

**Modelling the Distribution of Services for People with
Physical and Sensory Disabilities in Northern Ireland
(The PoC7 Programme of Care)**

Paul Dixon

(MSA/Ferndale – Secta, University of York)

Martin Spollen

(MSA/Ferndale – Secta)

Roy McConkey

(University of Ulster)

Jim Jamison

(Independent consultant)

Karen Kenny

(Secta)

September 2004

Acknowledgements

We are most grateful to all those throughout the Health and Social Services in Northern Ireland who went to such great lengths to locate and supply data for use in this study. Particular thanks are due to members of the CFRG who were very supportive of all phases of the project.

Finally, we must acknowledge the extensive help and support we received from members of the DHSSPS, especially Martin Mayock, Philip Spotswood and Donna Ruddy.

List of contents

EXECUTIVE SUMMARY **v**

MAIN REPORT

1. INTRODUCTION AND CONTEXT	1
1.1 Introduction	1
1.2 Context	1
1.3 Definition of Programme of Care	3
2. ACTIVITY DATA FOR THE MODELLING	4
2.1 The area base for the modelling	4
2.2 SOSKARE (PSAB) extracts for PoC7	6
2.3 PoC7 data from activity systems such as LCID and CLAN	12
2.4 TROJAN data for PoC7	17
2.5 Data on deaths due to the troubles (for PoC7)	18
2.6 Hospital Incidence Data for PoC7	21
3. NEEDS DRIVERS AND SUPPLY VARIABLES	26
3.1 Need drivers for services for people with physical disability and sensory impairment	26
3.2 Effects of Population density	26
3.3 Supply side factors	27
4. THE MODELLING	29
4.1 Modelling client numbers from the SOSKARE data set	30
4.2 Modelling the data from LCID and CLAN	39
5. TESTING THE MODELS	41
5.1 Sensitivity tests	41
5.2 Unmet need tests and adjustments	44

6. CONCLUSIONS AND RECOMMENDATIONS	49
GLOSSARY	52
ANNEX 1 Needs Drivers in Combined Set	59
ANNEX 2 Trust Data Request Letter and Specification	64

EXECUTIVE SUMMARY

The funding for many Programmes of Care in Northern Ireland is determined by age-sex weighted capitation methods supplemented by formulae for predicting local levels of additional need. Such formulae have been developed from empirical research in Northern Ireland, but prior to the present project no such research has been undertaken for the Physical & Sensory Disability Programme (PSD). In the absence of an empirically based formula, the Standardised Mortality Ratio (SMR) for deaths of people aged under 75 is used as the additional needs index for this Programme.

Research to develop additional needs formulae usually requires extensive data on the numbers of people receiving relevant services and the volume of services delivered. Such data are typically obtained from the client record systems of health and social services, but experience from other programmes has shown that data can be incomplete and difficult to extract.

Given these potential difficulties, a two-phase approach was proposed for the current exercise.

Phase I: a short scoping study to determine whether available information on the population of physically and sensorially disabled persons in Northern Ireland is of sufficient quality and coverage to support development of an index of additional need (i.e. over and above known age and gender effects); and

Phase 2: pending acceptance by the Capitation Formula Review Group of the findings and recommendations of Phase I, to collate the required information and undertake appropriate statistical analysis to derive an index suitable for implementation.

Phase I of the project identified the following five data sources with the potential to provide costed information on the services delivered to PoC7 clients.

- SOS CARE extracts for PoC7
- TROJAN DATA for PoC7

- community health and domiciliary care activity and finance data from LCID and CLAN
- data on deaths due to the Troubles (for PoC7).
- Hospital Incident Data for PoC 7

For the purposes of small area modelling, each data set must include details of the services supplied to individual clients and the post- or ward code of origin of each client. They should also contain a unique client identifier that can be used to guard against double counting. Many of these details were missing from these data and at one stage there were doubts on the feasibility of developing any models. Subsequently it has proved possible to develop robust formulae for this PoC, but they are based on only one of the available data sets. The following notes describe the main characteristics of the five potential data sets and the reasons for excluding most of these from the modelling.

Activity data from SOS CARE

The activity data provided from SOS CARE consisted of a snapshot of clients on a single date and included 29105 persons in PoC7. Of these, approximately 20% were aged over 65. We have been advised to exclude them from the modelling. It also included 11.3% of clients aged 18 or under. We have retained these.

Table EX.1 Numbers of clients by age, sex and Trust (PoC7 SOS CARE extract)

	Males aged:			Females aged:			All
	0-24	25-44	45-64	0-24	25-44	45-64	
S&E Belfast	277	255	607	196	367	841	2543
N&W Belfast	621	574	1065	565	618	1050	4493
Down&Lisburn	207	165	399	172	217	432	1592
NDown&Ards	107	127	301	101	238	556	1430
Causeway	113	147	333	82	202	439	1316
Homefirst	205	351	925	131	473	1253	3338
Armagh&Dungannon	170	202	387	98	214	447	1518
Newry&Mourne	129	146	343	120	222	408	1368
Craigavon&Banbridge	164	244	509	124	298	777	2116
Foyle	155	107	305	112	193	328	1200
Sperrin Lakeland	208	181	424	142	213	505	1673
	2356	2499	5598	1843	3255	7036	22587

In this type of exercise we would expect to model both the numbers of clients per ward and the numbers of clients weighted by the cost of services they individually consume. The cost weights would be computed from details of the services supplied to each client or by assigning clients to costed groups (as is the case for the cost weights in the Learning Disability and Family and Child Care modelling). The principal difficulties we encountered when trying to apply either approach to PoC7 were due to the variations in the services that Trusts record on SOS CARE and the different units used to record activity. When these problems are combined with the differences in the methods used to compile the FR22 costings, we felt that any resulting cost weights would be subject to variations that would not be eliminated by the use of Trust dummies and reluctantly abandoned attempts to attach costs weights to these data. Hence we concentrated on modelling the (unweighted) numbers of PoC7 clients per ward.

A further limitation of these data is the lack of information on the ward of origin for the 253 clients in residential and nursing care. As a result, these clients have to be excluded from the analyses. We have also to remove those receiving domiciliary care for whom there is no valid ward code. To compensate for these exclusions, the data set is weighted-up (retaining the age and sex distributions) to a total of 22587. The corresponding rates per 1000 persons in the population are shown in Table EX.2. As these rates vary with age and sex they will need to be standardised before being used in the modelling.

Table EX.2 Numbers of clients per 1000 persons in relevant age and sex group in population

	Males aged:			Females aged:		
	0-24	25-44	45-64	0-24	25-44	45-64
S&E Belfast	8.4	8.8	30.6	5.8	11.8	39.8
N&W Belfast	22.2	33.0	77.5	20.7	29.4	69.7
Down&Lisburn	6.3	6.7	21.6	5.6	8.4	22.8
NDown&Ards	4.4	6.1	16.4	4.4	11.0	29.3
Causeway	6.4	10.8	30.1	4.7	14.2	38.8
Homefirst	3.5	7.3	25.6	2.3	9.6	33.9
Armagh&Dungannon	8.6	13.8	35.8	5.2	14.9	41.7
Newry&Mourne	7.1	11.5	37.8	7.0	17.4	44.7
Craigavon&Banbridge	7.4	13.5	40.1	6.0	16.6	59.5
Foyle	4.6	4.5	19.2	3.5	7.8	20.3
Sperrin Lakeland	9.0	10.7	32.5	6.6	12.9	40.3

Activity data from systems such as LCID and CLAN

Most Trusts use total activity systems such as LCID and CLAN to record both administrative data and details of services. We hoped to use data from these systems as an alternative to SOS CARE; though we expected some variation in the scale and quality of these data as Trusts use different systems for their costed activity recording (some only keep manual records) and may use different units to record the same activity.

Table EX.3: Service coverage of data from LCID and CLAN

In the table, community nursing includes: community nursing, district nursing, health visiting, twilight nursing, practice nursing and school nursing

Trust	Comm nursing	Physio	OT	Speech & Lang	Aids & adapts
Causeway	x	x	x	x	
Craigavon & Banbridge	x				
Down & Lisburn	x	x	x	x	
Foyle	x		x	x	
Homefirst	x	x		x	
North & West Belfast	x	x	x	x	x
South & East Belfast	x	x	x	x	x
Sperrin Lakeland		x	x	x	
Ulster Community	x		x	x	

Table EX.3 shows the main services covered in these datasets. Apart from their partial coverage, a basic problem with these data is that they were supplied as separate files for each service in each Trust, but do not contain a unique client identifier. So if we add the files there will be multiple counting of clients. To avoid this problem, several strategies were used to combine the datasets. When Trusts supplied date of birth, age, sex and full postcode we could create an identifier that has a high probability of being unique within Trusts, but more approximate methods had to be used when the data sets contained less detail.

Further difficulties were encountered in the processing of these data. For example, nearly half the ward codes were missing for one Trust and this had to be dropped. After extensive cleaning and translation we could only produce data relating to 374 of

the 524 synthetic wards in Northern Ireland. Moreover, a relatively high percentage (11%) of these wards accounted for five or fewer clients, which could create instabilities in the modelling. Nevertheless we persisted to the point of constructing an age-sex standardised variable from these data.

There are some notable differences between these data and the SOS CARE extract. In particular, there are less 0-24 year olds in the LCID/CLAN material.

TROJAN DATA for PoC7

In preliminary discussions of data quality, it was suggested that the lack of a pre-care address on SOS CARE for those in residential care might be resolved by obtaining address information from TROJAN. (We could have used the next-of-kin (NoK) ward data from SOS CARE but this seemed unwise without knowing the client relation to NoK. Moreover, we were discouraged from pursuing this route, as it would have required an additional SOS CARE extract and a new set of permissions).

We translated the TROJAN PoC7 client data into an analysable format, but only (approx) 118 (34.8%) of the 339 PoC7 (or PSD) clients on Trojan appear to have useable pre-care home postcodes. Almost all of these are in three Trusts: Armagh and Dungannon, Foyle and Homefirst. Pre-care postcodes are missing for most cases in the remaining 8 Trusts. We also do not know if the clients already appear on SOS CARE. Hence we have not used the Trojan data in the modelling.

Data on deaths due to the Troubles (for PoC7).

The Department's *Needs & Effectiveness Evaluation* (NEE) estimates that the additional cost burden of the 'Troubles' amounts to £18 million per annum. Part of this will be the costs of caring for patients with physical and sensory disabilities arising as a consequence of the 'Troubles' – and such costs may fall disproportionately on Trusts and Boards across Northern Ireland.

We have acquired a dataset giving ward of residence and ward of incidence for 3696 deaths associated with the Troubles. We hoped to use these data as a proxy for more general morbidity associated with the Troubles, resulting in a need for PoC7 services.

Unlike most of the other data sets described here, this is not a cross-sectional snapshot, but a record of incidents over more than 30 years (Table 23). If persons injured, rather than killed, had the same age distribution, mainly aged 18-25 in 1970-9 and 1980-9, we would expect a peak in the numbers of men with physical disabilities aged 45 and over in 2002 and that this would be most noticeable in North and West Belfast. Such an effect can be detected in the SOS CARE extract where N&W Belfast is the only Trust for which the proportion of males aged 45-64 exceeds the rate for females in the same age group.

As the age-sex distribution of SOS CARE clients suggests that it already includes people whose disabilities were related to the Troubles we thought it would be inappropriate to further weight the SOS CARE material to take account of these effects.

Hospital Incidence Data for PoC 7

The definition for PoC 7 stipulates that hospital expenditure can only be reported under this Programme where services are provided from wards or other inpatient facilities solely for use of physical and sensory disabled patients. This effectively restricts reporting to the costs of providing specialist services – as mildly disabled patients are generally treated, where such is required, in general acute wards. Such costs are included under PoC 1 (Acute Services), and these are not capable of disaggregation to isolate patients otherwise presenting for PSD reasons.

In order to map the need for services for physical and sensory disability, regardless of the present service funding arrangements, we have tried to identify people with needs typical of PoC7 who are receiving hospital care outside specialist units.

We have devoted considerable time and effort devising criteria based on diagnostic profiles and length of stay that could identify this group, but have reluctantly closed this line of enquiry having been unable to arrive at a suitable specification. Consequently we have excluded all the data from hospital sources, although some hospital patients will be included if they are recorded on SOS CARE. We recognise

that this is a high cost group, but their numbers are so small that their omission should have no impact on the existing modelling which is based on client numbers not costs.

Choice of Needs drivers

A great deal of effort went into the construction of the dataset of potential needs drivers for the modelling – parts of which have also been used for the Learning Disability and Family and Child Care research. The full data set contains 81 variables of which the majority are derived from small area data from the 2001 Census, though there are also a number of indicators based on claimant counts and other administrative sources. The potential drivers for the PoC7 modelling fall into three main groups

General deprivation and social structure

In addition to the components of the Noble index, the variables chosen under this heading included measures of (and proxies for) low income as well as indicators of housing conditions. We also included several measures of social structure, such as one-parent households, that that are often associated with material deprivation.

Illness and general morbidity

The morbidity variables are based on self-report data from the census. SMRs are also included in the data set.

Child deprivation and circumstances of families with children

A selection of these variables was included because we were recommended to retain clients aged 17 and under in the analyses.

Supply variables

As for many health and social services, there is a possibility that the supply of, and access to, services for people with physical and sensory disabilities, may influence the volume of services used. Ward level measures of access to services are not always easy to obtain or compute. One ‘supply’ variable was tested initially: the minimum travel time to social services offices. This was intended for the FCC modelling, but used in the absence of more specific information on the location of services for PoC7 clients. Subsequently we have received more details of the social work offices that

support PoC7 clients and have, eventually, managed to compute a suitable supply variable from these data.

Modelling methodology

The overall approach to the modelling has been to start with a large number of potential needs drivers and progressively eliminate the less powerful or unstable predictors by a combination of statistical and theoretical criteria. The main steps are as follows:

- All possible variables were included in the regression model.
- Variables with both counter-intuitive signs and standard errors greater than their respective coefficients were eliminated. In the initial stages, due to the large number of variables, two or three variables were deleted at each re-estimation of the model.
- Variables with counter-intuitive signs irrespective of their significance level were rejected.
- Variables with intuitively correct signs were rejected on the basis of lack of statistical significance (selection criteria: $p > 0.05$). At each re-estimation of the model any variables resulting in counter-intuitive signs were eliminated prior to searching for non-significant variables.
- Where the model appeared to be too narrowly based, attempts were made to “force” variables relating to other types of phenomenon into the equation.
- modelling was repeated with and without the inclusion of the two supply side variables.
- the supply side variables were tested for endogeneity with the need drivers in the final models.
- variables relating to phenomenon not covered by variables already in the model were forced into the model when there was evidence of poor specification.
- other strategies used to improve specification included: changing the functional form and testing the influence of outliers.

The above methodology was applied to two data sets: one derived from SOSKARE, the other from the LCID and CLAN type systems. Only the work with SOSKARE was extensively pursued.

