Additional Needs Analysis for the Family and Child Care Programme (POC3)

Final Report

Paul Dixon

(MSA/Ferndale – Secta, University of York)

Roy Carr-Hill

(University of York)

Martin Spollen

(MSA/Ferndale – Secta)

September 2004

Acknowledgements

We are most grateful to all those throughout the Heath 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.

Arriving at an appropriate set of cost estimates was a crucial part of the work. Many Trust and Board finance staff provided help with this, but we would particularly like to thank Paul Ballentine, Cletus McCloskey, Clifford Coulter, Carol Anderson and Christine Fraser for their assistance with both methods and figures.

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

Table of Contents Page

EXECUTIVE SUMMARY

EX.1 Introduction

This project has the aim of updating the existing formula for allocating resources to the Family and Childcare Care Programme (POC3). In many respects it replicates the methods of the study that produced the existing formula. That is, it uses multivariate regression to model POC3 activity against potential need drivers and supply variables.

The modelling is conducted with ward rather than local authority level data. Models built from local authority level activity data have often been criticised for their tendencies to be biased by the peculiarities of local authority behaviour and for reproducing historic inequities in funding. By working at a ward level, the modelling can control for such influences.

In more detail, the main activities of the study were as follows:

- updating activity data and finance data;
- construction of an updated cost dependent variable using the above data and the workload estimates from the work that generated the current model;
- development of access related supply variables;
- construction of new needs variables using the 2001 Census and updated administrative sources;
- modelling of the activity data, potential need drivers and supply variables.
- assessing whether there is any evidence of unmet need, manifest as non-linearities in the fit of the model to the data
- applying any unmet need corrections as appropriate
- new dependent variable based on variations in the workload estimates;

EX.2 Activity data for the modelling

The activity data for the study was obtained entirely from SOSCARE. On the advice of the project commissioners, six files were requested from each of the 11 Trusts. These consisted of three snap-shots of activity on two dates: 31/2/02 and 30/9/02. The three snapshots relate to: all clients with open cases; all pending clients; and all FRC clients. Extracts covering open and pending cases were received for all Trusts. The information on FRC users was limited to 3 or 4 Trusts and was too incomplete to include in the modelling.

The minimum data requirement, in order to replicate the analysis that generated the current formula, was to have sufficient information on the type of care to be able to assign a client to one of five groups. These groups were chosen both for the earlier Northern Ireland study and the project that produced the model that is currently used in England. They met the two requirements of being identifiable from basic fields on client record systems and being costable from budgetlines on annual finance returns, albeit with the help of some additional data to apportion social work costs.

The five client groups used in the analyses are as follows. The percentages of clients in these groups are shown in Table EX.1.

- In residential care (and type of home)
- Being fostered
- In other forms of care (such as being placed back with families or relatives under supervision)
- On the CPR and not in care
- Other clients

In addition to being able to assign a client to one of these groups, we needed to know the client's age and sex and current or originating ward-code.

Table EX.1 Percentages in the five client groups – by Trust

(The total in this table refer to the combined downloads – i.e. they are double the numbers on would expect at any one time)

		Percent of clients in each of these groups					
		Resid	Foster	Other BLA	CPR Not BLA	Other	Total number of clients(=100%)
South & East Belfast	%	3.35	11.60	6.35	9.70	69.00	2897
North & West Belfast	%	2.91	9.21	6.11	10.56	71.20	3778
Down & Lisburn	%	2.80	8.97	4.21	8.33	75.68	2496
North Down & Ards	%	2.13	10.83	4.61	10.42	72.01	1929
Causeway	%	1.46	15.28	6.87	7.38	69.01	1165
Homefirst	%	1.71	10.84	4.12	7.02	76.30	4899
Armagh & Dungannon	%	0.68	6.39	4.54	3.27	85.11	2049
Newry & Mourne	%	2.12	9.42	2.65	14.60	71.22	945
Craigavon & Banbridge	%	1.44	7.59	3.32	4.64	83.00	1594
Foyle	%	2.12	8.20	5.72	7.28	76.68	4339
Sperrin Lakeland	%	0.87	5.42	1.64	5.78	86.29	2750
All N Ireland	%	2.05	9.26	4.70	7.88	76.10	28841

Clients can only be included in the modelling if we have a code for their ward of residence. After the two stages of cleaning, ward codes were present on the records of all but 9.0% of the 28847 clients on the combined downloads. However, we require a pre-care address for clients in care. As this was not routinely available from SOSCARE, the client ward code was been replaced by the next of kin (NOK) ward code. Because address details of next of kin may be incomplete on client record systems, we could only attach these codes to a limited number of clients and just under a third (31%) of the 4625 clients in care have no NOK ward code. The data set was weighted to compensate for these cases lacking ward codes.

EX.3 Estimates of client costs

Client record systems do not, as yet, contain information on the cost of services provided to POC3 clients, or, in most cases, sufficient detail of services to enable them to be multiplied by unit costs. Hence we are forced to adopt a less precise methodology and assign clients to five broad groups for which we can estimate unit costs. The cost of providing services to a client in each of these five groups was computed from: the latest FR22s; data (from a previous study) on the relative social worker time required by each type of client; and further information from, and judgements of, Board and Trust finance staff. The process is described in full in the main report and led to the "signed-off" estimates in Table EX.2.

Because cost is being used as a proxy for need, we have some concern at the variation in unit costs between Trusts and have agreed with the commissioners that we will use Northern Ireland average costs rather than the local figures; not least because the latter may reflect local and historic variations in provision.

Table EX.2 Unit Costs expressed as relativities

	Resid	Foster	Oth BLA	CPR	Other	Base cost
Down & Lisburn	24.2	5.3	1.6	1.3	1.0	3642
North Down &Ards	30.6	5.3	1.8	1.6	1.0	3560
N&W Belfast	25.0	8.0	1.7	1.5	1.0	3249
S&E Belfast	20.2	7.3	1.9	1.7	1.0	2889
Causeway	27.4	7.0	2.0	1.7	1.0	3218
Homefirst	35.7	6.2	2.9	1.9	1.0	2561
Armagh & Dungannon	47.8	9.0	3.2	2.4	1.0	2563
Craigavon & Banbridge	35.7	5.4	1.8	1.3	1.0	3358
Newry & Mourne	18.8	4.9	1.4	1.1	1.0	4797
Foyle	46.8	9.5	2.2	1.9	1.0	1944
Sperrin Lakeland	56.6	9.2	1.6	1.9	1.0	2168
N Ireland average	30.94	7.09	2.07	1.70	1.00	

EX.4 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 Physical and Sensory Disability studies. 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 subset of variables selected for the POC3 work, mostly belong to one of the following four groups:

General deprivation and social structure

The literature on the likelihood of a child or young person being a client of social services points to a relation with deprivation. We therefore considered the inclusion of a variety of both census and administrative data based measures of deprivation as a priority. The variables chosen 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

These were obviously important to include.

Educational attainment of the population

The 2001 census recorded that approximately 42% of the population of Northern Ireland aged 16-74 had no educational qualifications. Even though the overall annual rate of contact between children or young people in the population and the social services is only 2.5%, the number of twenty year olds who will *ever* have been in contact with social services will be considerably larger. We have therefore included the rates of people without educational qualifications and several other variables relating to education.

EX.5 Supply variables

As for many health and social services, there is a possibility that the supply of, and access to, services may influence the take-up and volume of services used. Ward level measures of access to services are not always easy to obtain or compute. One ‘supply’ variable was found to be sufficiently robust and intuitively credible to include in the regressions: the distance from the population centroid of the client’s ward to the nearest FCC social worker base. This variable was also found to be endogenous with an appropriate set of instruments in the context of modelling the most likely sub-set of potential need drivers. Hence we had to adopt two-stage least squares methods in order to identify the variables for final models.

EX.6 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.

EX.7 Results of the modelling

Much of the work on this project has been related to the successive development of the dependent variables. These developments have either followed the release of improved ward coding, additional activity data, or address (ward code) details for the next of kin of those being looked after in care. There have also been several iterations in the production of cost estimates, until finance directors agreed the figures presented in this report.

Not achieving good specification has been a problem throughout. In the early runs, the costed dependents tended to perform slightly better than their uncosted counterparts, but in these early analyses we failed to achieve models that approach specification. The breakthrough occurred when we decided to compensate for missing ward codes directly, rather than presume that these would be covered by the Trust dummies.

The final set of dependent variables provides much better specified models. When these were initially explored with a LOG-LIN form we noticed that the type of misspecification was consistent with incorrect functional form. As there is a case (based on their distributional properties) for log transforming a significant subset of the dependent variables, we re-ran the modelling with LOG-LOG transformations and arrived at well-specified models.

