

The modelling process also makes use of a range of independent, supply and dummy variables. There are many potential variables and the field was initially narrowed using a number of criteria:

1. *The measures had to be available at SOA level*
   This was a pre-requisite so the final measure could align with the remainder of the PNA.

2. *The measures had to be intuitive*
   This meant that there had to be a clear logical basis for their inclusion.

3. *The measures had to be as current as possible*
   This was necessary so that the model could provide the most accurate picture available of additional needs across Northern Ireland. In practice, measures which were last updated prior to 2017 were generally avoided. There were, however, some statistics which are only measured as part of the Census and so it was necessary to utilise data from 2011.

4. *The measures should be based on publicly available data*
   Given the importance of pharmaceutical services, it was felt that it was important to provide maximum transparency concerning the additional needs model and its construction.

5. *The measures should be updatable*
   This criterion was intended to ensure that the model could be refreshed as required.

During the selection process, 58 potential variables which could act as proxies for additional need were identified. A summary of these variables by date and type is provided below.

Variable Category	Data Year				Total
	2011	2017	2018	2019	
Demographics	11	-	-	1	12
Dummy Variables	-	-	-	-	6
Health Variables	5	-	3	9	17
Other	-	-	-	2	3
Social Security	-	-	8	-	8
Supply Variables	-	-	-	5	5
Vital Statistics	-	7	-	-	7
Total	**16**	**7**	**11**	**17**	**58***

* Year total does not add up to 58 because dummy variables and an 'Other' variable are not all date-related

In many cases, the raw data had to be standardised into rates. This is necessary to permit comparisons between SOAs with different population sizes.

For example, more populous areas often have, for example, more deaths because they contain more people. But the number of deaths as a proportion of their population may be lower than those of less populous SOAs which have fewer deaths in absolute terms.

Full details of the variables and the denominators used to standardise them are provided in Appendix 6.2 of the main report.

Supply Variables
Some of the indicators identified were supply variables i.e. factors which can explain demand variations among different areas but don't reflect an actual population need.

For example, an individual's proximity to their GP practice might encourage them to visit their GP more frequently, possibly resulting in them receiving more prescriptions for a given level of need than patients who live further away from their GP and may be more hesitant to seek assistance for minor ailments given the greater distance involved.

Such factors can artificially increase population need in certain areas if they are not controlled for during the modelling process. This is achieved by neutralising their impact in the final models by substituting the Northern Ireland average value for those variables for the individual SOA figures.

Five key supply factors were identified following stakeholder consultation and consideration by the project team. These were:

- Average distance to nearest GP practice
- Average Distance to nearest pharmacy;
- Prescription Scan Rate;
- Prescribing Rate; and
- List Discrepancy.

Average distance to nearest GP practice / pharmacy
The potential impact of proximity to a community health service is discussed above.

Prescription Scan Rate
Patient information for prescriptions is currently obtained through computer scanning of the barcodes on the original prescriptions. Due to the limitations of the scanning process, patient

information is not captured in all cases. In 2019/20, patient information could not be obtained from around 10% of all prescriptions dispensed.

The scan rate is variable across Northern Ireland and, if not corrected for, could artificially reduce the number of prescription items attributed to patients in the affected areas. The prescription scan rate variable prevents this situation arising.

Prescribing Rate

There are also variations in prescribing behaviour across GP practices. This may not necessarily represent differing levels of need but could reflect differences in treatment and therapy preferences between practices and GPs.

Given that the dependent variable is based upon prescription volumes, there is a possibility that observed demand in some areas might be influenced by these variations in medical practice rather than population need. The prescribing rate variable provides a means of removing this potential influence.

List discrepancy

List discrepancy is the difference between the number of patients registered with GPs in an SOA and its actual resident population.

This discrepancy can arise due to factors such as:

- Delays in removing departing patients from practice lists;
- Delays in registering new births with practices;
- Delays in removing deceased patients from practice lists;
- Use of addresses of convenience by users from Ireland to access free GP services

High levels of list discrepancy can lead to situations where areas are unduly advantaged within needs indices. It is therefore necessary to calculate list discrepancy so that its impact can be managed within the modelling process. The calculation used was based on the proportional difference between the latest MYE and similar period NHAIS GP registration data.