Models from LCID and CLAN data

After cleaning and translation, the merged activity data from LCID/CLAN and similar systems had limited geographical coverage: 374 of the 524 synthetic wards. Moreover, the client numbers were so small in 11% of these wards, that the data might not be capable of supporting robust modelling. Because of these limitations, we reduced our expectations of these data and mainly used them as a check on the models derived from SOSKARE.

There were two main problems in trying to model the LCID/CLAN material: models had low explanatory power and were consistently misspecified. A single variable LOG-LOG model (Table EX.4) came closest to being specified, but even here the reset test coefficient was significant at 3%). Adding further logged variables might improve the specification, but we decided to abandon this line of enquiry as the performance of these models were consistently inferior to those based on the SOSKARE data. The tests had, however, proved the point that despite the relatively low correlation between the SOSKARE and LCID/CLAN derived dependents ($r=0.497$), much the same variables, the Census derived health variables, or very close correlates, were emerging as the best predictors.

Table EX.4 Model based on LOG transformation of LCID/CLAN derived variable

	Unstandardised coeff		Std coeff Beta	T	Signif T	R squared
	B	S.e.				
(Constant)	-0.180	0.017		-10.549	0.000	
Reset	-0.347	0.160	-0.084	-2.167	0.031	
GNGHS	0.944	0.070	0.597	13.473	0.000	
Trust dummies						29.8
Model						54.5

GNGHS is a logged version of the proportion of people reporting “not good” health in the Census.

Models from the SOS CARE data set

Much of the effort in this project has been directed at translating the many and diverse sets of activity data. In the event, only one has been sufficiently comprehensive or robust to support modelling and, even here, it has not been possible to include cost weights in the modelling, which has been based on client numbers. In order to control for age-sex effects and improve the distributional properties of the dependent variable, we have used the log of the age-sex standardised numbers of clients per ward in all the final stages of the modelling.

Although we have not been able to develop a costed dependent, the resulting models are well specified and parsimonious – and include a plausible set of variables.

We were provided with several possible supply variables prior to a final version, that we computed, based on the location of social work facilities for PoC7 clients. This proved to be endogenous when tested with likely models and we therefore adopted a Two Stage Least Squares process (2SLS) to select the variables for inclusion in the final model. This process identified three variables for the final model. Of these, the top two in Table EX.5 (GLLTU65 and GADDLA) are by far the stronger predictors of activity and are the more robust when the model is re-run with data from subsets of Boards and Trusts.

Table EX.5 Variables in the Final Model

	Short definition - all are logged
GLLTU65	Limiting long-term illness amongst under 65s
GADDLA	Adults on disability allowance
GINSC	Noble income score

The final stage in the two stage process is to recompute the models with OLS, without the supply variable. This will provide the coefficients for the allocation formula. The results of this stage are shown in Table EX.6. We are torn between recommending the two or three variable models in the table. Their explanatory power is very similar. Our calculations have found that both models are well specified (when tested in a 2SLS context), though the peer reviewer had some doubts on the specification of the two-variable model and we certainly found that the test statistic is somewhat stronger

for the three variable version. The three variable version provides a wider range of themes in the need drivers, and, importantly, the addition of the income score variable broadens the responsiveness of the model to income deprivation effects. On the other hand, the third variable, the income score, is both less robust than the other two and has had to be computed by translating data from an old ward base. Whilst we would be happy to see either model adopted and leave the final decision to the CFRG, the three variable model is the slightly stronger candidate on the grounds of specification. Our own calculations have shown that the two and three variable models have very similar allocational implications.

Table EX.6 Coefficients when variables selected by 2SLS are re-used in OLS log-log modelling.

Variable	Dummies only	Coeff in 3-var model (and signif of t-value)	Coeff in 2-var model (and signif of t-value)
GLLTU65		0.559 (.000)	0.750 (.000)
GADDLA		0.376 (.000)	0.377 (.000)
GINCSC		0.114 (.004)	
(Constant)		0.198	0.357
Adjusted R-squared (%)	29.6	79.1	79.0

Given the strong correlations between the various needs drivers, it is very likely that other combinations of variables could perform as well as the models we are suggesting. However when we have substituted alternative variables covering similar topics and forced in variables when the range of drivers appears to limited, the suggested models proved surprisingly resilient to these tactics. In particular, the GLLTU65 and GADDLA variables emerge strongly in almost all the modelling.

The models are robust to exclusions of entire Board's data and the correlations between GLLTU65, GADDLA and activity are remarkably consistent across most of the Trusts. On these bases we are happy to recommend either model.

We cannot find evidence of unmet need related to deprivation that requires any correction to the original coefficients, nor of obvious non-linearities in the model itself, which might be corrected by a spline adjustment.

We would have liked to have included a greater range of data sources on activity in the modelling, but the weakness of the individual data sets and the possibility of double counting were two major obstacles. We have devoted considerable time and effort to attempts to specify and obtain data on PoC7 clients in specialist hospital units and other patients whose needs are similar, but who are in more general hospital wards. We have reluctantly closed this line of enquiry having been unable to arrive at a suitable specification. Consequently we have excluded all the data from hospital sources, although some hospital patients will be included if they are recorded on SOS CARE. We recognise that this is a high cost group, but their numbers are so small that their omission should have no impact on the existing modelling which is based on client numbers not costs.

Because the existing allocations are computed from SMRs, we took particular care to compare the predictive ability of SMRs with the Census derived self-report health variables that emerged so strongly in the modelling. We conclude that the Census variables are much better predictors of service activity than SMRs and therefore have no hesitation recommending these new models over the existing formula.

**Modelling the Distribution of Services for People with
Physical and Sensory Disabilities in Northern Ireland
(The PoC7 Programme of Care)**

THE MAIN REPORT

1. INTRODUCTION AND CONTEXT

1.1 Introduction

On 14th October 2002, the Department of Health, Social Services & Public Safety (DHSSPS) on behalf of the Capitation Formula Review Group (CFRG) commissioned MSA Ferndale to undertake research on an additional needs index for formula funding to HSS Boards for services delivered within the Physical & Sensory Disability Programme of Care (PoC7).

A two-phase approach was agreed:

Phase I: a short scoping study to determine whether available information on the population of physically and sensorially disabled persons in Northern Ireland is of sufficient quality and coverage to support development of an index of additional need (i.e. over and above known age and gender effects); and

Phase 2: pending acceptance by CFRG of the findings and recommendations of Phase I, to collate the required information and undertake appropriate statistical analysis to derive an index suitable for implementation.

1.2 Context

The DHSSPS and the four HSS Boards in Northern Ireland have been engaged in a rolling programme of work since 1994 aimed at improving the weighted capitation formula for the allocation of revenue resources to the HSS Boards. Prior to this, allocations were made using the old Proposals for the Allocation of Revenue Resources (PARR) formula, itself derived from earlier work by the Resource Allocation Working Party (RAWP) in England.

The programme of work is being taken forward by the Capitation Formula Review Group (CFRG), which comprises a multi-disciplinary team drawn from the Department and Boards. The CFRG's terms of reference are:

“...to ensure that the resource allocation formula, for use in the distribution of resources to Board level, provides the best measure of relative need for health and social care in Northern Ireland”

The aim of the CFRG research programme is, therefore, to ensure that as far as possible the allocation mechanisms (usually expressed as formulae) are consistent with the ‘fair shares’ principle. This principle requires funding distribution to take account of:

- “ relative differences in the levels of need within each Board’s population for relevant health and social care programmes; and
- “ legitimate differences in the unavoidable costliness of providing services to meet need, *ceterus paribus*.

The CFRG published their Third Report in October 2000, setting out the very significant progress made in developing formulae for each of the nine Programmes of Care in Northern Ireland. In relation to differential need, the formula for each Programme contains adjustments in respect of:

- “ the relative needs of different age and gender sections of the general population - and the resulting impact on per-capita costs of provision; and
- “ the additional per-capita effect arising from factors known to be linked with variations in levels of need from particular individuals or communities – such as low birth weight or deprivation.

The additional per-capita costs for many Programmes of Care are reflected in the formulae through ‘additional needs indices’, which were established through empirical research in Northern Ireland. To date, however, no such research has been undertaken for the Physical & Sensory Disability Programme – and the Under 75 Standardised Mortality Ratio is currently used as a default additional needs index for this Programme.

In the Third Report, the CFRG “emphasised the importance of developing a research-based needs weight for this Programme”.

The Phase I Report of this study explored the feasibility of developing such an empirically based needs index for the Physical & Sensory Disability Programme of Care in Northern Ireland. The index, if developed, would replace the interim needs weighting currently in use.

1.3 Definition of Programme of Care

The Physical and Sensory Disability Programme of Care is defined fully in Section 8.1 and 8.2 (page 73) of the Third Report of the CFRG, as follows:

Definition of patients and clients who can be allocated to the PoC:

"A permanent physical impairment resulting in a dependency in areas such as mobility, self-care, communication and social/leisure activities. Examples of services provided might be rehabilitation for independent living, employment rehabilitation, care services and family support. The patient/client should be under 65 years old."

It is important also to note that, according to the definition in the CFRG Report, hospital activity and related costs can only be allocated to the Physical & Sensory Disability PoC on the basis of entire wards, clinics or hospitals which treat only physically and/or sensory disabled patients. On the face of it, this means that even if a hospital episode or out-patient attendance occurs as a direct consequence of a physical or sensory disability, it should not be included in the PoC unless all patients using the ward or clinic concerned have such a disability.

The definition makes clear that “there will be persons with disabilities who will for definitional reasons, particularly related to age, be included in other PoCs”, and that the needs of such persons should be excluded. These demarcations could pose problems for the modelling if they create a data set that is defined by criteria relating to the source of funds rather than either the need for services or existing supply.

2. ACTIVITY DATA FOR THE MODELLING.

Phase I of the project identified the following five data sources with the potential to provide costed information on the services delivered to PoC7 clients:

- SOSKARE extracts for PoC7
- TROJAN DATA for PoC7
- community health and domiciliary care activity and finance data from LCID and CLAN
- data on deaths due to the Troubles (for PoC7)
- Hospital Incident Data for PoC7.

For the purposes of small area modelling, these data must include the services supplied to individual clients and the post- or ward code of origin of that client. They should also contain a unique client identifier that can be used to guard against double counting. Unfortunately, these details were not present in all of the above data sets. The problems this creates are described in sections 2.2 to 2.6.

2.1 The area base for the modelling

Effective small area modelling must be based on areas that have sufficiently large populations to have robust values for activity data and needs drivers but not be so large that they cover a variety of socio-economic circumstances that could obscure any relations between activity and conditions.

Electoral wards are the obvious choice as most Census and administrative statistics are available by ward, but rural wards may have populations that are too small to generate robust values for indicators of social conditions and service activity. To avoid the problem of small rural wards, we considered using the next larger group of administrative units, District Electoral Areas, but tests with both Census values and deprivation indicators have shown them to be socially heterogeneous and unlikely to produce conclusive models. Hence we felt we had no alternative but to construct a new set of “synthetic wards” Previous modelling had relied on a set of synthetic wards based on the 1984 ward configurations, but current Census and administrative

statistics are mostly based on 1992 wards, which had to be the basis for our new synthetic wards.

In devising these wards, the aim was to produce areas with populations that are always greater than 2000 persons. The algorithm to produce these was based on the grid references of ward population centroids (computed from the postcodes of GP registrations). Wards with populations under 2000 were combined with the nearest ward that was not already part of a synthetic ward. Using these criteria, and a subsequent manual check for the coterminosity of combined wards, we reduced the 582 wards (based on 1992 boundaries) to 524 synthetic wards. A final set of checks ensured that no synthetic wards crossed Local Health and Social Care Group (locality) boundaries. The minimum population of these synthetic combinations is 2026 and the maximum is 9572 with a mean of 3216. Of the 524 synthetic wards, 473 are uncombined wards with the same boundaries as the 1992 wards, 44 are combined with one other ward and 7 with two other wards. The numbers of synthetic wards, and their average populations in each Trust, are shown in Table 1. Average ward populations range from 2500 to 3500 in all except the two Belfast Trusts.

Table 1: Numbers of synthetic wards – and average ward populations (by Trust)

	N synth wards	Popln	Av pop per ward
Armagh&Dungannon	39	101963	2614
Causeway	33	99196	3006
Craigavon&Banbridge	41	119760	2921
Down&Lisburn	53	172482	3254
Foyle	51	162267	3182
Homefirst	109	327762	3007
N&W Belfast	27	143491	5314
NDown&Ards	48	149629	3117
Newry&Mourne	31	89338	2882
S&E Belfast	47	200361	4263
Sperrin Lakeland	45	118935	2643
Total	524	1685184	3216

The next sections describe how we have attempted to construct summaries of service activity from the different data sets for these 524 synthetic wards.

2.2 SOSKARE extracts for PoC7

We received SOSKARE extracts for all eleven Trusts. These contained details of 42,685 clients of whom 29,105 were in PoC7. In theory, PoC7 clients should be aged under 65, but the SOSKARE extract contained approximately 20% of clients aged 65 and over (see Table 2). Being outside the definition of PoC7, we have been advised to exclude these from the modelling. It also included 11.3% of clients aged 18 or under. We have retained these.

Table 2: Numbers of PoC7 clients aged 65 and over

Trust	Clients in PoC7 :	
	Aged 65	Aged over 65
South & East Belfast	39	329
North & West Belfast	98	3111
Down & Lisburn	35	388
North Down & Ards	2	31
Causeway	35	299
Homefirst	60	168
Armagh & Dungannon	4	21
Newry & Mourne	41	199
Craigavon & Banbridge	66	684
Foyle	6	13
Sperrin Lakeland	31	131
Total	417	5374

Missing ward codes

After a preliminary round of postcode/address cleaning, ward codes could not be attached to 9.6% of client addresses and the percentages of missing codes were particularly high in two Trusts: Causeway (40%) and Homefirst (27%) (see Table 3).

A subsequent set of cleaning reduced the overall level of missing codes to 4% (see Table 4). It is worth noting that all these codes are based on current address, which will not be the originating address for those in care.

Table 3: Missing ward codes after first round of cleaning

Trust	Number without ward code	Percent	Total(all clients)
South & East Belfast	103	3.50	2943
North & West Belfast	387	4.90	7895
Down & Lisburn	86	4.13	2082
North Down & Ards	7	0.48	1468
Causeway	675	39.89	1692
Homefirst	969	27.00	3589
Armagh & Dungannon	60	3.80	1581
Newry & Mourne	103	5.73	1796
Craigavon & Banbridge	118	3.99	2955
Foyle	57	4.52	1260
Sperrin Lakeland	221	11.98	1844
Total	2786	9.57	29105

Table 4: Missing ward codes after second round of cleaning

Trust	Percent Missing
South & East Belfast	3.50
North & West Belfast	4.90
Down & Lisburn	4.13
North Down & Ards	0.48
Causeway	0.65
Homefirst	0.61
Armagh & Dungannon	3.80
Newry & Mourne	5.73
Craigavon & Banbridge	3.99
Foyle	4.52
Sperrin Lakeland	11.98
Total	4.04

Information on Service Use from SOS CARE

Data on use of services was included in the SOS CARE extract, but the coverage varies between Trusts. Services covered by at least some of the Trusts include: nursing homes, residential centres, non-residential centres and day care service, adaptations, aids and equipment, home helps days, other domiciliary services, social worker days and occupational therapist days. The numbers of clients receiving these services are shown in Tables 5-7.