As already noted, the selection of need drivers had to be achieved by two-stage least squares methods because the supply variable, the minimum distance to a children's social services facility, was found to be endogenous with the most likely set of need drivers for both the costed and uncosted dependents. The two stage selection process identified the same need drivers for both the costed and uncosted modelling. The log-log form of the 2SLS models with these variables are well-specified.

The second stage of the 2SLS process is to re-model the variables with OLS (excluding the supply variable) in order to compute the coefficients for the allocation formulae. The results of the OLS runs are shown in Table EX.3. The need drivers included in these runs are as follows (all are logged):

- Children in income support households (GKIS)
- Proportion of 16-18 year olds not in full-time education (GKPROPNST)
- Social environment score (GSOCENV)
- Proportion of children in households in owner occupation (GKOWNOCC).

Possible substitutes for all of these were explored. The Noble child poverty score was substituted for the income support variable, but its broader definition tended to make the remaining variable choice less stable and increased misspecification. The Noble education score was a possible substitute for the proportion not in full-time education, but was rejected for similar reasons to those applying to the poverty score. The social environment score covers a range of phenomena, such as overcrowding and has previously been found to be a substitute for the (negative) use of a measure of children in two car households. Various combinations of social and private rental variables were used as alternatives to the (again negative) owner-occupation variable, but were not found to perform as well.

Table EX.3 Coefficients when variables selected by 2SLS are re-run in an OLS model without the supply variable.

Var name	Variable short definition	Coefficients for the costed and uncosted models	
		Uncosted	Costed
GKIS	Children in income support households	0.3127	0.371
GPROPNST	Prop of 16-18 yr olds not in full time education	0.1985	0.389
GSOCENV	Social environment score	0.1361	0.760
GKOWNOCC	Children in owner occupation	-0.1511	-0.626

EX.8 Testing the model – and subsequent recommendations

We have applied a wide range of sensitivity tests and followed the Deloitte and Touche recommendations for unmet need. In all, we have investigated:

- Robustness at Trust level
- The impact of excluding Boards from the modelling
- The effects of using an enhanced set of dummy variables
- The consequences of excluding the clients in residential care
- Possible unmet need effects relating to population density, deprivation and non-linearities associated with variables already in the model

None of the unmet need tests gave strong or unambiguous results so we do not think there is a case for amending the coefficients in the core models.

As noted above, in an attempt to improve the robustness of the model, we have tried to substitute similar variables for the less robust components of the model and have repeated much of the analysis without these variables. In both cases we have been unable to produce a more robust model and are therefore happy to recommend the four variable models shown above without any corrections.

The variables in both models are associated (either positively or negatively) to the types of factors that are cited as creating a need for social services support for families and children. The variables cover a good range of such factors: i.e. poverty, housing conditions, social environment and educational involvement. The statistical properties of the models with the costed and uncosted dependent are similar in many respects. As they were derived by 2SLS techniques, it is not possible to make a simple comparison of their explanatory powers from these stages of the analyses, but we get some indication from the values of r-squared for the final OLS runs. The OLS run for the costed dependent has an overall adjusted r-squared of 47.8% of which 5.2% is due to the Trust dummies. The corresponding figures for the uncosted dependent are 60.4% and 12.4%. On this basis, the uncosted model looks to have approximately

5.5% greater explanatory power, which combined with its greater robustness, gives it better overall statistical characteristics than the costed version.

The other issues to consider when choosing between the two are the nature and derivation of the cost weights. The argument for using a cost weighting is that it goes some way to taking account of the severity of individual cases, but this advantage largely relies on the validity of the weightings. While these were the best that could be done with the resources to hand, they do rely on social worker activity data collected in an earlier survey and the CFRG may want to consider whether the resulting weights will fully reflect the current distribution of social worker time between different client groups.

As both models use the same variables, the issue of how regularly these can be updated applies equally to both. In the past, not all these variables would have been capable of regular updating, but we suspect that it should be possible to develop acceptable approximations that will be revised more regularly than the decennial censuses.

- The income support variable should be regularly updateable, although in common with other claimant counts it is susceptible to changes in benefit regulations and levels of up-take.
- The proportion of 16-18 year-olds not in full-time education is derived from the census, (and performs better than the substitute in the modelling, the Noble education score) . However, it may be possible to estimate the numbers not in education on a regular basis and not rely on decennial censuses.
- The Noble social environment score may still contain values from the 1991 census, and has been converted from older ward boundaries for the purposes of this exercise. In England the equivalent suite of deprivation indices has just been revised to take account of the new census data and ward boundaries, so we hope that the same will be true for the Northern Ireland version.
- The proportion of children in owner-occupied properties derives from the census, but as with the education variable, there may be some opportunity to regularly update this from trends in housing tenure.

Main report

1. Introduction

1.1 Background

The current Family & Child Care programmes (POC3) needs formula covers all aspects of Family Support Services within the constraints of available data and in line with the guiding principles of the Children Order (1995).

Research undertaken in 1999/2000 involved the analysis of activity data at ward level. Both the numbers of clients and the cost of services per ward were modelled against a wide range of potential socio-economic need drivers to determine the extent to which variation in numbers and costs could be explained by differences in socio-economic conditions. Costs were estimated by assigning clients to one of five groups and computing an average per capita cost for each of the groups from the Trusts' annual finance returns and a survey of social work workload. The need drivers were derived from a variety of sources including 1991 census data and administrative data sets.

Some limitations with the work conducted in 1999/2000 have been identified, vis:

- the modelling undertaken assumes that utilisation, after controlling for supply, is an adequate proxy for need. However, such an approach may not fully reflect the relative distribution of need where differential unmet need exists between areas;

- the original model used data that dated from early days of the Children (NI) Order, and it is recognised that there may have been instability in the underlying data from this period;

- the analysis did not include any access related supply variables;

- costs used in the model may not accurately reflect differences between local populations and categories of client because of the methods of estimation and apportionment; and

- the original formula was not comprised entirely of updateable variables and future modelling should more easily allow for periodic updating of coefficients and variables.

1.2 The present study

The present project has the aim of updating the existing formula. It does this by firstly collecting and cleaning POC3 activity data. Then it uses these data, in conjunction with needs drivers and supply variables, to attempt to develop a model that has sufficient explanatory power and specification to provide the basis of equitable allocations of revenue resources to this programme of care.

The modelling is conducted with ward rather than local authority level data. Models built from local authority level activity data have often been criticised for their tendencies to be biased by the peculiarities of local authority behaviour and for reproducing historic inequities in funding. By working at a ward level, the modelling can control for such influences.

In more detail, the main activities of the study were as follows:

- updating activity data and finance data;
- construction of an updated cost dependent variable using the above data and the workload estimates from the work that generated the current model;
- development of access related supply variables;
- construction of new needs variables using the 2001 Census and updated administrative sources;
- modelling of the activity data, potential need drivers and supply variables.
- sensitivity testing of the resulting models
- assessing whether there is any evidence of unmet need, manifest as non-linearities in the fit of the model to the data
- applying any unmet need corrections as appropriate

This report deals with each of these in turn.

2. Activity data

2.1 The downloads

The basic aim of the exercise is to model levels of service activity in POC3 against the socio-economic characteristics of the areas from which the recipients of the services originate. The first stage in this process is to check, clean and combine the various data files on service activity that have been extracted from the SOSCARE system.

On the advice of the project commissioners, six files were requested from each of the 11 Trusts. These consisted of three snap-shots of activity on two dates: 31/2/02 and 30/9/02.
The three snapshots are:

- All clients with open cases on the date
- All pending clients at the date
- All Family Resource Centre (FRC) clients at the date.

Extracts covering open and pending cases were received for all trusts. The information on FRC users was limited to 3 or 4 Trusts and was too incomplete to include in the modelling.

In relation to the subsequent analyses, the key fields in the open cases extract were:

- Client age and sex
- Ward or postcode of residence – or pre-care code for clients in care
- Type of care (including name of institution)
- Whether on Child Protection Register (CPR)

The specification asked for many other details, including information on services received, but the recording of these was too inconsistent for them to be used in the analysis. Information on pending cases was even more limited.

The minimum data requirement in order to replicate the analysis that generated the current formula was to have sufficient information on the type of care and services to be able to assign a client to one of five groups:

1) In residential care (and type of home)
2) Being fostered
3) In other forms of care (such as being placed back with families or relatives under supervision)
4) On the CPR and not in care
5) Other clients

In addition we needed to know client age and sex and current or originating ward-code. The originating ward code is required for those currently in care because the analysis aims to identify the characteristics of areas generating need, rather than the characteristics of the areas where clients are placed in care.

2.2 Main stages in the data handling

There were many stages to the data checking and cleaning. At various points during the project we received additional data, including updates to the client postcode details. When we noticed that many of the client in care postcodes seemed to refer to institutional addresses, we requested information on next of kin postcodes that were used to derive a ward code for all children in care.