The GP Prescribing Formula Review found that raw list discrepancy values could incorporate some legitimate needs. As a result, it used an adjusted list discrepancy variable, which estimated the real level of list discrepancy.

This adjustment was made at the end of the modelling process using linear regression with the raw list discrepancy variable as the dependent variable. The independent variables were the other components of each candidate model along with five factors deemed to be the drivers of 'actual' list discrepancy.

These factors were: number of births, number of deaths, number of students, proximity to the Irish border and migration levels. The estimate list discrepancy was then used in the final calculations in place of raw list discrepancy.

It was impossible to replicate this adjustment in this modelling exercise for two reasons. The first was that migration statistics are unavailable at SOA level. Secondly, cross-border users are a legitimate source of need for the purpose of this exercise because of the demand they place on pharmaceutical services in border areas.

After consultation with the peer reviewer, it was agreed to use raw list discrepancy values throughout our modelling. These values would then be controlled out of the model to remove the distortive effect associated with them. Meanwhile, distance to the border would be input into the model as an independent variable due to the desirability of quantifying its impact on pharmaceutical need.

Testing for collinearity

The variables were then tested for collinearity. Collinear variables are so closely correlated with each other that they are effectively measuring the same thing. The inclusion of two or more collinear variables into a model would distort the final results because such variables might act in tandem with each other, producing misleading indications of significance.

A series of bivariate correlations involved all elements was performed to identify groups of variables which had a high correlation with each other. There were six such groups identified. These were:

Group 1	Carers Allowance Rate; Unpaid Carers Rate; ESAC Rate; Housing Benefit Rate; Income Support Rate; Jobseekers Allowance Rate; LLTI Lot Rate; LLTI Not Rate; Lone Parent Dependent Children Household Rate; No Qualification Rate; Pension Credits Rate; Poor Health Rate.
Group 2	Male Death Rate; Female Death Rate
Group 3	Average Distance to Pharmacy; Average Distance to GP
Group 4	No Qualification Rate; Level 4 Qualification Rate
Group 5	Catholic Rate; Protestant Rate
Group 6	White Ethnicity; Other Ethnicity

Collinearity was managed by including only one variable from each group in the modelling process. These variables were selected partly on their Pearson correlation values and their interaction with other variables within a series of test models.

Modelling Strategy

The modelling strategy used sought to identify candidate models which were parsimonious but had the greatest possible predictive value using a combination of independent variables, supply variables and dummy variables.

Each model was produced through an iterative process. An initial version would be produced for each model using the full set of the supply, needs and dummy variables needed to test a particular scenario. This version would typically contain a large number of variables, some of which contributed little to the model and some had an apparently counterintuitive input into it.

The phenomenon of counterintuitive inputs arises because socio-economic indicators tend to be somewhat correlated with each other even if they are not fully collinear.

This means that there are certain combinations of variables which can have unforeseen interactions with each other during the construction of a model, leading to situations where, for example, an increased number of pension recipients is associated with a decreased level of pharmaceutical demand.

To produce a more compact and easily comprehensible model, this initial version was simplified over a series of iterations. Each iteration saw the removal of a single variable from the model before the regression was rerun. This gave all remaining variables an equal opportunity to influence the next version of the model.

The removal process was conducted using the following criteria in order of sequence:

1. Remove variable if counterintuitive and coefficient is significant;
2. Remove variable if counterintuitive and coefficient is insignificant;
3. Remove least significant variable from model

The removal of counterintuitive variables was prioritised because they were most likely the product of unforeseen interactions with other variables in the model.

During initial test regressions, it was noted that the dependent variable and some independent variables did not exhibit a fully normal distribution. This led to the creation of statistically misspecified models. This issue was addressed by log transforming the dependent variable and all other variables (apart from dummies) to ensure they were normally distributed and would produce correctly specified models.

Outlier handling

Outliers are data points which have numeric values which differ significantly from the rest of the points concerning a particular variable. They generally arise due to extreme variability within the measurement. There were a small number of outliers within the dependent variable, relating to areas where there were very few items dispensed per head or population.