Table 5: Client use of Nursing homes, residential centres, non-residential centres and day care service (Numbers of clients using these services)

	Nursing homes	Residential homes	Non-residential centres	Day Care
S&E Belfast	0	10	276	1
N&W Belfast	0	51	271	7
Down&Lisburn	4	18	135	3
Ndown&Ards	0	1	92	0
Causeway	0	1	94	0
Homefirst	3	31	253	12
Armagh&Dungannon	0	13	36	0
Newry&Mourne	0	2	86	2
Craigavon&Banbridge	0	0	31	4
Foyle	48	12	61	2
Sperrin Lakeland	8	51	179	0
	63	190	1514	31

Table 6: Adaptations, aids and equipment (Numbers of clients receiving these services)

	Adaptations	Equipment
S&E Belfast	411	773
N&W Belfast	16	103
Down&Lisburn	0	51
Ndown&Ards	0	29
Causeway	89	344
Homefirst	344	1356
Armagh&Dungannon	0	46
Newry&Mourne	0	74
Craigavon&Banbridge	0	85
Foyle	0	74
Sperrin Lakeland	35	0
	895	2935

Table 7: Home helps days, other domiciliary services, social worker days and OT days. (Numbers of clients receiving these services)

	Home help	Other Dom services	Social work	Occupational therapy
S&E Belfast	340	2	1704	2011
N&W Belfast	527	238	2264	1180
Down&Lisburn	219	22	1942	216
Ndown&Ards	79	425	1464	0
Causeway	229	8	916	1134
Homefirst	310	0	1768	2612
Armagh&Dungannon	230	0	1076	689
Newry&Mourne	179	136	1136	980
Craigavon&Banbridge	361	1	1784	1761
Foyle	252	66	1218	19
Sperrin Lakeland	336	53	1346	859
	3062	951	16618	11461

In this type of modelling exercise we would expect to model both the numbers of clients per ward and the numbers of clients weighted by the cost of service they individually consume. The cost weights might be computed from details of the services supplied to each client or by assigning clients to costed groups (as is the case for the cost weights in the Learning Disability and Family and Child Care modelling). The principal difficulties we encountered when trying to apply either approach to PoC7 were due to the variations in the services that Trusts cover with SOS CARE and the different units they use to record activity. When these problems are combined with the variations in the methods used to compile the FR22 costings, we felt that any resulting cost weights would be subject to variations that would not be eliminated by the use of Trust dummies and reluctantly abandoned attempts to attach costs weights to these data. Hence we concentrated on modelling the (unweighted) numbers of PoC7 clients per ward.

A further limitation of these data is that we have no information on originating ward for the 253 clients in residential and nursing care and these clients have to be excluded from the analyses. We also removed those receiving domiciliary care for whom there is no valid ward code.

To compensate for these exclusions, the data set is weighted-up (retaining the age and sex distributions) to a total of 22587. There are 22641 clients aged under 65 in PoC7 in the original dataset. A small number (54) of cases have been dropped due to missing information on age and sex, but as these are relatively evenly distributed across Trusts their omission will have no impact on the modelling. The resulting numbers of clients per Trust (by age and sex) are shown in Table 8. The corresponding rates per 1000 persons in the population are in Table 9. It is clear from the second of these Tables and from Figure 1 (the age distribution of the clients) that these rates vary with age and sex and that the dependent variable should be standardised by age and sex. Three age bands have been used for the standardisation and the national rates for the six age-sex groups are shown in Table 10. The distributional characteristics of the standardised variable are discussed at the start of section 4.1, but here it is worth noting that the number of clients per synthetic ward is never less than 5 and that 35% of wards each account for more than 45 clients (Table 11).

Table 8: Numbers of clients by age, sex and Trust

	Males aged:			Females aged:			All
	0-24	25-44	45-64	0-24	25-44	45-64	
S&E Belfast	277	255	607	196	367	841	2543
N&W Belfast	621	574	1065	565	618	1050	4493
Down&Lisburn	207	165	399	172	217	432	1592
NDown&Ards	107	127	301	101	238	556	1430
Causeway	113	147	333	82	202	439	1316
Homefirst	205	351	925	131	473	1253	3338
Armagh&Dungannon	170	202	387	98	214	447	1518
Newry&Mourne	129	146	343	120	222	408	1368
Craigavon&Banbridge	164	244	509	124	298	777	2116
Foyle	155	107	305	112	193	328	1200
Sperrin Lakeland	208	181	424	142	213	505	1673
	2356	2499	5598	1843	3255	7036	22587

Table 9: Numbers of clients per 1000 persons in relevant age and sex group in population

	Males aged:			Females aged:		
	0-24	25-44	45-64	0-24	25-44	45-64
S&E Belfast	8.4	8.8	30.6	5.8	11.8	39.8
N&W Belfast	22.2	33.0	77.5	20.7	29.4	69.7
Down&Lisburn	6.3	6.7	21.6	5.6	8.4	22.8
NDown&Ards	4.4	6.1	16.4	4.4	11.0	29.3
Causeway	6.4	10.8	30.1	4.7	14.2	38.8
Homefirst	3.5	7.3	25.6	2.3	9.6	33.9
Armagh&Dungannon	8.6	13.8	35.8	5.2	14.9	41.7
Newry&Mourne	7.1	11.5	37.8	7.0	17.4	44.7
Craigavon&Banbridge	7.4	13.5	40.1	6.0	16.6	59.5
Foyle	4.6	4.5	19.2	3.5	7.8	20.3
Sperrin Lakeland	9.0	10.7	32.5	6.6	12.9	40.3

Figure 1 – Age distribution of PoC7clients on SOS CARE

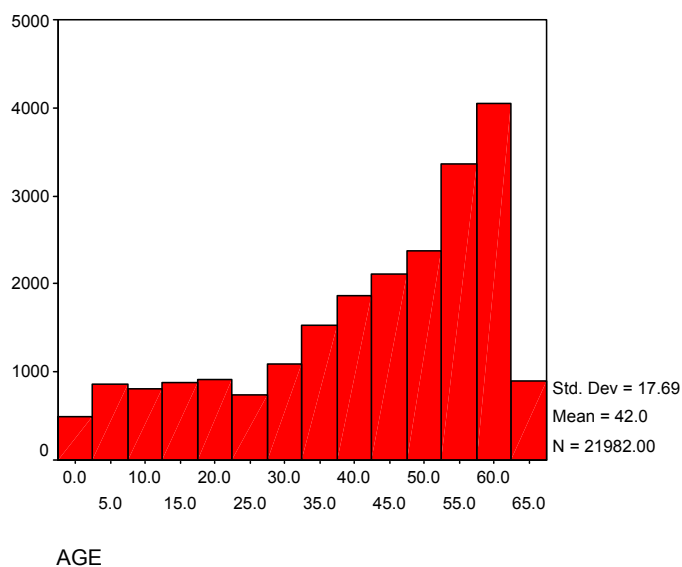


Table 10: Number of clients per 1000 in population (all N Ireland)

Age	Males	Females
0_24	7.58	6.17
25_44	10.41	13.06
45_64	29.61	36.25

Table 11: Numbers of clients in synthetic wards

Client No.s	Number of wards	Percent of wards
5 to 15	27	5.15
16 to 25	97	18.51
26 to 45	217	41.41
46&over	183	34.92
Total	524	100

2.3 PoC7 data from activity systems such as LCID and CLAN

Most Trusts use total activity systems (e.g. LCID and CLAN) to record both administrative data (e.g. patients' details) and details of services provided (see Table 12). The latter tends to be captured via professionals' diaries, though it is possible that some details will be based on recommended levels of service from care plans rather than actual care delivered.

We hoped to use data from these systems as an alternative to SOS CARE; though we expected some variation in the scale and quality of these data as Trusts use different systems for their costed activity recording (some only keep manual records) and may use different units to record the same activity.

Table 12 Systems used to keep records of community health and domiciliary care activity and costs

Trust	Information system used to record PSD community health contacts:		
	DN/HV	OT	PAMS
North & West Belfast	LCID	LCID	LCID
South & East Belfast	Paris/ CommWise	Paris/ CommWise	Paris/ CommWise
Down Lisburn	LCID	LCID	LCID
UCHT	LCID	LCID	LCID
Causeway	LCID	LCID	LCID
Homefirst	LCID	LCID	LCID
Armagh & Dungannon	ERS	ERS	ERS
Craigavon & Banbridge	Manual/DB	Manual/DB	Manual/DB
Newry & Mourne	Manual/DB	Manual/DB & Soscare	Manual/DB
Foyle	CLAN	CLAN	CLAN
Sperrin Lakeland	CLAN	PAS/PHS	PAS

Manual data collection onto in-house developed DataBase

DN – District Nurse; HV – Health Visitor; OT – Occupational Therapist; PAMS – Professions Allied to Medicine.

Table 13 shows the topics covered by the datasets supplied to the project and the number of clients receiving each group of services (or in some cases the number of

instances of each service). It is worth noting that some of these data will not be usable as they lack valid ward- or postcodes.

In order to use these data in the modelling, we need to combine the data sets on separate Trusts and services into a single file and devise a method for generating an overall measure of the service input to each ward.

Table 13: Service coverage of data from LCID and CLAN

In the table, community nursing includes: community nursing, district nursing, health visiting, twilight nursing, practice nursing and school nursing

Trust	Comm nursing	Physio	OT	Speech & Lang	Aids & adapts	Total
Causeway	1418	1916	746	478		4558
Craigavon & Banbridge	7340					7340
Down & Lisburn	493	528	737	181		1939
Foyle	30		1309	137		1476
Homefirst	545	73		249		867
North & West Belfast	27	54	1340	71	6825	8317
South & East Belfast	229	115	1115	87	499	2045
Sperrin Lakeland		97	806	30		933
Ulster Community	81		972	21		1074
Total	10163	2783	7025	1254	7324	28549

Apart from their variable coverage, a basic problem with these data is that they were supplied as a separate file for each service in each Trust, but do not contain a unique client identifier. Hence if we simply add the files there will be multiple counting of clients.

Three strategies were applied to obtain a ward based count of clients from these data.

- 1) When Trusts supplied date of birth, age, sex and full postcode we could use these to create an identifier that has a high probability of being unique within Trusts (though it fails in the case of same sex twins).
- 2) Where we have age, sex and postcode, these were used to produce an identifier that may merge a few cases, but should give reasonable approximations to client counts.

- 3) Where we neither had date of birth nor full postcode on the file there is no satisfactory method for merging files for different services. For these Trusts we have taken the file for the most widespread service (occupational therapy) and used this as the basis for the client count.

The available identifiers and strategies for merging are summarised in Table 14.

Table 14: Identifiers and Merge Strategies for the datasets

Trust	Code no.	DOB	Full pcodes	Strategy for merging
Causeway	2	Y	Y	1
Craigavon & Banbridge	3	Y	Y	1
Down & Lisburn	4	N	N	3
Foyle	5	Y	Y	1
Homefirst	6	N	Y	2
North & West Belfast	8	N	Y	2
South & East Belfast	9	N	N	3
Sperrin Lakeland	10	Y	Y	1
Ulster Community	11	N	Y	2

1=merge by Trust dob age sex and postcode

2=merge by age sex and postcode

3=just use OT

Further problems emerged as these data were being processed. For example, nearly half the ward codes were missing in Homefirst and the only postcode details provided were of the form BT36 and BT37 – which could not be converted to wards. Consequently Homefirst was dropped from the analysis.

After combining the datasets and weighting-up to compensate for cases with partially missing information, we arrived at the client numbers shown in Table 15. The corresponding rates per 1000 persons in the population are shown in Table 16 and overall rates for Northern Ireland in Table 18.

Because some Trusts could not supply all the information required, the merged data set only covers 374 of the 524 synthetic wards. Admittedly, all of these 374 wards account for at least one PoC7 client, but there is a relatively high percentage (11%) of wards with no more than 5 clients (Table 17), which could create instabilities in the modelling.

From the figures in Table 18 it is clear that these client rates will need to be standardised by age and sex prior to any modelling. Trust average values for the standardised variable are shown in Table 19. As there are considerable variations between Trusts we expect that Trust dummies will account for a major part of the overall variance.

Comparing the overall rates with those derived from the SOS CARE data set (Table 10) it is noticeable that the LCID/CLAN activity material produces lower rates for the 0-24 year olds.

Table 15: Numbers of PCO7 clients (by age, sex and Trust)

	Males aged:			Females aged:			All
	0-24	25-44	45-64	0-24	25-44	45-64	
S&E Belfast	20	102	287	22	184	424	1039
N&W Belfast	38	128	398	34	170	578	1346
Down&Lisburn	30	78	183	17	96	227	631
NDown&Ards	38	81	261	30	125	390	925
Causeway	95	107	178	62	126	227	795
Craigavon&Banbridge	44	74	176	23	127	201	645
Foyle	209	134	271	133	199	374	1320
Sperrin Lakeland	108	72	192	84	106	241	803
All in data set	582	776	1946	405	1133	2662	7504

Table 16: Numbers of PoC7 clients per 1000 in population

(By age sex and Trust)

	Males aged:			Females aged:		
	0-24	25-44	45-64	0-24	25-44	45-64
S&E Belfast	0.61	3.52	14.48	0.65	5.91	20.08
N&W Belfast	1.36	7.35	28.95	1.24	8.09	38.37
Down&Lisburn	0.92	3.15	9.92	0.55	3.71	11.97
NDown&Ards	1.57	3.89	14.21	1.31	5.80	20.55
Causeway	5.41	7.85	16.06	3.56	8.84	20.07
Craigavon&Banbridge	1.97	4.10	13.87	1.11	7.08	15.40
Foyle	6.22	5.66	17.10	4.14	8.05	23.11
Sperrin Lakeland	4.66	4.25	14.70	3.92	6.41	19.21

Table 17: Numbers of clients in synthetic wards

N of Clients	N wards	Percent of wards
1 to 5	42	11.23
5 to 15	112	29.95
16 to 25	113	30.21
26 to 45	76	20.32
46&over	31	8.29
All	374	100

Table 18: Number of clients per 1000 in population (all N Ireland)

Age	Males	Females
0_24	2.72	1.96
24_44	4.72	6.55
45_64	15.82	20.93

Table 19: Mean values for unweighted standardised LCID/CLAN derived dependent (by Trust)

Trust	Mean	Wards
Causeway	1.21	33
Craigavon&Banbridge	0.84	41
Down&Lisburn	0.58	53
Foyle	1.32	51
N&W Belfast	1.55	27
NDown&Ards	0.90	48
S&E Belfast	0.85	47
Sperrin Lakeland	1.00	45
All in data set	0.99	345

2.4 TROJAN data for PoC7

In preliminary discussions of data quality, it was suggested that the lack of a pre-care address on SOS CARE for those in residential care might be resolved by obtaining address information from TROJAN. (We could also use the next-of-kin (NoK) ward data from SOS CARE but this seems unwise without knowing the client relation to NoK. Moreover, we were discouraged from pursuing this route, as it would have required an additional extract and a new set of permissions).