The main tasks in this process were as follows:

- All open and pending files from both extract dates were merged.
- FRC files were excluded as too few Trusts have been able to provide these data.
- Placement codes were attached to the data to identify the type of placement.
- The improved postcode to ward code files were matched with the client data, including the second round of improvements for Causeway and Homefirst.
- The client ward code was removed for all clients in care and replaced by the next-of-kin ward code.

2.3 More details of the ward coding

After the two stages of cleaning, ward codes were present on the records of all but 9.0% of the 28847 clients on the combined downloads. (Table 1). For clients in care, the client ward code was been replaced by the next of kin (NOK) ward code. Because address details of next of kin may be incomplete on client record systems, we could only attach these codes to a limited number of clients. Just under a third (31%) of the 4625 clients in care have no NOK ward code and the percentage missing varies from 14% in S&E Belfast to 57.8% in Foyle. (Table 1).

Table 1 Extent of missing ward codes – all clients in combined download

	All client groups			Clients in care		
	N	N without Ward code	Percent W/o code	N	N without Ward code	Percent W/o code
South & East Belfast	2897	254	8.8	617	85	13.8
North & West Belfast	3778	355	9.4	689	149	21.6
Down & Lisburn	2496	227	9.1	399	149	37.3
North Down & Ards	1929	87	4.5	339	64	18.9
Causeway	1165	78	6.7	275	65	23.6
Homefirst	4899	318	6.5	817	282	34.5
Armagh & Dungannon	2049	126	6.1	238	70	29.4
Newry & Mourne	945	106	11.2	134	71	53.0
Craigavon & Banbridge	1594	157	9.8	197	41	20.8
Foyle	4339	530	12.2	696	402	57.8
Sperrin Lakeland	2750	355	12.9	218	40	18.3
All N Ireland	28847	2593	9.0	4619	1418	30.7

The 28847 clients in the combined file consist of the 14602 clients in the March downloads and 14245 in the September extracts. The former represent 2.92% of the 2001 census population aged 0-19 and the latter 2.85%. The highest ratios of clients to population are in North and West Belfast and Foyle; the lowest are in Causeway and Newry and Mourne. (Table 2). We have used the numbers aged 0-19 in the population as the denominator in this table, and for subsequent work, as the very small numbers of (rather arbitrarily distributed) clients aged 20 and over do not justify using 20 or 21 as the upper limit for the population denominator.

Table 2 Numbers of clients per 1000 people aged 0-19 in the population

	All clients		BLA clients		CPR clients	
	31/3/02	30/9/02	31/3/02	30/9/02	31/3/02	30/9/02
Trust	Per 1000 Popln	Per 1000 Popln	Per 1000 Popln	Per 1000 Popln	Per 1000 Popln	Per 1000 Popln
S&E Belfast	29.78	28.26	6.13	6.13	3.53	3.98
N&W Belfast	41.61	40.38	7.58	7.50	5.65	5.83
Down&Lisburn	24.40	22.72	3.81	3.74	2.62	2.51
NDown&Ards	23.83	26.03	4.34	4.47	3.51	3.67
Causeway	19.96	20.83	4.59	5.04	2.52	2.24
Homefirst	25.64	26.25	4.37	4.28	2.15	2.44
Armagh&Dungannon	31.61	31.70	3.61	3.80	1.67	1.70
Newry&Mourne	15.49	16.61	2.24	2.31	2.48	2.58
Craigavon&Banbridge	21.71	22.52	2.63	2.86	1.22	1.16
Foyle	41.55	38.47	6.29	6.55	4.22	3.56
Sperrin Lakeland	39.48	34.12	2.97	2.86	2.06	2.49
All	**29.20**	28.48	**4.60**	4.65	**2.93**	2.99

When we subdivide the clients into the groups that will be used in the analysis, we see differences in the distributions of types of care provided across the 11 Trusts. For example, the proportion being looked after varies from 7.9% in Sperrin Lakeland to 21.3% in S & E Belfast and 23.6% in Causeway.

Table 3 Numbers and percentages in the five client groups used for the analysis (by Trust)

		Resid	Foster	Other BLA	CPR Not BLA	Other	All Groups
		1	2	3	4	5	6
South & East Belfast	n	97	336	184	281	1999	2897
	%	3.35	11.60	6.35	9.70	69.00	100.00
North & West Belfast	n	110	348	231	399	2690	3778
	%	2.91	9.21	6.11	10.56	71.20	100.00
Down & Lisburn	n	70	224	105	208	1889	2496
	%	2.80	8.97	4.21	8.33	75.68	100.00
North Down & Ards	n	41	209	89	201	1389	1929
	%	2.13	10.83	4.61	10.42	72.01	100.00
Causeway	n	17	178	80	86	804	1165
	%	1.46	15.28	6.87	7.38	69.01	100.00
Homefirst	n	84	531	202	344	3738	4899
	%	1.71	10.84	4.12	7.02	76.30	100.00
Armagh & Dungannon	n	14	131	93	67	1744	2049
	%	0.68	6.39	4.54	3.27	85.11	100.00
Newry & Mourne	n	20	89	25	138	673	945
	%	2.12	9.42	2.65	14.60	71.22	100.00
Craigavon & Banbridge	n	23	121	53	74	1323	1594
	%	1.44	7.59	3.32	4.64	83.00	100.00
Foyle	n	92	356	248	316	3327	4339
	%	2.12	8.20	5.72	7.28	76.68	100.00
Sperrin Lakeland	n	24	149	45	159	2373	2750
	%	0.87	5.42	1.64	5.78	86.29	100.00
All N Ireland	n	592	2672	1355	2273	21949	28841
	%	2.05	9.26	4.70	7.88	76.10	100.00

3. Data requirements for small area modelling

3.1 Introduction

As described above, the basic aim of the exercise is to model levels of service activity against the socio-economic characteristics of the areas from which the recipients of the services originate. If activity is regarded as a proxy for need, the relation between activity and area characteristics can be used as a model to predict need and, hence, to allocate resources consistent with need.

The study has to address several problems with this approach, especially the assumption that current activity is a valid proxy for need. Firstly, for a variety of reasons, local service providers may provide different levels of service to the same client groups (especially where provision is non-mandatory). They may also provide services in different ways, which will have cost implications. Access to, and availability of services outwith the activity data may also have an impact on the uses or need for the designated service. For example, the availability of in-patient health care facilities may have an impact on the need for health and social services to provide care in the community.

Moreover, it is widely argued that some need for most health and social services goes unmet and that unmet need is not evenly distributed across the population. The most common positions are that services are most likely to fail to meet the needs of people in the most deprived and most rural areas. If such arguments are valid, current activity data will understate the needs of such populations and some independent measure of need (possibly survey based) may be required as an adjustment.

Finally, there is the problem of what set of areas to use in the modelling. This is potentially problematic as the populations of the areas must be sufficiently small to be socio-economically distinctive, but sufficiently large to give robust estimates of both activity and needs drivers.

These problems and proposed solutions are summarised in Table 4.

Table 4 Potential problems for the modelling – and possible solutions

Potential problem	Possible solution
Differences between Trusts in patterns or levels of service delivery	Dummy Trust variables
Supply of other types of services	Include variables measuring access to other services and tests for endogeneity of these variables
Unmet need in certain types of areas	Consider/apply unmet need adjustments
Area base too small for reliable estimates of activity	Use combinations of wards to ensure no area in modelling has population of less than 2000

The questions of how to introduce Trust level dummy variables and test for supply side effects are considered at appropriate points in the description of the modelling; as is the issue of unmet need. The next section of the report provides a short note on the choice of small areas for the modelling.

3.2 The development of synthetic wards for the modelling

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" 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. 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 on Table 5. The average ward populations range from 2500 to 3500 in all except the two Belfast Trusts.

Table 5 Numbers of synthetic wards – and average ward populations (by Trust)

	N synthetic wards	Population	Average population per ward	Average population 20 and under per ward
Armagh&Dungannon	39	101963	2614	863
Causeway	33	99196	3006	914
Craigavon&Banbridge	41	119760	2921	914
Down&Lisburn	53	172482	3254	1041
Foyle	51	162267	3182	1108
Homefirst	109	327762	3007	904
N&W Belfast	27	143491	5314	1783
Ndown&Ards	48	149629	3117	843
Newry&Mourne	31	89338	2882	984
S&E Belfast	47	200361	4263	1143
Sperrin Lakeland	45	118935	2643	862
Total	**524**	**1685184**	**3216**	**998**

4. Constructing measures of activity for use as dependent variables in the modelling

4.1 Introduction

The purpose of creating synthetic wards was to reduce the number of areas with such small populations that the estimates of activity and socio-economic characteristics might be unreliable. The tactic was mostly successful in that only one of the 524 synthetic wards had zero activity, although 34 did have less than 5 clients. On average there were 27.6 clients per ward. Nearly half of the wards accounted for 10-30 clients. The maximum number of clients per ward is 170, but we do not treat this as an outlier as it is a large ward (more than twice as many people under the age of 20 than the average) with by no means the highest rate of clients per head of population.