The inclusion of outliers within a dependent variable can affect the accuracy of the resultant model. In the worst case, variables which might not otherwise be significant might be included in the model.

To eliminate this risk, it was decided to exclude all SOAs with extreme dependent variable values (±3 standard deviations from the mean) from the modelling process. This figure was selected because it would encompass approximately 99.7% of all potential values where the dependent variable is normally distributed. This ensured that the SOAs selected for exclusion were genuine outliers.

SOA Code	LGD	Name	Value
95AA01S1	Antrim & Newtownabbey	Aldergrove 1	-2.32
95GG12S3	Belfast	Botanic 3	-1.40
95GG12S2	Belfast	Botanic 2	-1.35
95GG42S3	Belfast	Stranmillis 3	-1.14
95GG42S2	Belfast	Stranmillis 2	-0.94

Mean: -0.02

Std. Deviation: 0.294

Mean ± 3 SD values: -0.90; +0.86

Figure 2: SOAs excluded from modelling exercise due to outlier status

It should be emphasised that these SOAs were excluded only during the model creation process. An additional needs index value was generated for each using the model produced using the 885 SOAs which did not have outlying values.

LCG Dummy Variables
The modelling exercise incorporated dummy variables for Local Commissioning Groups (LCGs) due to a perception that LCG might, through local policies, influence service provision and need in their particular area.

This assumption has been questioned in recent years. Firstly, Northern Ireland now has a single Health and Social Care Board, which significantly reduced the scope for local policy variation.

Secondly, the review of the Northern Ireland Weighted Capitation Formula revealed an inexplicable juxtaposition between one of the dummies, the WHSCT dummy, and admission rates during its additional needs modelling.

In that instance, the peer reviewer suggested the exclusion of the LCG dummies to eliminate this effect. The Department of Health (DoH) decided to retain them on the basis that "it is known that on the ground there remain differences in policy and practice across the LCGs". The DoH then forced[2] the LCG dummies into its models to account for these perceived policy differences.

This modelling exercise initially followed the DoH's approach but it soon emerged that two LCG dummies (SHSCT and WHSCT) were contributing around 10% of any model's explanatory power. This would suggest an extremely powerful local policy effect which had never been seen in previous modelling exercises.

After discussion with the peer reviewer, it was concluded that this situation was improbable and that the dummies were probably masking other less geographically defined sources of need. A regression of the dummies against the need proxies identified a definite relationship between dummies and needs, which suggested that this was a distinct possibility.

It was agreed with the peer reviewer to cease forcing the LCG dummies into the model but to input them as though they were normal variables relating to need proxies. This led to the removal of all LCG dummies from the final refined model.

It was then decided to investigate the hypothesis that the presence of LCG dummies was masking additional need by offsetting the effects of some needs proxies. This was achieved by constructing a second model which was identical to the first but did not incorporate LCG dummies.

[2] This involves instructing the modeller to include these variables in the final model regardless of their statistical significance.

While early test models indicated that this might be the case[3], the exclusion of LCG dummies did not result in any change in terms of the final model.

Model testing

Each of the final models produced were tested to ensure that they were within specification and met the basic assumptions of multiple regression.

[3] These test models indicated that the exclusion of LCG dummies produced models which had a higher predictive value than the first and led to the inclusion of an extra further additional needs variable.

Modelling Outputs

The methodology outlined in the previous section was used to generate two models. Based on initial test modelling, it was decided that the following variables would be selected from each of the collinear groupings:

Group 1	ESAC Rate
Group 2	Female Death Rate
Group 3	Average Distance to nearest GP Practice
Group 4	Level 4 Qualification Rate
Group 5	Catholic Rate
Group 6	Other Ethnicity

For both models, the significant supply variables were Average Distance to Pharmacy, Prescriber Rate Index and Scan Rate.

Model A

The first model was a baseline model produced with LCG dummies included among the independent variables. It produced an adjusted-R^2 value (predictive value) of 0.887. This indicates a very strong correlation between the model and the dependent variable.

The initial version of the model utilised 19 variables, including one LCG dummy – SHSCT. Twenty-one iterations were required to produce the final refined model.