We have translated the TROJAN PoC7 client data into an analysable format, but only (approx) 118 (34.8%) of the 339 PoC7 (or PSD) clients on Trojan appear to have useable pre-care home postcodes (see Table 20). Almost all of these are in three Trusts: Armagh and Dungannon, Foyle and Homefirst. Pre-care postcodes are missing for most cases in the remaining 8 Trusts. We also do not know if the clients already appear on SOS CARE. Hence we have not tried to incorporate the Trojan data in the modelling.

Table 20: Availability of postcode of pre-care home address of PoC7 clients from TROJAN datasets

Trust	Notes on extent of prior address info and postcodes	PoC7 Clients	Approx no. with prior address postcode
Armagh & Dungannon	Only 14 clients, but prior postcodes available for 13.	14	13
Causeway	33 clients . For approx 25 of these the prior postcode is the same as the care home postcode.	33	8
Craigavon & Banbridge	No data	0	0
Down & Lisburn	Less than ten clients have relevant postcode.		<10
Foyle	54 clients approx 48 have postcode that differs from care home postcode.	54	48
Homefirst	72 clients. All have an original postcode, but at least half of these are the postcode of the current care home.	72	~30
Newry & Mourne	No data	0	0
North & West Belfast	49 clients, but no column giving a pre-care home postcode	49	0
South & East Belfast	141 LD and PD care home clients, but only 16 prior postcodes for these 141	70	8
Sperrin Lakeland	22 clients - only one has a prior postcode	22	1
Ulster Community	25 clients - none have a prior postcode...	25	0
ALL		339	118

2.5 Data on deaths due to the Troubles (for PoC7)

The Department's *Needs & Effectiveness Evaluation* (NEE) estimates that the additional cost burden of the 'Troubles' amounts to £18 million per annum. Part of this will be the costs of caring for patients with physical and sensory disabilities arising as a consequence of the 'Troubles' – and such costs may fall disproportionately on Trusts and Boards across Northern Ireland.

We have acquired a dataset giving ward of residence and ward of incidence for 3696 deaths associated with the Troubles. We hoped to use these data as a proxy for more general morbidity associated with the Troubles, resulting in a need for PoC7 services. In the original file, home ward details are missing for 955 cases and incident ward details for 717 cases, but we have been able to improve slightly on these (see Table 21).

Table 21: Trust of residence of people killed in the troubles

Trust	Deaths	Percent
No ward details	915	24.76
Armagh&Dungannon	245	6.63
Causeway	46	1.24
Craigavon&Banbridge	154	4.17
Down&Lisburn	151	4.09
Foyle	233	6.30
Homefirst	272	7.36
N&W Belfast	935	25.30
NDown&Ards	63	1.70
Newry&Mourne	134	3.63
S&E Belfast	377	10.20
Sperrin Lakeland	171	4.63
Total	3696	100

Unlike most of the other data sets described here, this is not a cross-sectional snapshot, but a record of incidents over more than 30 years (Table 22). If persons who were injured, rather than killed, had the same age distribution, mainly aged 18-25 in 1970-79 and 1980-89, we would expect a peak in the numbers of men with physical disabilities aged 45 and over in 2002. From the Trust distribution of deaths related to the Troubles (Table 23), we would expect this peak to be most noticeable in North and West Belfast. Such an effect can be detected in Table 24 (showing the age-sex

breakdown of PoC7 clients on SOS CARE) where N&W Belfast is the only Trust for which the proportion of males aged 45-64 exceeds the rate for females in the same age group.

This comparison with the age-sex distribution of SOS CARE clients suggests that these already include people whose disabilities were related to the Troubles, so that the SOS CARE figures do not need to be weighted-up to include the latter group. The data on deaths due to the Troubles could be modelled separately, but for the following reasons such a model is unlikely to be a good basis for PoC7 allocation:

- distribution of age-sex rates differs from those based directly on PoC7 activity;
- the numbers in North Down and Ards and Causeway are too small for synthetic ward level modelling;
- ward level effects will almost all be explained by Trust dummies.

Table 22: Numbers of deaths related to the Troubles (per year)

	Frequency	Percent
Before 1970	19	0.51
1970 to 1979	2192	59.31
1980 to 1989	889	24.05
1990 to 1999	554	14.99
2000 to 2003	42	1.14
Total	3696	100

(81.7% of the deceased were men)

Table 23 Numbers of deaths related to the troubles (by Trust)

Trust	N deaths	N per 1000
S&E Belfast	377	2.40
N&W Belfast	935	9.17
Down&Lisburn	151	1.21
NDown&Ards	63	0.55
Causeway	46	0.63
Homefirst	272	1.12
Armagh&Dungannon	245	3.38
Newry&Mourne	134	2.14
Craigavon&Banbridge	154	1.77
Foyle	233	2.06
Sperrin Lakeland	171	2.01
All	2781	2.25

**Table 24 Age and sex distribution of PoC7 clients on SOS CARE
(rates per 1000 in the population)**

SOS CARE data	Males aged:			Females aged:		
	0-24	25-44	45-64	0-24	25-44	45-64
S&E Belfast	8.4	8.8	30.6	5.8	11.8	39.8
N&W Belfast	22.2	33.0	77.5	20.7	29.4	69.7
Down&Lisburn	6.3	6.7	21.6	5.6	8.4	22.8
NDown&Ards	4.4	6.1	16.4	4.4	11.0	29.3
Causeway	6.4	10.8	30.1	4.7	14.2	38.8
Homefirst	3.5	7.3	25.6	2.3	9.6	33.9
Armagh&Dungannon	8.6	13.8	35.8	5.2	14.9	41.7
Newry&Mourne	7.1	11.5	37.8	7.0	17.4	44.7
Craigavon&Banbridge	7.4	13.5	40.1	6.0	16.6	59.5
Foyle	4.6	4.5	19.2	3.5	7.8	20.3
Sperrin Lakeland	9.0	10.7	32.5	6.6	12.9	40.3

2.6 Hospital Incidence Data for PoC 7

The definition for PoC7 stipulates that hospital expenditure can only be reported under this Programme where services are provided from wards or other inpatient facilities solely for use of physical and sensory disabled patients. This effectively restricts reporting to the costs of providing specialist services – as mildly disabled patients are generally treated, where such is required, in general acute wards. The costs of care for these patients are included under PoC1 (Acute Services), and are not capable of disaggregation to isolate patients otherwise presenting for PSD reasons.

In order to map need for services for physical and sensory disability, regardless of the present service funding arrangements, it would be helpful to be able to identify people receiving care outside specialist units whose diagnostic profile and length of stay approximates those in the units.

Table 25 presents a breakdown of the profile of specialist PSD hospital-based services – both activity and cost. It shows that the *Regional Spinal Injury Service* at Greenpark accounts for circa 75% of the hospital-based service input across Northern Ireland. This disaggregates as £1.765m (26%) inpatient services, and £4.986m (74%) outpatient care on a hospital site. The other major provider is the *Regional Brain Injury Unit* at Thompson House (Down & Lisburn), costing £1.7m in 2000/01.

Table 25: Analysis of specialist PSD hospital services (2000/01)

HOSPITAL SERVICES	IP Episodes	OP Attendance	Cost (£m)	Unit Cost (£)
Regional Spinal Injuries Service (Greenpark)				
Inpatients - Amputee Rehab.	75		0.588	7,840
Inpatients - Spinal Cord Injuries	58		1.177	20,293
Outpatients - Joss Cardwell		2,019	0.542	268
Outpatients - Regional Disablement Service		15,763	3.914	248
Outpatients - PAMs Service to Special Schools		n/a	0.53	
Regional Brain Injury Unit (Down & Lisburn)				
Inpatients - Thompson House	182		1.696	9,319
M.S. Centre (Causeway)				
Inpatients	248		0.636	2,563

Source: Trust Financial Returns (FR22 Ii) for year ending 31st March 2001.

Hospital inpatient services cater to a small number of patients – at a high cost per episode. The Regional Disablement Service at Greenpark Trust undertakes the bulk of outpatient activity.

Hospital information systems

All three units capture information on inpatients and inpatient services using the Patient Administration System (PAS). Information captured includes:

- ◆ patient's name and address;
- ◆ age and gender;
- ◆ postcode;
- ◆ clinical information (activity coded to HRG group); and
- ◆ relevant dates of admission, discharge/death, etc.

The presence of good quality postcode coverage and service information would support a small-area based analysis, though the low number of inpatients per annum in this Programme may necessitate analysis at a synthetic ward level (or higher).

In addition to providing some basic data on PoC7 clients, Thompson House produced a profile of the diagnoses and length of stay characteristics of these clients that could be used as a filter for HIS extracts, in order to estimate the numbers of potential PoC7 clients in hospitals.

We have discussed with relevant physicians what groups of diagnoses and lengths of stay could identify potential PoC7 clients amongst general hospital activity. The first attempt to apply a profile to three years HIS data produced the small numbers shown in Table 26. A profile based on the primary diagnosis codes listed in the Phase I report, when applied to a full year of HIS data (without any restrictions based on length of stay) identifies the 9105 admissions shown in Table 27 – though the small number admitted for diabetes suggests that the codes may only partially cover the named conditions.

We have explored other sets of diagnoses – examples are shown in Tables 28 and 29. From these attempts we tend to the conclusion that it is not possible to specify a set of diagnostic groups (combined with length of stay) that will identify people in the general acute hospital environment who have similar needs to those in PoC7. We have reached this conclusion both from our efforts to select cases from HES data and from discussions with physicians and other consultants who have doubted whether unambiguous criteria can be devised.

Given these difficulties we feel that it would be counterproductive to pursue this further and are recommending that the modelling should not try to include hospital patients who are treated outside the PoC7 specialist units.

We further recommend that the patients in these special units be excluded, because

- of difficulties in obtaining pre-care addresses;
- of possible geographical bias if these patients are included and those in other hospitals cannot be identified; and
- the modelling will be based on client numbers rather than costs and the special hospital group are a very small proportion of the total PoC7 clients. Moreover, there is no reason to assume that all those receiving relevant hospital care will have a different geographical distribution to those being cared for in the community.

Table 26: Numbers of people on HIS conforming to first attempt to construct a POC7 diagnostic profile

Trust	1999	2001	2002
Belfast City	1	0	0
Royal Group	2	5	14
Green Park	5	2	4
UC&HT	4	3	3
Down Lisburn	3	1	6
United	6	3	17
Causeway	4	3	2
Armagh & Dungannon	3	1	5
Newry & Mourne	3	9	2
Altnagelvin	6	3	11
Sperrin Lakeland	1	1	0
Total	38	31	64

Table 27: Numbers of persons admitted to hospital over one year with diagnosis codes listed in Phase I report

AIDS - including HIV infection	220
Alcohol abuse	59
Arthritis - including rheumatoid and osteoarthritis	11
Asthma	297
Back pain - not specified	61
Behavioural disorder - including enuresis and hyperactivity	1080
Blood disorders - including anaemia, clotting disorder and not specified	1
Bowel and stomach disease - including ulcer, constipation and not stated	198
Cerebrovascular disease - including stroke, CVA and hemiplegia	74
Chest disease - including bronchitis, emphysema, COAD and bronchiectasis	1956
Chronic fatigue syndromes - including M.E./post viral	30
Cystic fibrosis	60
Deafness	12
Dementia - including senile, pre-senile, Alzheimer's and traumatic	155
Diabetes mellitus	11
Disease of the muscles, bones or joints	146
Epilepsy	2
Frailty - including senility/not specified	5
Haemophilia	301
Heart disease - including coronary, ischaemic, myocardial or heart attack	66
Hyperkinetic syndrome	43
Inflammatory bowel disease - including Crohn's, ulcerative and colitis	7
Major trauma (other than D35)	99
Malignant disease - including cancer, carcinoma and leukaemia	244
Mental subnormality	41
Metabolic disease - including PKU	40
Motor Neuron disease	24
Multi system disorders, i.e. sarcoidosis, SLE/DLE and scleroderma	76
Multiple allergy syndrome	20
Multiple sclerosis	254
Neurological diseases - not specified	515
Parkinson disease	612
Peripheral vascular disease including thrombosis and claudication	776
Personality disorder	656
Psychoneurosis - including anxiety, depression, phobia and hysteria	413
Psychosis - including schizophrenia and manic depression	36
Renal disorders - including dialysis (not deemed)	13
Spondylosis - including disc disease, cervical/lumbar etc	8
Terminally ill	10
Total parenteral nutrition - any cause	1
Trauma to limbs - loss of fingers/toes/amputation	340
Traumatic paraplegia/tetraplegia	132
All in above list	9105

Table 28: Alternative selection based on (Thompson House?) recommendations (no restriction on length of stay)

	Frequency Percent	
	Frequency	Percent
1 M.S.	1147	7.9
2 Hypoxic & Anoxic Brain Injury	59	0.4
3 Traumatic Brain Injury	27	0.2
4 Epilepsy	3922	27.0
5 Parkinsons	1709	11.8
6 Huntingtons	44	0.3
7 C.V.A.	2183	15.0
8 Diabetes	4137	28.5
10 Sub Arachnoid Haemorrhage	26	0.2
11 Spinal Injury	17	0.1
12 Tardive Dyskineasia	6	0.0
13 Muscular Dystrophy	5	0.0
14 Cerebral Vasculitis	30	0.2
16 Hydrocephalus with Spina Bifida	512	3.5
17 Cerebral Palsy	708	4.9
Total	14532	100.0

Table 29: Alternative selection based on (Thompson House?) recommendations (Lengths of stay from 10 to 60 days)

Condition/diagnosis	Stays of these number of days or more		
	10 days	20 days	60 days
	1 M.S.	403	203
2 Hypoxic & Anoxic Brain Injury	40	32	15
3 Traumatic Brain Injury	12	5	5
4 Epilepsy	810	382	122
5 Parkinsons	940	597	229
6 Huntingtons	13	7	6
7 C.V.A.	1623	1261	647
8 Diabetes	1093	501	115
10 Sub Arachnoid Haemorrhage	19	6	0
11 Spinal Injury	1	1	0
12 Tardive Dyskineasia	2	2	1
13 Muscular Dystrophy	3	2	0
14 Cerebral Vasculitis	2	1	0
16 Hydrocephalus with Spina Bifida	93	49	4
17 Cerebral Palsy	70	31	4
Total	5124	3080	1233

3. Needs drivers and supply variables

3.1 Need drivers for services for people with physical disability and sensory impairment

A great deal of effort went into the construction of the dataset of potential needs drivers for the modelling – parts of which will also be used for the learning disabilities and family and childcare work. It contains 81 variables of which the majority are derived from small area data from the 2001 Census, though there are also a number of indicators based on claimant counts and other administrative sources. A full list of the variables and their distributional properties can be found in Annex 1.