Table 6 Distribution of client numbers by synthetic wards

Numbers of clients	Numbers of synth wards	Percent of synth wards
1	1	0.2
2 to 10	121	23.1
10.5 to 20	156	29.8
20.5 to 30	83	15.9
30.5 to 60	109	20.8
60.5 and over	53	10.1
All	523	100

The basic measure of activity is the number of clients per head of population. In this case, the relevant population consists of 0-19 year olds. Given the relatively small age range and negligible numbers of deaths in this group, there is an argument for not standardising for age and sex - such differences between areas usually appear only at adulthood when people are mobile in their own right, or due to differential mortality rates. However, we have used age-sex standardisation in preparing the dependent variables at the request of the project commissioners.

Because of the relatively small numbers of clients per ward, we have chosen to use indirect standardisation when computing the dependent variables. Whereas direct standardisation adjusts the number of clients per ward to the number that would be present if the ward had the average national age-sex composition, indirect standardisation uses the national rates for clients per head of population in each age/sex group to compute an expected number of clients for each ward. The dependent variable is then the ratio of the actual to the expected numbers of clients per ward.

In fact, we have computed two dependent variables, the client ratio (as above) and a similar ratio based on the numbers of clients where each client is weighted by the national average cost of the type of care they receive.

4.2 Estimating the costs of care

The point of modelling cost of care, rather than client numbers, is that cost is presumed to represent a more accurate proxy for need. There are a number of difficulties with this argument, not least that the type and cost of care may in part reflect local capacity and supply, as well as additional need. Historic differences in provision and policy may indeed explain some of the variation between Trusts in the proportions of children in care. (Table 3).

The major difficulty with this approach is how to estimate costs. We are still a long way from being able to retrieve costed information on care from client records for this programme of care. Hence we have had to divide the clients into 5 very broad groups and will attempt to estimate an average cost for each of these groups, first within each Trust and then for Northern Ireland as a whole.

The method used to estimate costs is basically the same as for the previous exercise in 1999/2000, with some improvements in the handling of Board and other centrally funded clients and some corrections to exclude known non-POC3 funded activities. The main principles are as follows:

1. for residential and fostering, we have allocated the corresponding budget lines shown in the current FR22.
2. Where Trusts have responsibility for Board and other centrally funded clients in special schools and residential training schools, we have added the average cost of these placements into the residential line of the Trust FR22s.

3. We have tried to exclude elements in the FR22s that are funded from budgets other than the POC3 – such as SureStart initiatives.
4. Social worker cost is attributed as before, using the results of a survey that was conducted for the previous exercise(See Table 7). This gives an estimate of the social worker time devoted to each type of client and is used to distribute the total Social Work budget on the current FR22.
5. the amount for other services have been distributed between other BLA, CPR and Other Not BLA according to the number of cases in each category.
6. the amount allocated for administration has been distributed pro-rata.

Interim calculations were circulated to finance directors on two occasions and the final principles for this year's estimates were agreed at a workshop with Board finance representatives. Estimates based on these principles were subsequently circulated and either confirmed, or, in a couple of cases, amended.

In the previous exercise, we used the unit costs shown in Table 8. We remarked at the time on the relatively large variations between the Trusts – between two and four fold - which was put down to relatively small numbers and the estimation method used.

Table 7 Average social worker time (minutes per year) by client/care group

From survey during 1999 modelling study.

	Average social work time per case type (minutes per year) from 1997 survey					
	Residential	**Foster**	**Other BLA**	**CPR not BLA**	**Other**	**All**
Down & Lisburn	6592	5107	4505	3391	2097	4308
North Down &Ards	5098	6413	5783	5008	2369	4483
N&W Belfast	8297	7377	5619	4923	2340	4725
S&E Belfast	9795	8904	5870	5031	2285	5038
Causeway	7788	6123	7726	6206	2972	5180
Homefirst	4903	4022	7254	4150	1403	3729
Armagh & Dungannon	13421	11611	9261	6779	2226	6531
Craigavon & Banbridge	9009	6321	5747	3888	2821	4666
Newry & Mourne	9564	9288	6285	4573	4104	5837
Foyle	3174	3187	2273	1902	689	1775
Sperrin Lakeland	4310	5725	1968	2588	906	2404
All	**7129**	**6675**	**5671**	**4368**	**2177**	**4308**

Table 8: Unit Costs used in 1999 Modelling Exercise

	Residential	Fostering	Other BLA	CPR	Other
Armagh&Dungannon	79346	22638	3399	2717	1528
Causeway	62690	12237	5938	5801	3298
Craigavon&Banbridge	175821	12950	4542	3461	2946
Down&Lisburn	53526	12055	4066	3290	2483
Foyle	50633	11543	3688	3425	1898
Homefirst	76399	9420	5393	3545	2042
N&W Belfast	106736	11897	4451	4281	3080
Ndown&Ards	53892	11874	6841	6355	4836
Newry&Mourne	200748	10085	4070	3595	3264
S&E Belfast	81030	14399	4536	4017	2395
Sperrin Lakeland	75513	10746	2472	3020	1587
Range	**£51K-£201K**	**£9K-£23K**	**£2.5K-£7K**	**£2.7K-£6.4K**	**£1.5K-£3.3K**
Northern Ireland Av	**71508**	**11952**	**3889**	**3354**	**2360**

The results of the present costing exercise are shown in Tables 9 and 10. These were produced by a multi-stage process, finally involving a workshop with Trust and Board finance representatives, which resolved issues such as how to handle the costs of residential care and whether to include certain income streams, such as SureStart in the calcuation of social service unit costs: a working note from this meeting is included as Annex 3. Because cost is being used as a proxy for need, we have some concern at the variation in unit costs between Trusts and would suggest using Northern Ireland average costs rather than the local figures; not least because the latter may reflect local and historic variations in provision. There are also difficulties in using locally specific costs when constructing an indirectly standardised dependent variable. Local costs are easier to introduce with direct standardisation, but directly standardised variables are less likely to generate robust models.

We also, note that there are in principle two possible sources for the cost of each case: either the costs of the providing agency or the average costs of the Trust in which the client resides. Whilst we have made separate estimates, there is in fact very little difference between the two.

Table 9 Unit Costs – as agreed with Finance Directors (see Annex 3 for method of derivation)

UNIT COSTS (£s per annum)	Residential	Foster	Other BLA	CPR not BLA	Other
14.00 Down & Lisburn	88248	19169	5805	4805	3642
15.00 North Down & Ards	109087	18910	6507	5838	3560
13.00 North & West Belfast	81065	26117	5506	5027	3249
6.00 South & East Belfast	58375	21114	5559	4934	2889
21.00 Causeway	88189	22512	6423	5398	3218
23.00 Homefirst	91453	15771	7408	4836	2561
31.00 Armagh & Dungannon	122407	23156	8150	6179	2563
34.00 Craigavon & Banbridge	119747	18176	5912	4289	3358
33.00 Newry & Mourne	90092	23603	6478	5158	4797
46.00 Foyle	90979	18461	4272	3727	1944
48.00 Sperrin Lakeland	122697	19987	3404	4126	2168
Maximum	122697	26117	8150	6179	4797
Minimum	58375	15771	3404	3727	1944
N Ireland weighted average cost	87741	20095	5863	4826	2836

Table 10 Unit Costs (from Table 9) expressed as relativities

	Resid	Foster	Oth BLA	CPR	Other	Base cost
Down & Lisburn	24.2	5.3	1.6	1.3	1.0	3642
North Down &Ards	30.6	5.3	1.8	1.6	1.0	3560
N&W Belfast	25.0	8.0	1.7	1.5	1.0	3249
S&E Belfast	20.2	7.3	1.9	1.7	1.0	2889
Causeway	27.4	7.0	2.0	1.7	1.0	3218
Homefirst	35.7	6.2	2.9	1.9	1.0	2561
Armagh & Dungannon	47.8	9.0	3.2	2.4	1.0	2563
Craigavon & Banbridge	35.7	5.4	1.8	1.3	1.0	3358
Newry & Mourne	18.8	4.9	1.4	1.1	1.0	4797
Foyle	46.8	9.5	2.2	1.9	1.0	1944
Sperrin Lakeland	56.6	9.2	1.6	1.9	1.0	2168
N Ireland weighted average ratio	30.94	7.09	2.07	1.70	1.00	

The Northern Ireland weighted average relative costs (from the bottom line of Table 10) have been used in constructing the cost weighted variable. In order to compute the standardised dependents we have computed national average age-sex rates for clients and cost weighted clients. These rates are shown in Table 11 and are used to compute the expected client numbers for each ward.