All counterintuitive contributions were removed at the sixth iteration. At that point, the model contained 19 variables. It did not include any LCG dummies.

The final iteration contained 11 variables, including the three significant supply variables and the list discrepancy variable. It did not include any LCG dummies.

The basic composition of the final model is outlined below.

	Variable	Adj-R^2	Contribution[1]
1	Employment Support Allowance Claimant Rate	0.632	71.2%
2	Supply Variables *(a) Average GP Distance* *(b) Prescriber Rate Index* *(c) Scan Rate*	0.159	17.9%
3	List Discrepancy	0.035	3.9%
4	Stroke / Transient Ischaemic Attack Prevalence	0.018	2.0%
5	Nursing Home Residency	0.013	1.5%
6	Proximity to Border	0.013	1.5%
7	Level 4 Qualification Rate	0.012	1.4%
8	GP Registrations – New from outside UK/ROI	0.003	0.3%
9	Mental Health	0.003	0.3%
Total		**0.887**	**100%**

[1] Contribution to model. Co-efficients are provided in detailed output.

Model B

The second model was identical to the first apart from the exclusion of LCG dummies. It produced an adjusted-R^2 value of 0.887. This indicates a very strong correlation between the model and the dependent variable.

The initial version of the model utilised 20 variables. Eighteen iterations were required to produce the final refined model. All counterintuitive contributions were removed at the sixth iteration. At that point, the model contained 19 variables.

The final iteration contained 11 variables, including the three significant supply variables and the list discrepancy variable.

The exclusion of LCG dummies simplified the iterative process by eliminating a major source of interactions with other variables. The final model, however, was identical to Model A.

The basic composition of the final model is outlined below.

	Variable	Adj-R^2	Contribution[1]
1	Employment Support Allowance Claimant Rate	0.632	71.2%
2	Supply Variables *(d) Average GP Distance* *(e) Prescriber Rate Index* *(f) Scan Rate*	0.159	17.9%
3	List Discrepancy	0.035	3.9%
4	Stroke / Transient Ischaemic Attack Prevalence	0.018	2.0%
5	Nursing Home Residency	0.013	1.5%
6	Proximity to Border	0.013	1.5%
7	Level 4 Qualification Rate	0.012	1.4%
8	GP Registrations – New from outside UK/ROI	0.003	0.3%
9	Mental Health	0.003	0.3%
Total		**0.887**	**100%**

[1] Contribution to model. Co-efficients are provided in detailed output.

Model C

An attempt was made to produce a third model which excluded all variables from the 2011 Census. The advantage of this model would have been that it would have been more readily updatable. But it failed post-production testing.

Commentary

The two models were produced using the same dependent variable and the same input variables with the exception of the four LCG dummies removed for Model B.

In terms of final composition, the models are identical. It was felt, however, that the construction of Model B was superior to that of Model A. This was primarily due to concerns about the interaction between the LCG dummies and need variations, which had led to the apparent masking of demand when they were forced into the modelling process. After consultation with the peer reviewer, it was agreed that Model B would be the preferred choice.

Appendix 1: Model Details

Model A

Model using ESAC Rate with LCG dummies

Dependent Variable: PU_Items_NM_LN
Weighting Variable: MYE_Index

Supply Variables: avg_GP_distance; Prescrib_Rate_Index; Scan_Rate

Independent Variables: BHSCT; SEHSCT; SHSCT; WHSCT; List_Discrep; AAR_Rate; ESAC_Rate; Pension_Rate; Level4_Qual_Rate; Only_Person_aged_65years_plus_HH_Rate; Catholic_Rate; Other_Religion_Rate; Outside_UK_ROI_Rate; LLTI_Little_Rate; Unpaid_Care_Rate; Pop_Density; Asthma; Cancer; CHD; COPD; DM; HYP; MH; STIA; Birth_Rate; UM_Birth_Rate; Death_Rate; Female_Death_Rate; U75_Death_Rate; Med_Age_At_Death; Malignant_Neoplasm_Death_Rate; Circulatory_Disease_Death_Rate; Respiratory_Disease_Death_Rate; Non UK/ROI GP Registrations Rate; MYE_1yr_growth; MYE_3yr_growth; nursing_home; Rural; Mixed; Dist_to_border; Other_Ethnicity_Rate.