Broadly speaking they fall into three groups.

General deprivation and social structure

In addition to the components of the Noble index, the variables chosen under this heading included measures of (and proxies for) low income as well as indicators of housing conditions. We also included several measures of social structure, such as one-parent households, that are often associated with material deprivation.

Illness and general morbidity

The morbidity variables are based on self-report data from the Census. SMRs are also included in the data set.

Child deprivation and circumstances of families with children

A selection of these variables was included because we were recommended to retain clients aged 17 and under in the analyses.

3.2 Effects of Population density

There has been extensive debate on the role of measures of population density and sparsity in resource allocation formulae. The pressure to include some measure of sparsity tends to come from rural authorities who judge that there will be extra costs in providing some types of services to sparsely distributed populations. Even where there is clear evidence of these extra costs, our argument is that they should be

addressed by specific weightings rather than by attempts to include sparsity in an allocation formula. The problem with the formulaic approach is that in most areas of the UK, sparsity is strongly and negatively correlated with deprivation – because the greatest deprivation is usually found in the densest parts of metropolitan areas. Hence any measure of density will tend to serve as a proxy for urban deprivation and direct money away from rural areas. For this reason we are reluctant to include measures of sparsity or density directly in the modelling, but we recognise two concerns:

- there may be extra costs of service delivery in rural areas; and
- there may be unmet need in (rural) areas with limited access to services or difficulties of supply.

We think that the former should be addressed with targeted cost weights. The latter could be considered when the unmet needs adjustment is applied to the final formulae, but the recommended test and adjustment encounters similar problems to those described above – that it cannot prevent measures of density acting as inverse measures of deprivation.

3.3 Supply side factors

In any modelling of activity there is always the risk that geographical patterns of activity will be distorted by the location of service delivery points. In this case we have been given four datasets of distances from wards to services in order to test for supply side effects. They provide travel times to:

- general practices
- learning disability hospitals
- nursing and residential homes
- social work bases (for Family and Child Care).

Given that 64% of PoC7 expenditure relates to social services, we were most interested to test whether activity was (negatively) related to distance from social work base – though we need to confirm that these bases provide services that are relevant to PoC7 as well as FCC.

Very late in the project we have also received more details of the location of services for PoC7 clients – though only for seven Trusts. These have been translated to grid references and a new variable has been computed: average distance from PoC7 social work bases to ward centroids.

4. The modelling

The overall approach to the modelling is as follows:

- All possible variables were included in the regression model.
- Variables with both counter-intuitive signs and standard errors greater than their respective coefficients were eliminated. In the initial stages, due to the large number of variables, two or three variables were deleted at each re-estimation of the model.
- Variables with counter-intuitive signs irrespective of their significance level were rejected.
- Variables with intuitively correct signs were rejected on the basis of lack of statistical significance (selection criteria: $p > 0.05$). At each re-estimation of the model any variables resulting in counter-intuitive signs were eliminated prior to searching for non-significant variables.
- Where the model appeared to be too narrowly based, attempts were made to “force” variables relating to other types of phenomenon into the equation
- Modelling was repeated with and without the inclusion of the two supply side variables.
- The supply side variables were tested for endogeneity with the need drivers in the final models.
- Variables relating to phenomenon not covered by variables already in the model were forced into the model when there was evidence of poor specification.
- Other strategies used to improve specification included: changing the functional form and testing the influence of outliers.

Careful checks were made to ensure that the statistical model was well specified. The specification of a regression model consists of a formulation of the regression equation and of statements or assumptions concerning the regressors and the disturbance term. A “specification error” in the broad sense occurs whenever the formulation of the regression equation or one of the underlying assumptions is incorrect. Specification errors can occur for various reasons:

- omission of a relevant explanatory variable.
- inclusion of an irrelevant explanatory variable.
- incorrect mathematical form of the regression equation.
- incorrect specification of the covariance structure such that the error term is not normally distributed.

Only well-specified models were considered acceptable, as measured by the widely recognised “Reset” test for specification that involves testing the significance of the squared predicted values of the basic model when added to the model as an independent variable. The intention of the modelling procedure was therefore to derive a model of utilisation, which was (i) plausible; (ii) parsimonious; and (iii) statistically acceptable.

Details of the modelling are reported in the next two sections. The first considers the variables derived from the data set on clients in the community and the second describes the modelling of the variables from the combined data set.

Modelling has been attempted with data derived from SOSKARE and from the LCID/CLAN activity systems. The two sets of models are described in Sections 4.1 and 4.2.

4.1 Modelling client numbers from the SOSKARE data set.

As mentioned earlier, it proved impossible to sensibly weight the clients in the SOSKARE data set to reflect the service each received because the recording of services and expenditure was too incomplete and inconsistent across Trusts. Therefore the modelling had to be based on the numbers (rather than costs) of clients per ward. These data have been aggregated by Trust and client ID to ensure that there is no double counting and have been standardised by age and sex.

The resulting variable has values for both skewness and kurtosis (Table 30) that suggest it will produce poorly specified models. Applying a log transformation produces a distribution that is much closer to normal (Table 31).

Table 30: Properties of unlogged dependent

Minimum		Maximum		Mean		Skewness		Kurtosis	
Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
0.19	3.19	1.00	0.02	1.19	0.11	2.10	0.21		

Table 31: Properties of logged dependent

Minimum		Maximum		Mean		Skewness		Kurtosis	
Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
-0.73	0.50	-0.05	0.01	-0.16	0.11	-0.18	0.21		

The correlations between potential needs drivers and the standardised numbers of clients (after removing the effect of the Trust dummies) show a preponderance of Census derived health variables amongst the strongest correlates (Table 32). Three versions of self-report limiting long-term illness – LLTI, ADLLTI, LLTU65 (all age standardised, 18-64 unstandardised, under 65 standardised) and two measures of self-report poor health (ADNGH NGHS) are amongst the 8 strongest correlates. Measures of the proportion of persons in social rented housing are also some of the stronger correlates, though, as we saw in the modelling of learning disability, these tend to substitute for, rather than complement, the influence of the poor health variables.

Given the significance of the health variables, it is worth noting that standardised mortality ratios, as used in the current allocations, are less strongly correlated with activity than both the Census derived health variables and most of the proxies for deprivation.

Table 32: Correlation coefficients between residual variance of SOS CARE activity variable (after removing effects of Trust dummies) and variables in the standard modelling set. (See Annex for variable definitions)

SOCREN	0.657	0.000	SINGLE	0.290	0.000
LLTU65	0.647	0.000	KJAS	0.272	0.000
KSOCRENB	0.642	0.000	POPDENS	0.242	0.000
ADLLTI	0.641	0.000	KNOCENTH	0.215	0.000
CHILDPSC	0.641	0.000	NOCENTH	0.161	0.000
ADNGH	0.634	0.000	CATHOL	0.130	0.003
LLTI	0.625	0.000	KCATHOL	0.097	0.027
NGHS	0.620	0.000	NWRDS	0.081	0.063
EDUCSC	0.618	0.000	HH3K	0.075	0.085
LOWSEG	0.612	0.000	KPRVRENB	0.061	0.164
NOCAR	0.608	0.000	PRIVREN	0.047	0.281
KIS	0.607	0.000	NONWHT	0.046	0.294
NOCARER	0.597	0.000	HOUSSC	0.042	0.333
ADDLA	0.594	0.000	KNBASIC	0.029	0.513
ADIS1864	0.593	0.000	KPRVRENA	0.012	0.776
LONPU24	0.592	0.000	KLPP	-0.050	0.251
ADIS	0.588	0.000	NOBASIC	-0.060	0.170
INCSC	0.587	0.000	HHDEPK	-0.067	0.124
KNOCARC	0.582	0.000	SNGCARER	-0.075	0.087
LPDEPK	0.574	0.000	PRIVREM	-0.085	0.052
EMPSC	0.572	0.000	HHALLK	-0.093	0.034
U75SMR	0.566	0.000	PROTST	-0.130	0.003
NOQUAL	0.563	0.000	KPROT	-0.146	0.001
HLTHSC	0.562	0.000	ACCSSC	-0.350	0.000
U65SMR	0.561	0.000	MARRIED	-0.520	0.000
KSOCRENA	0.560	0.000	THREECAR	-0.523	0.000
KNOCARB	0.558	0.000	KOWNOCCA	-0.523	0.000
OVSC	0.554	0.000	HIGHSEG	-0.570	0.000
PROPNST	0.543	0.000	KTWOCARA	-0.584	0.000
KNOCARA	0.540	0.000	KTWOCARB	-0.598	0.000
ADUNEMP	0.534	0.000	KTWOCARC	-0.601	0.000
SWIDRCD	0.531	0.000	KHGHSEGA	-0.620	0.000
KWFTC	0.524	0.000	OWNOCC	-0.621	0.000
KLOWSEGB	0.511	0.000	KHGHSEGB	-0.621	0.000
KDLA	0.510	0.000	TWOCAR	-0.633	0.000
ADJSA	0.497	0.000	KOWNOCCB	-0.637	0.000
KLOWSEGA	0.493	0.000			
KLLTI	0.492	0.000			
SOCENVSC	0.491	0.000			
EXMAR	0.476	0.000			
ADWFTC	0.471	0.000			
ADLPP	0.463	0.000			
KOWCRAB	0.459	0.000			
ADIS65OV	0.447	0.000			
KNGH	0.389	0.000			
ALLSMR	0.369	0.000			
KOWCRAA	0.362	0.000			

Indicators of geographical variation in access to services (“supply variables”)

One ‘supply’ variable was tested initially: the minimum travel time to social services offices. This was intended for the FCC modelling, but used in the absence of more specific information on the location of services for PoC7 clients. It was included because of the high level of social services expenditure on PoC7 clients, but we were not expecting a strong correlation, as access to secondary health care may be a more important factor. In the event, the minimum time to social services office variable proved to be statistically significant at >1%, and added approximately 4% of explanation to the model containing only the Trust dummies. As expected, the coefficient is negative, suggesting that client numbers decrease with increased distance from a social services base.

A similar, though less powerful, effect was noted when we used a variable (AVDIST) more specifically related to the location of PoC7 social services. This was available for seven Trusts, was significant at 2%, and increased the explanatory power from 25.2% to 26.4%.

The problem with the second of these variables is that it ceases to be significant when almost any relevant need driver is added to the model. For example, the t value for the coefficient of AVDIST changes from -2.4 to -1.1 when the under-65 standardised limiting long-term illness variables is included. From this we conclude that it may be acting as a proxy for stronger effects that are better represented by the needs drivers.

Towards the end of the project we were able to get additional information on the location of social services for people in the PoC7 programme from all the N Ireland Trusts and compute a new supply variable based on the distance from these facilities. As we already had some provisional LOG-LOG models based on SOS CARE data, we conducted the standard test for endogeneity using these models and the new supply variable.

The test for endogeneity is intended to establish whether the supply variable is, itself, predicted by factors other than legitimate need drivers. The peer reviewer had already posed this question and found that: *“the system is endogenous with regard to a one variable log-log solution ($F_{1,510} = 6.72 p < 0.01$) and is still endogenous when GADDLA is added to produce a two-variable log-log model ($F_{1,509} = 5.26 p < 0.05$). Switching off the weights still retains endogeneity ($F_{1,509} = 4.35 p < 0.01$). There are also 4 (out of 524) wards that exhibit outlying behaviour. Removing these causes endogeneity to increase ($F_{1,505} = 6.52 p < 0.05$) with the unweighted model, and ($F_{1,505} = 8.58 p < 0.01$) on my preferred weighted model. Moving on to a three-variable log-log model (with GKWFTC added) and endogeneity is retained ($F_{1,504} = 7.86 p < 0.01$)”*.

We have repeated these analyses, with similar results, and also find that other selections of instruments and modelling variables can increase the value of the test statistic to 20 or more (the critical 5% point is 3.8). The precise values of the endogeneity test statistic will depend on the set of instruments and the variables in the equation under test.

The basic point is that the new supply variable is endogenous with various sets of instruments when log-log models are used. Hence two-stage methods are required to identify the best predictor variables.

The main features of this two stage method are firstly that the normal OLS procedure for identifying the best predictor variables is replaced by a process in which the supply variable is, in effect, substituted by its residual after it has been regressed on a set of candidate variables or “instruments”. This technique is used to search for the best predictor set of need drivers, which are then put into a simple OLS model (the second stage) in order to compute the coefficients for the allocation formula.

There are various methods for whittling down the full set of potential variables to a plausible subset, but, basically, this was achieved by a combination of statistical and substantive criteria.

The main techniques used were partial correlations, OLS stepwise regressions (including runs with the residual of the supply variable on the full instrument set). These, and related techniques are used to identify a subset of variables for the manual search in 2SLS. Throughout this stage, we paid particular attention to the substantive plausibility of both individual and groups of variables and the effects that intercorrelations between the needs drivers, and between the need drivers and the supply variables, were having on the significance and signs of the coefficients. We also carried out some early 2SLS runs to confirm that they endorsed our preliminary conclusions as to the variables most likely to be selected for the final modelling.

By these means we identified a subset of the most significant need drivers that would be further explored via manual 2SLS techniques.

Derivation of the model by 2SLS

When we re-model the system from scratch using 2SLS, the variables in Table 33 emerge as the strongest predictors.

It is worth noting that, in all of these runs, a full set of Trust dummies are included and, for the sake of consistency, the four outliers identified by the reviewer have been removed, though the results are substantively the same if these wards are retained.

Table 33: Variables emerging in (manual) 2SLS log-log stepwise approach as best multivariate predictors - in descending order of statistical significance

	Short definition - all are logged	Hypothesised sign	Sign in modelling
GLLTU65	Limiting long-term illness amongst under 65s	+	+
GADDLA	Adults on disability allowance	+	+
GINSC	Noble income score	+	+
GPROTST	Proportion Protestants	?	+
GNOQUAL	Proportion with no qualification	+	-
GADJSA	Job seekers allowance	+	-
GNOCENTH	No central heating	+	+

A slightly different process of manual stepwise selection to that used for Table 33 results in a three variable model including: (the log of) adults in disability living households, (the log of) limiting long term illness amongst under 65s and (the log of) households without cars. This model fails on two grounds: it is mis-specified (chi-squared $df_6 = 27.6$) and includes a variable relating to car ownership that has potentially different meaning in urban and rural areas.

Returning to the variables in Table 33, two of these, persons with no qualifications and those claiming job seekers allowance, have counter-intuitive signs - resulting from high correlations with the top three variables in the table. There is no evidence that this counter-intuitive signs are indicative of unmet need effects. The presence of the variable relating to the proportion of Protestants does not influence the two counter-intuitive signs.

The first three variables in the table, GLLTU65 GADDLA GINSC, are all strong contenders for the modelling. GPROTST (the proportion of Protestants) may be hard to defend on theoretical grounds as it is unclear whether this should be positively or negatively correlated with health and social service activity. It also has relatively little impact on overall explanatory power, as does the "no central heating" variable.