Table 11 Age-sex client rates (uncosted and costed) per 1000 in population

	Males, aged:				Females, aged:			
	0-4	5-9	10-14	15&over	0-4	5-9	10-14	15&over
Rates	27.6	33.0	33.2	24.2	24.8	30.2	29.1	27.8
Costed rates	47.9	60.1	95.0	71.8	42.7	51.9	76.3	73.8

Rates for clients aged 15 and over are based on the numbers of 15-19 year olds in the population.

5. Needs drivers

5.1 Introduction

As this project is one of three modelling services for different programmes of care, we had an interest in constructing a set of potential needs drivers that could be used across all the projects.

One starting point was to include variables relating to the nine equality groups, viz:

Gender
Age
Marital Status
Ethnic Origin
Religion
Whether dependents or not
Whether disabled or not
Political Opinion
Sexual Orientation.

Given that the Programme of Care is concerned exclusively with children, we will include variables that predominately relate to children. However, we have been unable to include some potentially relevant variables, such as low birth weight, as the most recent values for these variables were not available (at ward level) in time for the modelling. We have also included some variables relating to adults, such as claimant counts and household composition as these may be better general markers of social conditions than some of the more specific children's variables; though in some cases we were able to obtain figures for the numbers of children in households where the adults are receiving these benefits.

We have some concerns regarding the census based data relating to children. The Northern Ireland census tables follow much the same structure as the Standard Tables for England and mostly provide information only for households with children, not for numbers of children in households. In England this has now been addressed by the release of special "Theme" tables relating to children, but we have not yet seen these for Northern Ireland. Consequently, some of the variables are less precise than their nearest 1991 equivalents.

There is also the more general point of whether to de-emphasise census based measures in favour of variables that could be more frequently updated. We have mixed feelings on this point, especially given the volatility of the definitions and payment structures of the benefit system, but have followed the commissioners instructions to make maximum use of frequently updated variables.

A great deal of effort went into the construction of the dataset of potential needs drivers for the modelling. 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 2. The main areas covered as follows:

- Demographic variables based on vital statistics (population and SMRs)
- Census data from the 2001 Census
- Claimant counts from Social Security
- Index of Multiple Deprivation and components thereof
- Health variables derived from the Northern Ireland Survey of Health and Social Well-Being

5.2 Choice of Independent Variables for this modelling exercise

The subset of variables selected from thePOC3 work, mostly belong to one of the following four groups:

General deprivation and social structure

The literature on the likelihood of a child or young person being a client of social services is substantial demonstrates a relationship with deprivation. We therefore considered the inclusion of a variety of both census and administrative data based measures of deprivation as a priority. The variables chosen 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

These were obviously important to include.

Educational attainment of the population

The 2001 census recorded that approximately 42% of the population of Northern Ireland aged 16-74 had no educational qualifications. Even though the overall annual rate of contact between children or young people in the population and the social services is only 2.5%, the number of twenty year olds who will *ever* have been in contact with social services will be considerably larger. We have therefore included the rates of people without educational qualifications and several other variables relating to education.

5.3 Supply Variables

In addition, two proxies for supply variables have been constructed, based on postcodes of relevant social work offices.

FCCSMIN the minimum travel time from ward centroid to any social work office.

FCCSAV the average travel time from ward centroid to all social work offices.

We have considerable doubts on the credibility of a variable that is based on the average distances from all social work offices in Northern Ireland to a client's ward of residence. Moreover, it was either not significant in a number of early trial models, or had a fluctuating sign. For example, in one set of runs FCCSAV was not significant, though it had the expected sign when entered on its own. When used with fccsmin it changes sign and is insignificant with the cost weighted dependent and significant with the unweighted version. Given these uncertainties regarding its meaning and behaviour, we have decided to abandon the average distance measure in favour of the variable that represents the distance to the nearest base for FCC social workers. From here on we only examine the role of FCCSMIN in the modelling.

5.4 Characteristics and inter-correlations of needs drivers

The basic characteristics of all the variables in the modelling exercise are shown in a table at the end of Appendix 2. The inter-correlations of a subset of the independent variables are shown in Table 12.

Unsurprisingly**,** there are high levels of inter-correlation between many of these independent variables and therefore problems of multi-collinearity. This means that the final choice of variables to enter the model is, to a certain extent, arbitrary. It also requires a cautious approach to avoid the modelling being destabilised when near equivalent variables are entered in the same run. For these reasons, we never entered all 81 variables in a single run. In any event, there are insufficient observations (520) to support the simultaneous use of all these variables.

Table 12 Correlations between components of deprivation scores and census and claimant count variables

	KLOWSEGA	KLOWSEGB	KHGHSEGA	KHGHSEGB	KNOCARA	KNOCARB	KNOCARC	KTWOCARA	KTWOCARB
OVSC	.813	.829	-.697	-.712	.804	.812	.820	-.593	-.631
INCSC	.899	.905	-.706	-.727	.792	.784	.768	-.553	-.598
EMPSC	.750	.771	-.698	-.710	.690	.706	.719	-.578	-.613
HLTHSC	.698	.714	-.639	-.658	.738	.750	.760	-.621	-.662
EDUCSC	.717	.737	-.778	-.798	.730	.745	.754	-.608	-.656
ACCSSC	.011	-.015	.115	.087	-.515	-.553	-.593	.737	.722
SOCENVSC	.358	.370	-.419	-.422	.671	.682	.695	-.714	-.738
HOUSSC	.080	.086	-.202	-.199	-.077	-.065	-.042	.074	.082
CHILDPSC	.843	.858	-.765	-.779	.823	.827	.823	-.665	-.705

	KTWOCARC	KOWCRAA	KOWCRAB	KPRVRENA	KPRVRENB	KSOCRENA	KSOCRENB	KOWNOCCA	KOWNOCCB
OVSC	-.668	.747	.847	.134	.131	.757	.778	-.475	-.792
INCSC	-.636	.831	.882	.156	.109	.773	.744	-.379	-.753
EMPSC	-.641	.652	.764	.173	.183	.639	.681	-.433	-.713
HLTHSC	-.690	.634	.744	.152	.169	.673	.711	-.460	-.739
EDUCSC	-.685	.577	.681	.069	.061	.721	.772	-.490	-.766
ACCSSC	.706	.053	-.110	.149	-.056	-.358	-.481	.661	.483
SOCENVSC	-.735	.315	.444	.027	.148	.578	.632	-.596	-.656
HOUSSC	.079	.047	.100	.243	.272	-.093	-.054	-.017	-.030
CHILDPSC	-.737	.746	.839	.119	.120	.796	.811	-.507	-.821

	HH3K	LONPU24	LPDEPK	KNBASIC	KNOCENTH	KLLTI	KNGH	HHDEPK	HHALLK	KCATHOL	KPROT
OVSC	.368	.765	.790	.201	.282	.645	.519	.146	.139	.422	-.438
INCSC	.508	.761	.773	.235	.332	.630	.481	.283	.278	.561	-.567
EMPSC	.350	.682	.689	.205	.268	.590	.467	.094	.093	.431	-.434
HLTHSC	.305	.682	.737	.174	.164	.641	.519	.103	.084	.396	-.421
EDUCSC	.138	.708	.722	.089	.302	.599	.447	.075	.073	.146	-.173
ACCSSC	.434	-.431	-.566	.216	.377	-.395	-.358	.437	.525	.142	-.022
SOCENVSC	-.046	.647	.700	-.052	.014	.502	.422	-.119	-.200	.158	-.230
HOUSSC	.177	-.049	-.055	.115	.314	.007	-.014	-.058	-.044	.057	-.029
CHILDPSC	.359	.797	.818	.169	.315	.668	.493	.141	.134	.441	-.471

6. The modelling – general issues and early attempts

6.1 Overall approach

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.

Although all the variables listed in the Family and Child Care Section of Annex 2, together with a large subset of the adult variables from the same Annex , were initially included in the work, the first five steps listed above, together with some partial correlations and analyses of residuals, were used to reduce the number of variables to a large but manageable sub-set. This set was then subjected to all the above stages.

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 from 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.

6.2 Difficulties in achieving specification with early versions of the dependent variable.

There have been several attempts at the modelling, using different versions of the activity data.

- The first proved to have incomplete data on client wards of residence and a set of supplementary fields were provided for several Trusts.
- The second was based on estimates of cost per client from the modelling that generated the current formula and provisional estimates of new costs, prior to obtaining a fully agreed set of costs.
- The third was able to use the agreed costs, but the correction for missing ward codes may have introduced some distortion into the costed version of the activity variable.
- The fourth, as reported here, incorporates all the above improvements to the data and costs, plus enhanced weighting to deal with cases lacking valid ward codes.