Initial Run: $R^2 = 0.901$ (19 variables)

Final result: $R^2 = 0.887$ (11 variables [3 supply])

Model composition

	R^2	*Co-efficient*
1. ESAC Rate	0.632	0.290
2. Supply Variables	0.159	
Average GP Distance		*-0.041*
Prescriber Rate Index		*0.140*
Scan Rate		*2.478*
3. List Discrepancy	0.035	6.712
4. STIA (Stroke)	0.018	0.329
5. Nursing Home	0.013	3.499
6. Distance to Border	0.013	-0.031
7. Level 4 Qualification Rate	0.012	-0.160
8. Overseas GP Reg Rate	0.003	-0.017
9. Mental Health	0.003	0.097
Constant		*-18.394*

Exclusions

The process followed when reducing the original model to eight independent variables was as follows:

Run	Exclusion Rule	Description	R^2 diff
2	Counterintuitive input	Only_Person_aged_65years_plus_HH_Rate	**-0.002**
3	Counterintuitive input	LLTI_Little_Rate	**-0.004**
4	Counterintuitive input	Death Rate	**+0.001**
5	Counterintuitive input	Circulatory Disease Death Rate	**-0.001**
6	Counterintuitive input	Pension Rate	**-0.001**
7	Least significant input	Unmarried Mothers' Birth Rate	**0.000**
8	Least significant input	Female Death Rate	**-0.001**
9	Least significant input	Catholic Rate	**+0.001**

10	Least significant input	SEHSCT	-0.001
11	Least significant input	CHD	0.000
12	Least significant input	Outside_UK_ROI_Rate	-0.001
13	Least significant input	Other Ethnicity Rate	-0.001
14	Least significant input	U75 Death Rate	-0.001
15	Least significant input	Asthma	0.000
16	Least significant input	Median Age At Death	+0.001
17	Least significant input	Hypertension	-0.001
18	Least significant input	SHSCT	0.000
19	Least significant input	MYE_1yr_growth	0.000
20	Least significant input	Population Density	-0.001
21	Least significant input	Birth Rate	-0.002

The following variables were never included in the model during any of the runs:

BHSCT; WHSCT; Other_Religion_Rate; Cancer; COPD; DM; Malignant_Neoplasm_Death_Rate; Respiratory_Disease_Death_RateMYE_3yr_growth; Rural; Mixed

Model B

Model using ESAC Rate with no LCG dummies

Dependent Variable:	PU_Items_NM_LN
Weighting Variable:	MYE_Index
Supply Variables:	avg_GP_distance; Prescrib_Rate_Index; Scan_Rate
Independent Variables:	List_Discrep; AAR_Rate; ESAC_Rate; Pension_Rate; Level4_Qual_Rate; Only_Person_aged_65years_plus_HH_Rate; Catholic_Rate; Other_Religion_Rate; Outside_UK_ROI_Rate; LLTI_Little_Rate; Unpaid_Care_Rate; Pop_Density; Asthma; Cancer; CHD; COPD; DM; HYP; MH; STIA; Birth_Rate; UM_Birth_Rate; Death_Rate; Female_Death_Rate; U75_Death_Rate; Med_Age_At_Death; Malignant_Neoplasm_Death_Rate; Circulatory_Disease_Death_Rate; Respiratory_Disease_Death_Rate; Non UK/ROI GP Registrations Rate; MYE_1yr_growth; MYE_3yr_growth; nursing_home; Rural; Mixed; Dist_to_border; Other_Ethnicity_Rate.
Initial Run:	$R^2 = 0.901$ (18 variables)
Final result:	$R^2 = 0.887$ (11 variables [3 supply])
ZPRED² significance:	0.828