By this route we arrive at a three variable model in which the two more powerful predictors are the same as those in the previously reported OLS model. The one difference is that the log of the Noble income score has substituted for the log of the children in working family tax credit households.

Table 34 shows the relative explanatory power and specification of the 2SLS models based on combinations of these three variables, to which has been added the variable used by the peer reviewer (taken from the variables in a previous OLS derived model). Several key points emerge from the figures in this table.

- Most of the explanatory power is provided by GLLTU65
- GADDLA and GINSC each add between 3% and 4% to r-squared
- GKWFTC is not or barely significant and makes little or no contribution to the explanation - it also impacts on the significance of the income score.
- The single variable model is very misspecified, but adding one or more variables results in specification.

Table 34: Combinations of variables in 2SLS models

	Variables in model				
Model number	1	2	3	4	5
GLLTU65	x	x	x	x	x
GADDLA		x	x	x	x
GINCSC			x		x (NS)
GKWFTC				x (NS)	x (NS)
Adjusted r-squared for 2SLS estimation (as %)	72.4	75.9	79.6	76.1	79.3
Specification approx value of chi-squared	~30	~5	~3	~1	~2
DF	4	3	2	2	1
Critical value for spec test	9.5	7.8	6.0	6.0	3.8
Model number	1	2	3	4	5

As these are 2SLS analyses, an alternative to the Reset test has to be used to check for specification. This alternative involves computing an auxiliary equation derived from an OLS regression of the residuals from the 2SLS model against the full instrument and variable set in the 2SLS. The test statistic is the ratio of the sum of squares explained by this auxiliary equation to the estimated error variance for the 2SLS estimated utilisation equation. It should have a chi-squared distribution with degrees of freedom equal to the number of instruments in the auxiliary equation minus the number of regressors in the utilisation equality.

Unlike the reviewer, our calculations suggest that the two variable model (model number 2 in Table 34) is just specified. The reviewer found that only the three variable version is specified. So it is a question of whether to use the two or three variables models (numbers 2 and 3 in Table 34).

Superficially, the r-squared values in the 2SLS analyses suggest that the income score adds approximately 3-4% to the level of explanation. but such values are notoriously difficult to interpret in the context of 2SLS. Consequently, we have carried out a related set of calculations to test the effect of including the income score in the OLS

equation used to predict allocations. Here we find that although the coefficient of the income score is significant, the variable only adds approx 0.3% to the r-squared.

On this basis, and given the potential difficulties with the age of the income score data and the method of transformation to Census wards, there is an argument for recommending the two-variable model - the same as suggested by the peer reviewer, albeit reached by a slightly different route. Against this, however, the three variable model has the advantage of being responsive to income deprivation.

When the variables are used in OLS estimation, without the supply variable and excluding the four outlying wards, we arrive at the (unstandardised) coefficients shown in Table 35.

Table 35: Coefficients when variables selected by 2SLS are re-used in OLS log-log modelling

Variable	Dummies only	Coeff in 3-var model (and signif of t-value)	Coeff in 2-var model (and signif of t-value)
GLLTU65		0.559 (.000)	0.750 (.000)
GADDLA		0.376 (.000)	0.377 (.000)
GINCSC		0.114 (.004)	
(Constant)		0.198	0.357
Adjusted R-squared (as%)	29.6	79.1	79.0

We have one slight concern with the work - that a relatively weak set of instruments is used in 2SLS estimations. To address this point we have repeated the analyses with larger and stronger sets of instruments. The variable choice in the models is nearly always the same and no convincing alternatives emerge. However, it can be harder to achieve specification with different sets of instruments.

4.2 Modelling the data from LCID and CLAN

Section 2.3 described the construction of this data set and noted that

Because of the lack of information from some Trusts the available data only relate to 374 of the 524 synthetic wards. All of these 374 wards account for at least one PoC7 client, but there is a relatively high percentage (11%) of wards with no more than 5 clients (Table 17), which could create instabilities in the modelling.

Given these concerns, our main purpose in modelling this material was to test whether it identified different needs drivers from those in the SOS CARE models. One reason for suspecting it might, was the relatively low correlation ($r=.497$) between the ward values of the dependent variables derived from the two data sets.

The first point of note when using the LCID/CLAN derived variable is that its distributional properties were far from normal and were barely improved by logarithmic transformation. A square root transformation produced a rather better distribution and this version was mainly used in the modelling.

As predicted, the Trust dummies explain a considerable proportion of the overall variance (32.7%). During the search for plausible models, specification was consistently poor and inspection of the residuals revealed several outliers. Three of these were removed from the analysis and the specification improved slightly, though the Reset test confirmed the strong misspecification.

The best predictor of these activity data was another Census derived health variable, the proportion of the adult population reporting “not good health” (ADNGH). Combined with the dummies this accounts for 59.9% of the overall variance. (Replacing this with the best variable in the SOS CARE modelling LLTU65 reduces this figure slightly to 57%) Stepwise and manual methods were used to test various combinations of variables, but no plausible combinations were able to increase the overall explanatory power beyond 67% and misspecification was a consistent problem.

One of these models is shown in Table 36. This adds children in working family tax credit households and the under-75 SMR to the Census poor health indicator. Its explanatory power is just under 65%, but it is very poorly specified, even after using a square root transformation on the dependent and excluding three outliers.

Table 36: Model based on square root transformation of LCID/CLAN derived variable

	Std coeff Beta	T	Signif T	R squared
ADNGH	0.355	5.472	0.000	
KWFTC	0.137	3.276	0.001	
U75SMR	0.256	4.793	0.000	
Reset Dummies	0.099	2.896	0.004	32.7
Model				64.9

Other functional forms and exclusions were explored. A single variable LOG-LOG model (Table 37) came closest to being specified (the significance of the reset test coefficient has now decreased to 3%). Adding further logged variables might improve the specification, but we decided to abandon this line of enquiry as the performance of these models were consistently inferior to those based on the SOS CARE data. The tests had, however, proved the point that despite the relatively low correlation between the SOS CARE and LCID/CLAN derived dependents, much the same variables, or very close correlates, were emerging as the best predictors.

Table 37: Model based on LOG transformation of LCID/CLAN derived variable

	Unstandardised coeff		Std coeff	T	Signif T	R squared
	B	S.e.	Beta			
(Constant)	-0.180	0.017		-10.549	0.000	
Reset	-0.347	0.160	-0.084	-2.167	0.031	
GNGHS	0.944	0.070	0.597	13.473	0.000	
Trust dummies						29.8
Model						54.5

GNGHS is a logged version of the proportion of people reporting "not good" health in the census.

5. Testing the models

Only the SOSCARE based models have been tested as we do not recommend the models from the LCID/CLAN dataset.

5.1 Sensitivity tests

The first test examines the robustness (as a predictor of activity) of the two variables that are common to both the recommended models: (the log of) age standardised limiting long-term illness for people aged 65 and under, and (the log of) adults in disability living allowance claimant households. Correlations between these variables and (the log of) standardised activity were computed separately for the wards in each of the eleven Trusts (See Table 38). The correlations are always highly significant and for the former variable are between .61 and .89, and for the latter between 0.60 and 0.87. They make a strong case for inclusion of both variables in any model.

Table 38: Single Trust correlations - between standardised activity and LLTI amongst 65s and under, and adults in disability living allowance households (all variables are logged).

	GLLTU65	GADDLA
Armagh&Dungannon	0.773	0.772
Causeway	0.648	0.720
Craigavon&Banbridge	0.822	0.853
Down&Lisburn	0.868	0.860
Foyle	0.590	0.718
Homefirst	0.873	0.823
N&W Belfast	0.885	0.868
NDown&Ards	0.857	0.828
Newry&Mourne	0.737	0.672
S&E Belfast	0.887	0.867
Sperrin Lakeland	0.615	0.605

The peer reviewer notes that the overall model obtained represents a pooling of all within Trust models and “for completeness” computes the individual within Trust model coefficients – i.e. runs the models on data from separate Trusts. The results are shown in Table 39. The Trust most at variance with the overall model is number 5. This is also the HSST contributing most to Trust dummies. However removal of data

for this Trust from the overall model has negligible effect on the final solution or on HSST allocations.

Table 39: Coefficients in models run on data from single Trusts

Trust	Adjusted R ²	Coefficient (SE) of GADDLA	Coefficient (SE) of GLLTU65
1	59.6	0.374 (0.272)	0.614 (0.436)
2	48.6	0.616 (0.255)	0.174 (0.493)
3	72.2	0.700 (0.247)	0.366 (0.245)
4	75.5	0.340 (0.221)	0.811 (0.357)
5	49.7	0.789 (0.194)	-0.132 (0.288)
6	77.1	0.261 (0.108)	1.071 (0.164)
7	77.3	0.133 (0.293)	1.117 (0.465)
8	75.9	0.369 (0.146)	0.941 (0.239)
9	51.3	0.181 (0.313)	1.014 (0.437)
10	78.8	0.254 (0.170)	0.901 (0.258)
11	37.2	0.303 (0.239)	0.651 (0.364)

The remaining sensitivity tests are less rigorous and mostly concentrate on the effects on the coefficients in the recommended model of removing the data for each of the Boards in turn. When this test is applied, the coefficients of all the needs drivers in the model fluctuate, but remain significant at better than 0.01%; with the exception of the coefficient of GINSC which ceases to be significant when Eastern Board data are removed.

**Table 40: Sensitivity tests– removing one Board at a time
(Three variable model)**

Table shows standardised coefficients

	Including all Boards	Excluding Board			
		1 (Eastern)	2 (Northern)	3 (Southern)	4 (Western)
GLLTU65	.356	.326	.301	.342	.440
GADDLA	.388	.386	.402	.405	.349
GINCSC	.141	.056 (NS)	.168	.168	.134

All coefficients (except where marked) are significant at least .01%

A further set of tests re-ran the three variable model with a modified set of Trust dummies. The initial set of dummies was based on the 11 Trusts. The second set increased this number to 14 by dividing Belfast into four sectors. The effect on explanatory power was negligible. For the three-variable log-log model the explanatory power of the dummies increased very slightly from 29.6% to 29.9% and the value for the overall model increased slightly from 79.4% to 80.3%. However, the new set of dummies does produce some changes in the variable coefficients: notably a reduction in the significance of the logged income score (see Table 41).

Table 41: Coefficients of three variable LOG-LIN model with standard and enhanced set of dummies

	Unstandardised coeff		Std coeff Beta	T	Signif T
	B	S.e.			
11 dums					
(Constant)	0.198	0.082		2.407	0.016
GLLTU65	0.559	0.112	0.356	5.010	0.000
GADDLA	0.376	0.057	0.388	6.596	0.000
GINCSC	0.114	0.040	0.141	2.831	0.005
14 dums					
(Constant)	0.294	0.084		3.492	0.001
GLLTU65	0.579	0.110	0.368	5.283	0.000
GADDLA	0.419	0.057	0.433	7.410	0.000
GINCSC	0.078	0.041	0.097	1.905	0.057

To sum-up, two variables in the model, the standardised limiting-long-term illness rates for under 65s and the proportion of adults in disability living allowance households were particularly robust, to the point of surviving tests on ward level data from individual Trusts.

5.2 Unmet need tests and adjustments

To test for unmet need effects, the project has applied the methods recommended in the Sutton and Lock paper “Inequalities in Health and Social Care Use: the Implications for Resource Allocation in the HPSS”. The same methods are employed in the Allocation of Resources to English Areas Report.

These works describes two approaches to testing and potentially correcting for unmet need. The first assumes that activity levels for a wide range of services will be strongly correlated with deprivation and that evidence of non-linearity in the relation between activity and deprivation, especially at high levels of deprivation, may be indicative of unmet need. This is referred to as "the shortfall approach".

The second is concerned with variations in the relation between activity and deprivation within each of the supplying authorities - "the variations approach". In resource allocation based on small area modelling it is customary to base final models on the average relation across all authorities. The Sutton and Lock paper argues that it may be preferable to use an average of the relations in a sub-set of authorities whose activity varies most in relation to need.

Both approaches are applied in the AREA report but for the present exercise we have only been asked to investigate the shortfall approach. In the Northern Irish context, this involves the following major steps (as set-out in Sutton and Lock)

- Construct a best-fit and parsimonious model of the relation between service activity and needs drivers and any measure of supply.
- Use a spline regression to test for evidence of non-linearity in the relation between the Noble deprivation score and residuals from the model. Significant values of the coefficient of the spline variable may be a clue to unmet need.

- Regardless of the result of the spline test, force a suite of variables into the model. The recommended suite is a measure of limiting long-term illness, the Noble score, and several health variables that were synthetically estimated from the Northern Ireland Health and Social Well-being Survey.
- Discard any non-significant variables.
- Re-compute the model retaining the original variables and any of the unmet need test variables that are significant – paying particular attention to significant negative associations with deprivation scores.
- Use the revised coefficients of the original variables to compute the allocations.

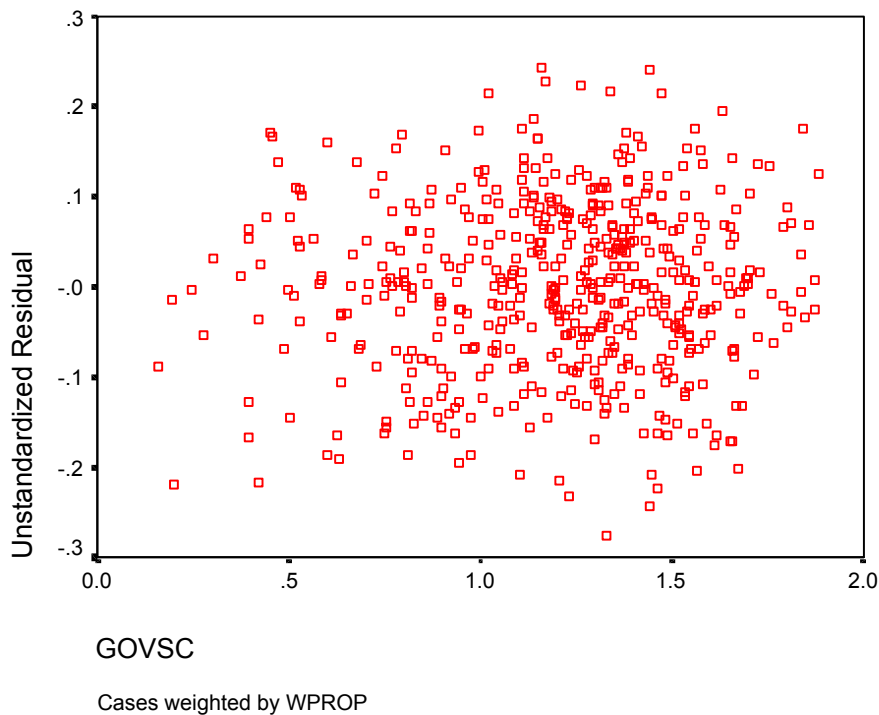
At the outset it has to be noted that the application of spline and other regression approaches to unmet need when models are derived via 2SLS is relatively unexplored territory; and is not described in either the Sutton and Lock paper nor the AREA report. Indeed the latter work subsumes all supply factors into a set of area dummies, possibly to avoid the need to use 2SLS.