The first three attempts at the modelling were plagued by problems of misspecification – suggesting that the formulation of the regression equation or one of the underlying assumptions is incorrect. Specification errors can occur for various reasons, including:

- Wrong functional form
- Missing need driver
- Incoherent or not fully explicable phenomenon
- Effect of outliers
- Unrecognised supply factors

6.2.1 Functional form

All the versions of the activity variables (including those described here) have had distributions that (as indicated by their values of kurtosis and skewness) considerably depart from normality. In all cases a log transformation improved these distributional properties.

A similar transformation is justified in the case of many of the independent variables. In searching for a well specified model the first three sets of modelling tried a wide range of functional combinations including those shown in Table 13. More esoteric strategies were also tried, but we consistently found that the LOG-LIN form (with untransformed independent variables) provided the closest approximation to specification.

Table 13: Possible Dependents and Functional Forms (all these have been explored)

	Choice of dependent, all per population 0-20 of the ward			
Functional Form	Client Numbers	Cost using 1997 estimates	Average Northern Ireland Cost	Cost using information supplied for 2002
Linear – no transformation				
Square root both sides				
Log-log				
Logged dependent, unlogged independents				

6.2.2 Further needs drivers

A very wide range of need drivers was included in the set of potential predictors. As expected we found that the residuals from a best fit four or five variable model did not significantly correlate with any of the remaining (potential) predictors – as expected if the modelling has selected the best subset of variables. Hence we were unable to identify a missing need driver that would improve the specification.

6.2.3 Inexplicable phenomenon

It is, of course, possible that the phenomenon cannot be fully modelled by linear or transformed linear relationships with measures of socio-economic characteristics. In which case we should be asking whether the type and extent of the misspecification should be a cause for concern in resource allocation.

6.2.4 Outliers

We have found that removing one or two wards with very low levels of activity improves the specification slightly with some versions of the dependent variable.

6.2.5 Supply factors

There are two types of supply effects that could lead to the type of misspecification that we observed with earlier versions of the dependent variables.

1) Supply factors that might explain excessive levels of activity in wards where activity is already high. Proximity to social services FCC bases is the most obvious factor, but other services, such as the presence of drug support teams, may direct additional clients to social services.

2) Conversely, in low activity areas, we are looking for supply factors that may suppress demand. Again, lack of access to social services could be an issue, but it may also be the case that health and education services are taking over clients who would otherwise have been referred to social services if these were more accessible. This would be a substitution effect in low activity areas.

6.3 Attempts to improve specification by recalculating the dependent variable

The most recent efforts to improve the specification have focussed on the properties of the dependent variable: the age-sex standardised level of activity (both cost weighted and unweighted). When checking the age-sex standardisation of this variable, we noticed that the variable was not fully corrected to take account of cases with missing ward codes. We had not considered this to be a serious problem as we had only intended to weight up cases at Trust level, so any correction would be counteracted by changes in the values of the Trust dummies. This was not entirely correct, especially in relation to the cost weighted variable.

The problem is that the proportion of missing ward codes are not evenly distributed across all types of clients. They tend to be higher for the groups in residential and foster care, for whom we have had to try to use a next-of -kin address as the basis for the ward code of origin. These groups are also the higher cost groups. If these were under-represented in the dependent variable, it would tend to have lower than expected values in the highest need areas and the model might well be misspecified.

To investigate this potential problem we established the number of cases with missing ward codes for each of the five client groups in each Trust. We then removed these cases and weighted up each of the groups in each Trust to correct for these exclusions.

We considered weighting-up by age, sex, Trust and client type, but there were too few clients in some of these combinations for this to be practical. Because we are only weighting by Trust and client type, there is a risk that the weighted data used in the modelling may have a significantly different age-sex distribution from that in the original data set. However, some basic checks, reported in Table 14, show this not to be the case: the age-sex distribution of the weighted data is very similar to that of the original.

Table 14 Impact on client age-sex distribution of correcting for missing ward codes (from which we conclude that the weighting has not led to distortion)

	Males, aged:				Females, aged:			
	0-4	5-9	10-14	15&over	0-4	5-9	10-14	15&over
Client numbers								
Uncorrected N	3219	4099	4506	3234	2786	3607	3769	3593

Corrected	N	3264	4174	4523	3179	2774	3622	3768	3538
Uncorrected	Percent of all	11.2	14.2	15.6	11.2	9.7	12.5	13.1	12.5
Corrected	Percent of all	11.3	14.5	15.7	11.0	9.6	12.6	13.1	12.3
Costed clients									
Uncorrected	N	5535	7325	12134	9729	5042	6275	9970	9827
Corrected	N	5674	7589	12918	9419	4787	6218	9867	9394
Uncorrected	Percent of all	8.4	11.1	18.4	14.8	7.7	9.5	15.1	14.9
Corrected	Percent of all	8.6	11.5	19.6	14.3	7.3	9.4	15.0	14.3

As for previously constructed dependent variables, the latest versions have values for both skewness and kurtosis (Table 15) that suggest it will continue to produce poorly specified models. Applying a log transformation produces distributions that are much closer to normal (Table 15).

Table 15 Properties of logged and unlogged versions of dependent (activity) variables

VN1 = standardised client number per ward
VC1 = standardised client cost per ward
GVN1 and GVC1 are logged (base10) versions.

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
VN1	523	0.0222	5.0933	0.9762	0.7389	1.5473	3.1424
VC1	523	0.0097	6.7922	0.9646	1.0255	2.2602	6.5792
GVN1	523	-1.6546	0.7070	-0.1326	0.3434	-0.4189	0.3792
GVC1	523	-2.0137	0.8320	-0.2345	0.4624	-0.2878	-0.0666

7. Results of the final modelling

7.1 General properties of the models

As noted above, in the previous attempts at modelling we found that the LOG-LIN form gave the results that were closest to being specified. In the final attempt at modelling, we found that specification could be achieved by using logged transformations of the independent variables. The statistical properties of these variables support the case for a number of them being logged, but, conceptually, a LOG-LIN model has advantages over its LOG-LOG counterpart.

Some of the well-known disadvantages of the latter are:

- Dealing with variables have values close to or at zero.
    - This can, of course, be circumvented by subtracting proportions from 1 (one) but tends to generate opacity rather than transparency; and for a few of the variables create further problems as their highest values are near 1.

- The resulting model (log-log) would be multiplicative which is not always easy to interpret in a practical context (whilst it provides direct estimates of the relation between marginal changes in an independent variable and those in the dependent variable of interest or 'point elasticities', the policy interest is not in marginal changes but in changes over the whole range).

Nevertheless, the nature of the misspecification – revealed by inspection of the residuals, suggested that a change in functional form would be the appropriate remedy and from here on all the modelling reported uses LOG-LOG forms.

7.2 Correlations with potential need drivers

An indication of the variables that are likely to be most important in the modelling can be had by correlating the potential need drivers with the residuals obtained by modelling the activity variables with the Trust dummies and supply variable.

The results of these correlations for the cost weighted variable are shown in Table 16. As noted above, several of these variables have been "inverted" prior to logging by subtracting their values from 1. These are identified in the third column of the table. Negative values for several other variables have been avoided by (arbitrarily) adding the smallest constant that removes negatives.

There are two main points of note. Firstly, almost all the variables in the data set are highly significant correlates of activity (though this is helped by the number of cases in the data). This suggests that many alternative models could be found, with different groups of variables, but producing very similar results.

Secondly, the strongest predictors roughly fall into the following groups:

- Measures of income for households with children, such as children in income support households, or the child poverty component of the Noble index.
- Proportion of young people not in full time education, or some measure of their educational attainment, such as the Noble education score
- Children in families with lone parents
- A measure of housing tenure – in this case a strong negative correlation with the proportion of households with children in owner occupation. (Variables relating to separate forms of rental were excluded from this test data set, as they perform less well than the owner occupation measure)

- Some measure of poor health in the population – either self-reported from the census, or a SMR

Again, because of the high level of correlation between many of these variables, there is no guarantee that representatives of all these groups will be present in the final models.