Model composition

	R^2	Co-efficient
1. ESAC Rate	0.632	0.290
2. Supply Variables	0.159	
Average GP Distance		*-0.041*
Prescriber Rate Index		*0.140*
Scan Rate		*2.478*
3. List Discrepancy	0.035	6.712
4. STIA (Stroke)	0.018	0.329
5. Nursing Home	0.013	3.499
6. Distance to Border	0.013	-0.031
7. Level 4 Qualification Rate	0.012	-0.160
8. Overseas GP Reg Rate	0.003	-0.017
9. Mental Health	0.003	0.097
Constant		*-18.394*

Exclusions

The process followed when reducing the original model to eight independent variables was as follows:

Run	Exclusion Rule	Description	R^2 diff
2	Counterintuitive input	Only_Person_aged_65years_plus_HH_Rate	**-0.001**
3	Counterintuitive input	LLTI_Little_Rate	**-0.005**
4	Counterintuitive input	Death Rate	**+0.001**
5	Counterintuitive input	Circulatory Disease Death Rate	**-0.001**
6	Counterintuitive input	Pension Rate	**-0.001**
7	Least significant input	Unmarried Mothers' Birth Rate	**0.000**
8	Least significant input	Female Death Rate	**-0.001**
9	Least significant input	Catholic Rate	**0.000**
10	Least significant input	CHD	**0.000**

Run	Exclusion Rule	Description	R^2 diff
11	Least significant input	Outside_UK_ROI_Rate	-0.001
12	Least significant input	Other Ethnicity Rate	-0.001
13	Least significant input	U75 Death Rate	-0.001
14	Least significant input	Asthma	0.000
15	Least significant input	Median Age At Death	0.000
16	Least significant input	MYE_1yr_growth	0.000
17	Least significant input	Population Density	-0.001
18	Least significant input	Birth Rate	-0.002

The following variables were never included in the model during any of the runs:

Other_Religion_Rate; Outside_UK_ROI_Rate; Cancer; COPD; DM; HYP; Malignant_Neoplasm_Death_Rate; Respiratory_Disease_Death_Rate; MYE_3yr_growth; Rural; Mixed

Appendix 2: Sample SOA level scores

1. SOAs with highest additional needs scores – Models A and B

SOA	SOA Name	LGD	Rank	Score
95GG19S1	Crumlin 1 (Belfast)	Belfast	1	1.88
95GG51S3	Woodvale 3	Belfast	2	1.76
95MM10S1	Creggan Central 1	Derry City and Strabane	3	1.72
95MM13S2	Culmore 2	Derry City and Strabane	4	1.71
95GG48S2	Whiterock 2	Belfast	5	1.71
95GG40S2	Shankill 2	Belfast	6	1.70
95SS06S2	Collin Glen 2	Belfast	7	1.69
95GG48S3	Whiterock 3	Belfast	8	1.69
95MM12S2	Crevagh 2	Derry City and Strabane	9	1.68
95MM06S2	Carn Hill 2	Derry City and Strabane	10	1.67

Scores are relative to the NI average (NI = 1.00)

2. SOAs with lowest additional needs scores – Models A and B

SOA	SOA Name	LGD	Rank	Score
95AA01S1	Aldergrove 1	Antrim and Newtownabbey	890	0.45
95GG42S3	Stranmillis 3	Belfast	889	0.51
95GG42S2	Stranmillis 2	Belfast	888	0.54
95GG42S4	Stranmillis 4	Belfast	887	0.55
95GG42S1	Stranmillis 1	Belfast	886	0.55
95II15S1	Knockbracken 1	Lisburn and Castlereagh	885	0.58
95GG41S2	Stormont 2	Belfast	884	0.59
95WW19S3	Jordanstown 3	Antrim and Newtownabbey	883	0.60
95WW23S2	Rostulla 2	Antrim and Newtownabbey	882	0.61
95SS09S1	Drumbo 1	Belfast; Lisburn and Castlereagh	881	0.62

Scores are relative to the NI average (NI = 1.00)

Appendix 3: Correlations with Multiple Deprivation Measure

Model	MDM Domains							Overall MDM
	Income	Employment	Health	Education	Services	Living Environment	Crime	
A/B	0.601**	0.910**	0.899**	0.791**	-0.346**	0.244**	0.642**	0.840**

** Correlation is significant at the 0.01 level (2-tailed).