In the following discussion we test both whether there is a case for applying corrections to the OLS version used for predicting allocations and whether the recommended "additional needs" variables would be retained in the 2SLS stages.

Spline test using overall Noble index

The purpose of a spline test using a measure of deprivation such as the Noble index is to test whether there are obvious non-linearities in the relation between deprivation and that part of service activity that is unexplained by the recommended models. The interpretation of a spline test is somewhat changed when dealing with log transformed variables, when it means we are looking for non-linearities in an otherwise linear relation between two log-transformed variables.

Figure 2 Residual from recommended two-variable model plotted against log of Noble overall score



As a preliminary to searching for possible spline points, we look at the relation of the log of the overall Noble score to the residuals from the basic two-variable model. A scattergram of this relation is shown in Figure 2. It demonstrates that there is so little evidence of a basic relation between the Noble score and the residuals from the model ($r=+.025$) that there is little purpose in looking for a spline point. As with the modelling of services for learning disabilities, there are stronger determinants of the phenomenon in question than the overall Noble score.

Although Figure 2 suggests we are unlikely to find unmet need effects in relation to the Noble measure of deprivation, we have followed the recommended methodology and checked the effects of adding the suite of variables recommended by Sutton and Lock, including the overall Noble score, to the basic models.

If the Noble score is forced into the equation used to predict allocations (including the Trust dummies but not the supply variable) it is not significant (see Table 42). The Noble score becomes significant if the supply variable is introduced, but only due to its interaction with the supply variable - and it has a positive sign, which is inconsistent with its functioning as a correction for unmet need. Moreover, if the overall Noble score is included in the list of variables in the 2SLS analysis, it is almost always not significant - indicating that it should not be retained for the final OLS allocation modelling.

Table 42: Effect of adding overall Noble score to basic OLS model used for predictions.

Variable	Unstand coeff	T-value	Signif of t
Without Noble score			
GLLTU65	.750	8.37	.000
GADDLA	.378	6.58	.000
With Noble score			
GLLTU65	.676	6.40	.000
GADDLA	.372	6.48	.000
Overall Noble score	0.037	1.32	.187

Much the same applies to the three individual morbidity variables that are recommended in the Sutton and Lock work. When a stepwise procedure is used to add them (in a logged form) to the OLS reworking of the basic model none of the three new variables is retained. None is significant at 5%, though the mental health factor is nearly significant (Table 43). It is unsurprising that the mental health variable has the greatest significance amongst these three as it has most substantive relevance to the social services activity being modelled.

Table 43: Effect of adding ward level values of individual morbidity variables to the basic OLS model to be used for the allocations

Variable	Unstand coeff	T-value	Signif of t
Without Noble score			
GLLTU65	.750	8.37	.000
GADDLA	.378	6.58	.000
log(genhealth factor)	0.049	1.38	0.168
log(circ disorder factor)	-0.016	-0.54	0.587
log(mental health factor)	0.052	1.83	0.067

Again, we find that if these variables had been entered into the 2SLS stages (Table 44), they would have been ruled out for the later OLS work by virtue of being insignificant.

Table 44: Effect of adding ward level values of individual morbidity variables in the 2SLS stages of the modelling

Variable	Unstand coeff	T-value	Signif of t
GLLTU65	0.675	2.29	.0041
GADDLA	0.396	3.80	.0002
log(genhealth factor)	-0.064	-0.063	0.950
log(circ disorder factor)	-0.143	-0.274	0.784
log(mental health factor)	0.469	0.48	0.630

From the results of all these runs, we cannot find evidence of unmet need related to deprivation that requires any correction to the original coefficients, nor of obvious non-linearities in the model itself, which might be corrected by a spline adjustment. We were also unable to find any effects relating to the individual morbidity factors and note these would have been rejected if they were included in the 2SLS phases of the work. Therefore we do not recommend making any corrections to the model.

6. Conclusions and recommendations

Much of the effort in this project has been directed at translating the many and diverse sets of activity data. In the event, only one has been sufficiently comprehensive or robust to support modelling and, even here, it has not been possible to include cost weights in the modelling, which has been based on client numbers. In order to control for age-sex effects and improve the distributional properties of the dependent variable, we have used the log of the age-sex standardised numbers of clients per ward in all the final stages of the modelling.

Although we have not been able to develop a costed dependent, the resulting models are well specified and parsimonious – and include a plausible set of variables.

We were provided with several possible supply variables prior to a final version based on the location of social work facilities for PoC7 clients. This proved to be endogenous when tested with likely models and we therefore adopted a 2SLS process to select the variables for inclusion in the final model. This process identified three variables for the final model. Of these, the top two in Table 45 (GLLTU65 and GADDLA) are by far the stronger predictors of activity and are the more robust when the model is re-run with data from subsets of Boards and Trusts.

Table 45: Variables in the Final Model

	Short definition - all are logged
GLLTU65	Limiting long-term illness amongst under 65s
GADDLA	Adults on disability allowance
GINSC	Noble income score

The final stage in the two stage process is to recompute the models with OLS, without the supply variable. This will provide the coefficients for the allocation formula. The results of this stage are shown in Table 46. We are torn between recommending the two or three variable models in the table. Their explanatory power is very similar. Our calculations have found that both models are well specified (when tested in a 2SLS context), though the peer reviewer had some doubts on the specification of the two-variable model and we certainly found that the test statistic is somewhat stronger

for the three variable version. The three variable version provides a wider range of themes in the need drivers, but the third variable, the income score, is both less robust than the other two and has had to be computed by translating data from an old ward base. We would be happy to see either model adopted. and leave the final decision to CFRG. Our own calculations have shown that the two and three variable models have very similar allocational implications.

Table 46 Coefficients when variables selected by 2SLS are re-used in OLS log-log modelling.

Variable	Dummies only	Coeff in 3-var model (and signif of t-value)	Coeff in 2-var model (and signif of t-value)
GLLTU65		0.559 (.000)	0.750 (.000)
GADDLA		0.376 (.000)	0.377 (.000)
GINCSC		0.114 (.004)	
(Constant)		0.198	0.357
Adjusted R-squared (%)	29.6	79.1	79.0

Given the strong correlations between the various needs drivers, it is very likely that other combinations of variables could perform as well as the models we are suggesting. However when we have substituted alternative variables covering similar topics and forced in variables when the range of drivers appears to limited, the suggested models proved surprisingly resilient to these tactics. In particular, the GLLTU65 and GADDLA variables emerge strongly in almost all the modelling.

The models are robust to exclusions of entire Boards' data and the correlations between GLLTU65, GADDLA and activity are remarkably consistent across most of the Trusts.

On these bases we are happy to recommend either model, though there is a good argument for using a three variable version, as the explanatory power is improved by adding the income score variable and it broadens the responsiveness of the model to income deprivation effects.

We cannot find evidence of unmet need related to deprivation that requires any correction to the original coefficients, nor of obvious non-linearities in the model itself, which might be corrected by a spline adjustment.

We would have liked to have included a greater range of data sources on activity in the modelling, but the weakness of the individual data sets and the possibility of double counting were two major obstacles. We have devoted considerable time and effort to attempt to specify and obtain data on PoC7 clients in specialist hospital units and other patients whose needs are similar, but who are in more general hospital wards. We have reluctantly closed this line of enquiry having been unable to arrive at a suitable specification. Consequently we have excluded all the data from hospital sources, although some hospital patients will be included if they are recorded on SOS CARE. We recognise that this is a high cost group, but their numbers are so small that their omission should have no impact on the existing modelling which is based on client numbers not costs.

Because the existing allocations are computed from SMRs, we took particular care to compare the predictive ability of SMRs with the Census derived self-report health variables that emerged so strongly in the modelling. We conclude that the census variables are much better predictors of service activity than SMRs and therefore have no hesitation recommending these new models over the existing formula.

Glossary

Endogeneity (of supply side variables)

(Notes on endogeneity are adapted from Carr-Hill et al (Sept 1994) – Modelling NHS Inpatient Utilisation)

In developing a resource allocation formula, we wish to correct for variations in supply between areas. Effectively this means assuming that all supply in an area is at some national average appropriate to the level of needs found in that area.

Measures of supply (“supply variables”) provide a means of testing this assumption. Supply can vary for many reasons in addition to those identified in the modelling as representing legitimate need. The analytic task is to find that part of the supply effect which is attributable to factors unrelated to the needs indicators in the model and to remove that part of the supply effect from the model.

The test of the assumption involves regressing the supply variables on the needs drivers, then including the residuals from these regressions alongside the needs drivers and supply variables as independents in a regression with the utilisation variable as dependent. The latter is sometimes known as the unrestricted question and its residual sum of squares is compared with that from a restricted equation – where the utilisation variable is simply regressed on the needs drivers and supply variables.

If the difference between the explanatory power of the two equations is significant then endogeneity has been demonstrated and two stage least squares methodology should be used.

The main features of this two stage process are firstly that the normal OLS procedure for identifying the best predictor variables is replaced by a process in which the supply variable is, in effect, replaced by its residual after it has

been regressed on a set of candidate variables or “instruments”. This process is used to search for the best predictor set of need drivers, which are then put into a simple OLS model (the second stage) in order to compute the coefficients for the allocation formula.

Forcing variables (into the model)

Multiple regression procedures in statistical packages use a variety of criteria to decide which variables should be included in an equation. The most commonly reported is the t value of the variable coefficient, but the tolerance (the extent of correlation with variables already in the equation) is usually considered when admitting or rejecting variables. Forcing a variable into an equation means suppressing these criteria and including a variable regardless.

FR22

A financial return from Trusts which gives details of provider expenditure by Programme of Care. Expenditure is split into three categories: Personal Social Services, Community, Hospital.

Functional form

The types of multivariate regression used in this modelling assume that both dependent and independent variables have approximately normal distributions and that relations between these variables will be linear. Where one or both of these conditions are not met, it is customary to apply algebraic translations to the variables to either improve their approximation to normality, or to improve the linearity of the overall relationship.

Two descriptive statistics, the kurtosis and skewness of variables, will give an indication of the approximation to normality. High values of skewness indicate

variables that have a long tail either side and kurtosis identifies variables that have distributions that are too flat or too steep.

Non-linearity is often indicated by low explanatory power, or poor specification. Transformations that are used to improve approximations to normality and linearity include converting variables to an exponential, logarithmic or a square root form.

It may be appropriate to apply these transformations to either or both the independent and dependent variables depending on the cause of the misspecification.

Specification (of a model)

Careful checks were made to ensure that the statistical models were well specified – i.e. that no systematic effects have been overlooked.

Specification errors can occur for various reasons:

- omission of a relevant explanatory variable
- inclusion of an irrelevant explanatory variable
- incorrect mathematical form of the regression equation
- incorrect specification of the covariance structure such that the error term is not normally distributed.

The test for misspecification employed here (the “Reset” test) involves testing the significance of the t value of the coefficient of the squared predicted values from the basic model when added to the model as an independent variable. In effect this is testing to see whether there are any systematic effects remaining in the residuals.

The Reset test should not be applied to 2SLS modelling. The preferred procedure involves computing an auxiliary equation derived from an OLS regression of the residuals from the 2SLS model against the full instrument and variable set in the 2SLS. The test statistic is the ratio of the sum of squares explained by this auxiliary equation to the estimated error variance for the 2SLS estimated utilisation equation. It should have a chi-squared distribution with degrees of freedom equal to the number of instruments in the auxiliary equation minus the number of regressors in the utilisation equality.

Standardisation

Where a phenomenon has been shown to be strongly related to basic demography, such as population age and sex, it will be inappropriate to make comparisons between populations without controlling for this relation.

When very different rates of physical and sensory disability are found in different age and sex groups in the population, when we want to construct a ward level measure of the numbers of people with physical and sensory disability, we have to take account of the age-sex composition of the ward population. There are basically two ways in which this can be done.

Direct standardisation.

Here the physical and sensory disability rates are computed for each age-sex group for each ward, then these rates are applied to the national average ward composition giving a corrected number of people with physical and sensory disability. The advantage of this approach is that the resulting metric can be numbers of people or services costs, rather than a dimensionless ratio; the disadvantage is that it tends to be unstable for small wards and for phenomena with very small numbers in some age-sex groups.

Indirect standardisation

The alternative (as used in this project) is to compute the national average RATES for each age sex group, then apply these to the ward population characteristics, giving an expected value for each ward, assuming the national rates universally applied. A ratio can then be constructed of the expected to the actual numbers per ward. This method has the advantage of being far more robust than direct standardisation and tends to produce more powerful models. The problem is that national rates have to be used to convert the ratio back to costs or activity in order to use the resulting equations for resource allocation.

Spline regressions (dealing with partial non-linearity)

General departures from linearity are usually handled by transforming the dependent or independent variables. Departures from linearity over part of the range of an otherwise linear relation can be handled by spline coefficients. The following example describes the application of spline coefficients when we believe that there may be some non-linearity between prevalence of physical and sensory disability and the proportion of people living in Disability Living Allowance Households.

The base model includes the proportion of people living in Disability Living Allowance Households and any other significant variables that have been selected. To this we add a further independent variable that is zero in all but the 10% of wards with the highest values for the proportion of people in DLA households (this is the 90% spline point). In the top 10% of wards the variable is given the value of the proportion in DLA households minus the proportion at the spline. If this new variable is significant we could interpret this as evidence of a change in slope at this point, which in certain circumstances might be interpreted as unmet need. In testing the models, we have re-run the regressions with spline points at the top 5%, 10%, 20% 40% 60% 80% values of the selected variables. Where the spline variable is significant and has an

intuitively correct sign, it can be argued that the coefficients of the non-spline variables in the model should be those that were computed with the spline variables added, although the spline variables themselves should not be retained in the allocation formula.

Synthetic wards

As previously noted, a key requirement for the modelling is to choose areas that are sufficiently large to have robust values for the activity data and socio-economic indicators but not so large that they each contain a variety of conditions and thereby obscure any relations between activity and conditions.

Electoral wards are the obvious choice for this work as Census and administrative statistics are available at this level, but rural wards may have populations that are too small to generate robust values for indicators of conditions and service activity. We considered using the next larger group of administrative units, District Electoral Areas, but tests with both Census values and deprivation indicators have shown them to be socially heterogeneous and unlikely to produce conclusive models. Hence we felt we had no alternative but to construct a new set of “synthetic wards”. In devising these wards, the aim was to produce areas with populations that are always greater than 2000 persons.

Trust dummies

Whenever activity data are used as dependent variables there is the risk that systematic supply influences may distort the modelling. Some of these phenomena are addressed with variables relating to the supply of specific services, but there may be more general or unmeasurable effects that we suspect are due to differences in policy and performance between agencies responsible for services. Although the current modelling is carried out at ward

level, we suspect there may be influences due to the behaviours and policy of Trusts.