Table 16 Correlation coefficients between residual variance of SOSCARE activity variable (after removing effects of Trust dummies) and variables in the standard modelling set. (See Annex 2 for variable definitions)

Var name	Abbrev description	inverted	Corr	Sig(2-tail)
GCHILD	child poverty score		0.5120	0.0000
GKIS	income support (children)		0.5054	0.0000
GEDSC	education score	*	0.4907	0.0000
GLOWSEG	Low SEG		0.4752	0.0000
GOVSC	Noble overall score		0.4749	0.0000
GPROPNST	Proportion 16-18 not in ft educ		0.4744	0.0000
GLPDEPK	lone parents(children)		0.4732	0.0000
GADIS84	adults income support		0.4729	0.0000
GINCSC	income score		0.4658	0.0000
GSWIDV	single widowed, divorced		0.4646	0.0000
GADIS	income support (adults)		0.4562	0.0000
GADNGH	not good health		0.4425	0.0000
GEMPSC	employment score		0.4400	0.0000
GADUNEM	unemployment		0.4389	0.0000
GADLLTI	limiting long-term illness		0.4325	0.0000
GU75SMR	smr		0.4310	0.0000
GU65SMR	smr		0.4277	0.0000
GHLTHSC	health score	*	0.4248	0.0000
GLLTI	limiting long-term illness		0.4242	0.0000
GADJSA	job seekers allowance		0.4183	0.0000
GADDLA	disability living allowance		0.4159	0.0000
GNOCARER	no carer		0.4132	0.0000
GKLOWSEG	Low SEG (children)		0.4082	0.0000
GKOWCROW	overcrowding		0.4026	0.0000
GEXMAR	divorced		0.4020	0.0000
GKWFTC	WF tax credit (children)		0.3971	0.0000
GNOQUAL	no qualifications		0.3853	0.0000
GKDLA	disab living all..ce (children)		0.3738	0.0000
GKLLTI	limiting l-t illness (children)		0.3712	0.0000
GSOCENV	social environment score	*	0.3703	0.0000
GADWFTC	WF tax credit		0.3502	0.0000
GDENS	Pop density		0.1248	0.0043
GACCSC	access score	*	-0.2011	0.0000
GKJAS	job seekers all..nce (children)	Y	-0.2436	0.0000
GKNGH	not good health (children)	Y	-0.2551	0.0000
GADLP	lone parents	Y	-0.2942	0.0000
GMAR	married		-0.3726	0.0000
GKTWOC	two cars		-0.4038	0.0000
GLONPU24	lone parent under 24	Y	-0.4377	0.0000
GKOWNOCC	owner occupied		-0.4692	0.0000

**Negative values of some of the Noble index components have been avoided by adding constants. The values added are: educsc 3.0; healthsc 2.5; socenvsc 2.5; accesssc 2.5.*

7.3 Testing the supply variable for endogeneity

At an early stage in the modelling it is necessary to test whether the supply variable is "endogenous", that is the extent to which it is, itself, predicted by the need drivers in the model, and candidate need drivers that were excluded during the modelling. Only the one supply variable FCCSMIN is being considered. This is not endogenous when tested with LIN-LIN and LOG-LIN models for both the costed and uncosted dependent, but with the LOG-LOG form the supply variable proves to be strongly endogenous with both the costed and uncosted dependent. A typical value of the test statistic $F(1,504)$ is 7.9, compared with the 5% threshold value of 3.8 Therefore a two stage least squares (2SLS) approach has to be be used.

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 and 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.

7.4 Derivation of the models by 2SLS

Modelling the costed dependent

This preliminary work suggested that the main group of variables to consider are those shown in Table 17.

Table 17 Main variables surviving the preliminary selection procedures.

Variable name	Definition	
GKIS	log(base10)of	Proportion of children in income support households
GPROPNST	log(base10)of	Proportion of 16-18 yr olds not in full-time education
GSOCENV	log(base10)of	Noble Social Environment score
GKOWNOCC	log(base10)of	Proportion of children in owner occupied housing
GKDLA	log(base10)of	Proportion of children in DLA claimant households
GLPDEPK	log(base10)of	Proportion of children in lone parent h/households
GLLTI	log(base10)of	All age standardisesd self-report lilmiting long-term illness
GNOCARER	log(base10)of	Persons in no carer households
GCHILD	log(base10)of	Noble child poverty score
GINCSC	log(base10)of	Noble income score
GMAR	log(base10)of	Persons who are married or living in a relationship
GEMPSC	log(base10)of	Noble employment score
GSWIDV	log(base10)of	Persons who are single widowed or divorced

Of these the standardised limiting-long-term illness variable and the Noble income score almost always have the wrong sign due to being highly correlated with better predictors of costed activity. Hence they are discarded at a relatively early stage.

The proportion who are single widowed or divorced (SWIDV) and those who are married or living in a relationship (GMAR), apart from being closely intercorrelated, tend to introduce similar effects to the owner occupation variable, and, again, are less strong predictors of costed activity. Hence they are both rejected.

The proportion of children living in lone parent households (GLPDEPK) is a weak predictor and ceases to be significant when almost any other variables are added.

Both the income score (GINCSC) and the child proverty score (GCHILD) are not significant when the proportion of children in income support households (GKIS) is already in the model; not surprising as the income support variable is one of the components of the child poverty score. Including GCHILD (the Noble child poverty score) reduces the significance of GKIS.

Although GNOCARER was suggested by the preliminary search process, is was not significant in the 2SLS runs.

The resulting model for the costed dependent therefore uses some combination of the variables listed in Table 18.

Table 18 Choice of variables for the final model

Variable name	Definition	
GKIS	log(base10)of	Proportion of children in income support households
GPROPNST	log(base10)of	Proportion of 16-18 yr olds not in full-time education
GSOCENV	log(base10)of	Noble Social Environment score
GKOWNOCC	log(base10)of	Proportion of children in owner occupied housing
GKDLA	log(base10)of	Proportion of children in DLA claimant households

Removing any of the first four of these reduces the adjusted R-squared by several percent. Adding GKDLA only increases r-squared from 49.2 to 49.6%.

Table 19 Output from 2SLS – with costed dependent and preferred four variables

```
Dependent variable.. GVC1

Listwise Deletion of Missing Data

Multiple R           .71150
R Square             .50623
Adjusted R Square    .49160
Standard Error       .32292

            Analysis of Variance:

                DF   Sum of Squares      Mean Square

Regression      15        54.111263        3.6074176
Residuals      506        52.779827         .1042756

F =      34.59504       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable             B         SE B       Beta         T  Sig T

FCCSMIN       -.016400     .004230   -.228759    -3.877  .0001
```

TRDUMA1	.064906	.068013	.035370	.954	.3404
TRDUMA2	-.060192	.071702	-.030620	-.839	.4016
TRDUMA3	-.108451	.061939	-.062101	-1.751	.0806
TRDUMA4	-.117407	.055802	-.079443	-2.104	.0359
TRDUMA5	-.089824	.057513	-.061807	-1.562	.1190
TRDUMA7	-.194165	.063062	-.124393	-3.079	.0022
TRDUMA8	.083431	.061421	.049355	1.358	.1750
TRDUMA9	-.164939	.072953	-.085938	-2.261	.0242
TRDUMA10	-.042921	.058586	-.028449	-.733	.4641
TRDUMA11	.209819	.072485	.122107	2.895	.0040
GPROPNST	.499161	.103745	.210022	4.811	.0000
GSOCENV	.518502	.149029	.168925	3.479	.0005
GKOWNOCC	-.480063	.178728	-.142151	-2.686	.0075
GKIS	.378083	.072927	.309152	5.184	.0000
(Constant)	.283705	.134378		2.111	.0352

To test for specification we take the explained sum of squares when the residual from this model is regressed on the set of instruments used in the test for endogeneity (which include the variables in the model) and divide this by the residual sum of squares from the 2SLS regression, divided by the degrees of freedom. The result should be distributed as chi-squared 9. The value of the test statistic is 11.77 (the 95% point of chi-squared 9 is 16.9 – so this model is specified. The instruments used in the tests for endogeneity and specification are as follows.

GADNGH	GKJAS	GKIS	GADJSA	GEDSC
	GMAR	GKTWOC	GKOWNOCC	GLPDEPK
GPROPNST	GSOCENV	GNOCARER	GADIS	GACCSC

The 2SLS modelling points to four variables that should be included in the final model for the costed dependent. All the four variables are associated (either positively or negatively) to the types of factors that are cited as creating a need for social services support for families and children. The variables cover a good range of such factors: i.e. poverty, housing conditions, social environment and educational involvement.

The supply side variable is removed and these four variables are re-modelled using OLS in order to arrive at the coefficients to be used for purposes of computing allocations. The results are shown in Table 20. It should be remembered that these are values prior to any adjustments for unmet need.

Table 20 Values of constant and unstandardized coefficients in recommended model for the costed dependent *(model computed with trust dummies but no supply variable)*

GKIS	Children in income support households	0.371
GPROPNST	Prop of 16-18 yr olds not in full time education	0.389
GSOCENV	Social environment score	0.760
GKOWNOCC	Children in owner occupation	-0.626

Uncosted dependent.

Table 21 shows the variables that result from a manual 2SLS search with the uncosted dependent. One variable in this set is particularly problemmatic: the proportion of children in lone-parent premium benfit households (GKADLP). Because this variable had been inverted to avoid losing cases in the log transformation, the positive coefficient that appears for all the variants in Table 21 is counterintuitive. This unexpected sign arises because of strong correlations between GKADLP and GKIS, GKOWNOCC and GKDLA. So if any of these three are in the model GKADLP takes the “wrong” sign. It will also have a counterintuitive sign if the other lone parent variable is forced into the formula.