Therefore we include dummy variables that take values of one for each observation in the Trust to which they relate and values of zero for all other cases. Such a dummy variable is created for all but one of the Trusts (if all 11 Trusts have a dummy there are no degrees of freedom left in the modelling). It is customary to not use a dummy for the Trust with the largest number of cases.

ANNEX 1 Needs Drivers in Combined Set

PoC7 and LD Modelling		Census source
Income		
Persons aged 18-64 in income support households	ADIS18_64	Claimant counts - for most of these indicators have proportion of persons in relevant benefit households as well as numbers of claimants
Persons aged 18-64 in income based JSA households	ADJSA	
Persons aged 18-64 in family credit households	ADWFTC	
Persons aged 18-64 in disability working allowance households	ADDLA	
Persons aged 18-64 in lone parent claimant households	ADLPP	
Persons aged 18-64 in JSA claimant households	ADJSA	
Unemployment (from census)		
Other Economic		
Households where head is in a low SEG	LOWSEG	S344
Households where head is in a high SEG	HIGHSEG	S344
Households with no car	NOCAR	KS17
Households with 2 or more cars	TWOCAR	KS17
Households with 3 or more cars	THREECAR	KS17
Social environment/facilities		
Index of MD Social environment score	SOCENVSC	
Index of MD Access score	ACCSSC	
Household Structure		
Households with 3+ dependent children	HH3K	S007
Persons in single carer households	SNGCARER	S027
Persons in no carer households	NOCARER	S027
Persons aged 18-64 who are single, widowed or divorced	SWIDRCD	KS04
Single- never married	SINGLE	KS04
Married or living in relationship	MARRIED	KS04
Ex married	EXMAR	KS04
Household Tenure and facilities		
Persons aged 18-64 in social rented housing	SOCREN	KS18
Persons aged 18-64 in owner occupied housing	OWNOCC	KS18
Persons aged 18-64 in private rented housing	PRIVREN	KS18
Persons aged 18-64 in private rented housing assoc with employment	PRIVREM	KS18
Persons aged 18-64 in households without basic amenities	NOBASIC	KS19
Persons aged 18-64 in households without central heating	NOCENTH	KS19
Illness		
SMR under 65	U65SMR	
All age SMR	ALLSMR	
SMR under 75	U75SMR	
Persons with "not good health"	ADNGH	S016

Persons with "not good health" (standardised)	NGHS	
Persons aged 18-64 with a limiting long term illness	ADLLTI	S016
Persons aged 18-64 with a limiting long term illness (standardised)	LLTI	S016
Index of MD Health Score	HLTHSC	
Education		
Working age adults with no qualifications	NOQUAL	KS13
Index of MD Education Score	EDUCSC	
Ethnicity		
Persons in non-white ethnic groups	NONWHT	KS06
Multiple Deprivation		
All components and overall score for Noble indicator	OVSC	
Religion		
Persons of Catholic origin	CATHOL	KS07B
Persons of Protestant/Christian origin	PROTST	KS07B

Children and Family PoC		Census or other source
Income		
Unemployed JSA claimants	ADUNEMP KJAS	KS09
Noble child poverty score	CHILDPSC	
Other Economic		
Dependent children in households where head is in a low SEG	KLOWSEGA,B,,	S345
Dependent children in households where head is in a high SEG	KHGHSEGA,B	S345
Children in no car households	KNOCARA,B	S062
Children in households with 2 or more cars	KTWOCARA,B	S062
Household Structure		
Lone parents aged under 24	LONPU24	KS20.
Children in lone parent premium households	KLPP	
Households with 3+ dependent children	HH3K	S007
Children in overcrowded households	KOWCRAA,B	S357
Proportion of households with children 0-15 and lone parents of all ages	LPDEPK	S007
Households with dep (ALL) children	HHDEPK & HHALK	
Household Tenure and facilities		
Children in social rented housing	KSOCRENA,B	S359
Children in owner occupied housing	KOWNOCCA,B	S359
Children in private rented housing	KPRRENA,B	S359
Children in households without basic amenities	KNBASIC	S359
Children in households without central heating	KNOCENTH	S359
Illness		
SMR under 65	U65SMR	
Persons aged 18 and under with "not good health"	KNGH	S016

Persons aged 18 and under with a limiting long term illness	KLLTI	S016
Children in DLA claimant households	KDLA	
Education		
Working age adults with no qualifications	NOQUAL	KS13
Persons aged 16-18 not in full-time education		S028
Ethnicity		
Proportion in non-white ethnic groups	NONWHT	KS06
Religion		
Children of Catholic origin	KCATHO	KS07A
Children of Protestant/Christian origin	KPROT	KS07A
Claimants		
Children in JSA households	KJAS	Claimant count data
Children in WFTC households	KWTFC	Claimant count data
Children in Lone Parent Supplement HHolds	KLPP	Claimant count data
Children in income support households	KIS	Claimant count data
Multiple Deprivation		
All components and overall score for Noble indicator	OVSC	

Descriptive statistics for variables in needs driver set.

	Minimum	Maximum	Mean	Std. Deviation
ADLLTI	0.078	0.381	0.186	0.052
ADNGH	0.033	0.283	0.110	0.042
SNGCARER	0.011	0.085	0.036	0.010
NOCARER	0.117	0.493	0.281	0.062
OWNOCC	0.203	0.983	0.756	0.157
PRIVREN	0.007	0.628	0.069	0.052
SOCREN	0.000	0.774	0.168	0.151
PRIVREM	0.000	0.366	0.006	0.024
HIGHSEG	0.070	0.721	0.284	0.121
LOWSEG	0.028	0.568	0.242	0.102
NOCENTH	0.003	0.172	0.052	0.029
NOBASIC	0.000	0.038	0.008	0.006
NOQUAL	0.080	0.745	0.426	0.098
PROPNST	0.030	0.583	0.255	0.097
CATHOL	0.009	0.990	0.433	0.325
PROTST	0.007	0.969	0.538	0.311
NONWHT	0.000	0.056	0.008	0.007
SWIDRCD	0.297	0.901	0.479	0.090
NOCAR	0.033	0.748	0.240	0.147
TWOCAR	0.023	0.682	0.314	0.144
THREECAR	0.000	0.201	0.064	0.040
SINGLE	0.177	0.742	0.293	0.063
MARRIED	0.174	0.750	0.563	0.085
EXMAR	0.207	0.797	0.363	0.078
ADUNEMP	0.010	0.125	0.041	0.021
UPTO15S	249.000	2637.000	760.310	310.730
UPTO17S	293.000	3002.000	862.386	349.214
AD18TO64	1144.000	8080.000	1928.449	715.735
ALLAD	1359.000	8538.000	2354.591	834.504
ADOV65	122.000	1390.000	426.141	192.515
NHHOLD	571.000	3669.000	1196.696	459.679
KLOWSEGA	0.008	0.378	0.129	0.065
KLOWSEGB	0.009	0.428	0.148	0.072
KHGHSEGA	0.025	0.373	0.148	0.065
KHGHSEGB	0.033	0.428	0.171	0.075
KNOCARA	0.000	0.425	0.076	0.071
KNOCARB	0.000	0.485	0.094	0.085
KNOCARC	0.000	0.590	0.119	0.112
KTWOCARA	0.010	0.445	0.229	0.109
KTWOCARB	0.013	0.541	0.284	0.131
KTWOCARC	0.023	0.595	0.335	0.142
KOWCRAA	0.003	0.241	0.052	0.035
KOWCRAB	0.007	0.344	0.100	0.057
KPRVRENA	0.000	0.291	0.035	0.024
KPRVRENB	0.000	0.497	0.071	0.049
KSOCRENA	0.000	0.635	0.085	0.084
KSOCRENB	0.000	0.792	0.170	0.157

KOWNOCCA	0.063	0.610	0.388	0.113
KOWNOCCB	0.199	0.995	0.759	0.162
HH3K	0.035	0.307	0.131	0.048
LONPU24	0.000	0.089	0.019	0.016
LPDEPK	0.012	0.522	0.127	0.082
KNBASIC	0.000	0.018	0.003	0.003
KNOCENTH	0.000	0.077	0.018	0.013
KLLTI	0.011	0.124	0.053	0.019
KNGH	0.000	0.045	0.013	0.009
HHDEPK	0.094	0.699	0.374	0.080
HHALLK	0.131	0.802	0.508	0.093
KCATHOL	0.000	0.974	0.423	0.319
KPROT	0.000	0.863	0.413	0.263
POPDENS	0.107	113.828	14.675	18.905
KIS	0.011	0.814	0.216	0.160
KWFTC	0.028	0.433	0.236	0.072
KJAS	0.000	0.117	0.017	0.014
KLPP	0.001	0.607	0.063	0.080
ADLPP	0.000	0.090	0.011	0.013
ADJSA	0.003	0.130	0.033	0.023
ADIS	0.018	0.566	0.167	0.096
ADIS1864	0.007	0.534	0.134	0.092
ADIS65OV	0.045	0.828	0.328	0.147
ADDLA	0.016	0.312	0.102	0.049
ADWFTC	0.003	0.145	0.071	0.026
KDLA	0.004	0.079	0.031	0.012
ALLSMR	0.317	2.281	1.000	0.245
U75SMR	0.376	2.404	0.986	0.294
U65SMR	0.310	3.203	0.992	0.376
OVSC	1.132	76.046	20.747	14.626
INCSC	3.260	68.701	26.815	13.521
EMPSC	2.843	29.912	12.112	4.218
HLTHSC	-2.197	2.267	-0.038	0.741
EDUCSC	-2.589	2.475	-0.071	0.877
ACCSSC	-1.995	2.477	-0.012	0.787
SOCENVSC	-2.385	1.899	-0.044	0.744
HOUSSC	0.131	0.441	0.260	0.047
CHILDPSC	2.381	91.584	37.602	19.207
NWRDS	1.000	3.000	1.111	0.354
LLTI	0.470	1.822	1.000	0.237
NGHS	0.356	2.248	0.988	0.346

ANNEX 2 Trust Data Request Letter and Specification

Capitation Research on Physical & Sensory Disability Research

The Department of Health, Social Services & Public Safety (DHSSPS) have commissioned MSA Ferndale to undertake research to refine the mechanism for annual revenue allocations to HSS Boards for the Physical & Sensory Disability PoC. This research was described in recent correspondence between Mr Andrew Hamilton, Director of Finance, DHSSPS and your Trust Chief Executive.

A Steering Group comprising representatives of DHSSPS and the four HSS Boards is overseeing the research.

To inform the work planned, the research team require access to a range of data on services provided to patients/clients for reasons of physical and sensory disability. This letter concerns requirements in relation to community health data only – as set out in the specification appended to this letter. The request related to service activity coded as PoC 7 (Physical & Sensory Disability PoC).

The study requires that services provided by the Trust can be analysed by the age and gender of users, and by electoral ward. For this reason, information to establish such is requested. For example, postcodes are requested to allow the research team to attribute each record to the service user's electoral ward of residence – but this data item can be omitted where the Trust is in a position to provide electoral ward directly on the files forwarded (both 1984 and 1992 ward reference codes are required).

Data is required at the level of the individual service user, but no information is requested that could be used to establish the identity of such users.

I would be grateful if you could:

- ◆ examine the attached community health services data specification;
- ◆ consider what could be provided from Trust IT systems, mindful that the timescale of the study will necessitate data being received by our team by w/e 4th July; and
- ◆ advise me by e-mail of your Trust's ability to provide such information, or detailing any other issues you may wish us and the Department to consider. I would be grateful for a reply by w/c 24th June.

If you wish to discuss the above directly, please feel free to contact:

Commissioning body	Lead researcher
Philip Spotswood Project Support Analysis Branch DHSSPS Annexe 2 Castle Buildings Stormont BT4 3UD Tel: 028 90 522197 Philip.spotswood@dhsspsni.gov.uk	Martin Spollen Director MSA Ferndale 393 Shore Rd Whiteabbey BT37 9SB Tel: 028 90 369366 Martin.spollen@msa-ferndale.com

We are grateful for any assistance you may provide, and I look forward to hearing from you in the near future.

Yours sincerely,

Martin Spollen
Director

PoC 7 Research – Information specification

Group 1	Community Nursing
Required	A download of 12 months worth of contact information (see fields below) for period to 31 December 2002. The following fields to comprise a single record for each patient in contact during the period.
Field #1	Patient's age @ 30 June 2002
Field #2	gender
Field #3	postcode of residence
Field #4	electoral ward of residence (both 1984 & 1992 wards)
Field #5	No of face to face <i>community nursing</i> contacts recorded in the period under Physical & Sensory Disability Programme

Group 2	Physiotherapy
Required	A download of 12 months worth of contact information (see fields below) for period to 31 December 2002. The following fields to comprise a single record for each patient in contact during the period.
Field #1	Patient's age @ 30 June 2002
Field #2	Gender
Field #3	postcode of residence
Field #4	electoral ward of residence (both 1984 & 1992 wards)
Field #5	No of face to face contacts with a community <i>physiotherapist</i> recorded in the period under Physical & Sensory Disability Programme

Group 3	Occupational Therapy
Required	A download of 12 months worth of contact information (see fields below) for period to 31 December 2002. The following fields to comprise a single record for each patient in contact during the period.
Field #1	Patient's age @ 30 June 2002
Field #2	Gender
Field #3	postcode of residence
Field #4	electoral ward of residence (both 1984 & 1992 wards)
Field #5	No of face to face contacts with a community <i>OT</i> recorded in the period under Physical & Sensory Disability Programme

Group 4	Speech & Language (inc. Audiology)
Required	A download of 12 months worth of contact information (see fields below) for period to 31 December 2002. The following fields to comprise a single record for each patient in contact during the period.
Field #1	Patient's age @ 30 June 2002
Field #2	Gender
Field #3	postcode of residence
Field #4	electoral ward of residence (both 1984 & 1992 wards)
Field #5	No of face to face contacts with a community <i>speech & language therapist</i> and/or <i>audiology services</i> recorded in the period under Physical & Sensory Disability Programme

Group 5	Aids & Adaptations
Required	A download of 12 months worth of information (see fields below) for period to 31 December 2002. The following fields to comprise a single record for each patient in contact during the period.
Field #1	Patient's age @ 30 June 2002
Field #2	Gender
Field #3	postcode of residence
Field #4	electoral ward of residence (both 1984 & 1992 wards)
Field #5	Number of aids and appliances provided to this service user
Field #6	Expenditure on aids and appliances for this service user.

NOTES:

- ◆ **Full postcode is required to allow the Department to append SAS electoral ward codes to the files, after which postcode information will be deleted from the files. However, postcode is not required IF the Trust can provide both the 1984 and 1992 SAS ward codes directly.**
- ◆ **ALL FILES TO BE PROVIDED IN MICROSOFT EXCEL FORMAT.**
- ◆ **EACH GROUP TO BE PROVIDED IN A SEPARATE & LABELLED FILE.**