Table 21 Standardized coefficients (and significance of coefficients)

		Model 1	Model 2	Model 3	Model 4	Model 5
		All 7 vars	Drop census lone parent var	Drop social environment score	Drop social census LP var and children in DLA h/holds	Drop social environment score and census LP variable
GKIS	Income support	.2815 (.0000)	.3813 (.0000)	.3249 (.0000)	.4378 (.0000)	.4356 (.0000)
GPROPNST	Not in f/t educ	.1252 (.0011)	.1419 (.0002)	.1209 (.0018)	.1485 (.0001)	.1385 (.0004)
GSOCENV	Soc env score	.0979 (.0169)	.1051 (.0113)	x	.1029 (.0138)	x
GKOWNOCC	Owner occupation	-.3762 (.0000)	-.3947 (.0000)	-.3961 (.0000)	-.3909 (.0000)	-.4175 (.0000)
GADLP	Lone parent premium	.3564 (.0000)	.3301 (.0000)	.3821 (.0000)	.3175 (.0000)	.3559 (.0000)
GLPDEPK	Children in LP households	.1756 (.0142)	x	.1877 (.0093)	x	x
GKDLA	Children in DLA h/holds	.1155 (.0033)	.1063 (.0073)	.1144 (.0040)	x	.1044 (.0094)
R-squared (as %)		64.9%	64.2%	64.3%	63.6%	63.3%

When the 2SLS runs are repeated, paying more attention to the counterintuitive sign of GADLP, we find that GLPDEPK is not significant – and assume it only entered before because of correlation with the other (incorrectly signed) lone parent variable.

The results of 2SLS models with various permutations of the strongest candidate need drivers are shown in Table 21. Just as for the costed dependent, we find we are left with a model based on four or five variables. The standardised coefficients (and their signficance) in these two models are shown in Table 22. Both are well-specified using the instrument set that was intital employed to test for endogeneity – and was used for the tests for the costed model. As little by way of explanatory power or specification is gained by adding the fifth variable, GKDLA, the four variable version (model 2 in Table 22) is recommended.

Table 22 2SLS models based on recommended four or five variables for the uncosted dependent

		Model 1	Model 2
GKIS	Income support	.3220 (.0000)	.3127 (.0000)
GPROPNST	Not in f/t educ	.1944 (.0000)	.1985 (.0000)
GSOCENV	Soc env score	.1383 (.0010)	.1361 (.0014)
GKOWNOCC	Owner occupation	-.1463 (.0015)	-.1511 (.0002)
GKDLA	Children in DLA h/holds	.0913 (.0239)	x
R-squared (in 2SLS models)		62.4	62.0
Spec test Chi-squared 9 (critical value 16.9)		14.75	14.07

In sum, the search for models for the uncosted activity variable was carried-out in a similar way to that for the costed option and the same combination of four core variables emerge as the best predictors.

As for the costed model, the supply side variable is removed and these four variables are re-modelled using OLS in order to arrive at the coefficients to be used for purposes of computing allocations. The results are shown in Table 23. It should be remembered that these are values prior to any adjustments for unmet need.

Table 23 Values of constant and unstandardized coefficients in recommended model for the uncosted dependent ***(model computed with trust dummies but no supply variable)***

GKIS	Children in income support households	0.3127
GPROPNST	Prop of 16-18 yr olds not in full time education	0.1985
GSOCENV	Social environment score	0.1361
GKOWNOCC	Children in owner occupation	-0.1511

8. Testing the models

8.1 Sensitivity tests

8.1.1 Tests with subsets of data

The basic sensitivity test investigates the effect of removing the data for each of the Boards in turn for the preferred four variable model of costed activity. When this test is applied to the costed dependent, the coefficients of all the needs drivers in the model fluctuated and in two cases ceased to be significant at 5% or more – though in all cases there was no change of sign. (Table 24a). Children in income support households and the social environment score are the two most robust variables in terms of variations in their coefficients. The same test applied to the uncosted dependent gave slightly better results (Table 24b), with rather less variation in the standardised coefficient, except for the effect on the student variable of removing Eastern Board.

Table 24a Sensitivity tests– removing one Board at a time

Table shows standardized coefficients (costed dependent)

		Excluding Board			
	Including all Boards	1 (Eastern)	2 (Northern)	3 (Southern)	4 (Western)
GKIS	0.394	0.254	.443	.434	.404
GPROPNST	0.459	0.209*	.549	.490	.398
GSOCENV	0.606	0.530	.523	.588	.730
GKOWNOCC	-0.500	-1.058	-0.345*	-0.432	-0.503

All coefficients are significant at least 5% - with exception of two marked with asterisk

Table 24b Sensitivity tests– removing one Board at a time

Table shows standardized coefficients (uncosted dependent)

		Excluding Board			
	Including all Boards	1 (Eastern)	2 (Northern)	3 (Southern)	4 (Western)
GKIS	0.370	0.258	0.434	0.397	0.396
GPROPNST	0.178	0.010*	0.224	0.193	0.166
GSOCENV	0.171	0.178	0.113*	0.180	0.183
GKOWNOCC	-0.171	-0.326	-0.117	-0.160	-0.165

All coefficients are significant at least 5% - with exception of two marked with asterisk

A more rigorous test, that is unlikely to be passed by all components of any models with this type of potentially variable activity data, is to examine the univariate correlations at Trust level. The former test, excluding Board level data, is genuinely multivariate, that is, changes in coefficients may say less about the significance of individual variables than the extent to which individual variables play a different role in a fixed set of predictors.

The results of the Trust level correlations for the costed and uncosted dependent are shown in Tables 25a and 25b. For the **costed dependent**, owner occupation is the most robust predictor. It is significantly and negatively correlated with costed activity in all the eleven Trusts. Income support and the social environment score are significant correlates (with consistent sign) in 10 of the Trusts. The least robust measure is the proportion of 16-18 year-olds in full-time education – this was not a significant correlate of activity in 4 of the Trusts. When the modelling was repeated excluding this variable from the initial set, a stepwise procedure immediately substituted another measure of educational attainment amongst young people, the Noble education score (GEDSC) – see Table 26. The obvious implication is that we should try to include at least one of these variables referring to education in the model. The two are strongly correlated (r=0.781) and the Noble component is slightly more robust (see the far right column in the table), but models using GEDSC rather than proportion not in full time education were not well-specified. The failure of models including the Noble education score to achieve specification may be due to the broader definition of the score than the census proportion of young adults in full-time education, leading to greater multicollinearity with the other variables in the model.

Trust level correlations for the uncosted dependent are shown in Table 25b. As for the test on regressions coeficients in Tables 24a and 24b, the uncosted version performs rather better than the costed. In this case, only the student variable has any coefficients that are not signficant. The attempt to substitute the Noble score for this variable failed in the same way as for the costed dependent.

Table 25aSingle Trust correlations with cost weighted dependent

		GKIS	GPROPNST	GSOCENV	GKOWNOCC	GEDSC
6 South & East Belfast	Corr coeff	0.8099	0.6616	0.5994	-0.7231	*.608*
	Sig	0.0000	0.0000	0.0000	0.0000	
13 North & West Belfast	Corr coeff	0.6759	0.5452	0.4405	-0.6623	*.607*
	Sig	0.0001	0.0033	0.0215	0.0002	
14 Down & Lisburn	Corr coeff	0.7379	0.4597	0.6331	-0.5585	*.619*
	Sig	0.0000	0.0005	0.0000	0.0000	
15 North Down & Ards	Corr coeff	0.7480	0.6999	0.6357	-0.7364	*.706*
	Sig	0.0000	0.0000	0.0000	0.0000	
21 Causeway	Corr coeff	0.5179	0.2640	0.5211	-0.5980	*.384*
	Sig	0.0024	0.1443	0.0022	0.0003	
23 Homefirst	Corr coeff	0.5018	0.3205	0.5032	-0.5112	*.447*
	Sig	0.0000	0.0007	0.0000	0.0000	
31 Armagh & Dungannon	Corr coeff	0.3564	-0.0261	0.3191	-0.5435	*.208*
	Sig	0.0260	0.8749	0.0477	0.0003	*ns*
33 Newry & Mourne	Corr coeff	0.2962	0.3007	0.5214	-0.5331	*.252*
	Sig	0.1057	0.1003	0.0026	0.0020	*ns*
34 Craigavon & Banbridge	Corr coeff	0.6589	0.4806	0.5999	-0.6126	*.450*
	Sig	0.0000	0.0015	0.0000	0.0000	
46 Foyle	Corr coeff	0.2921	0.4276	0.1638	-0.3962	*.340*
	Sig	0.0375	0.0018	0.2506	0.0040	
48 Sperrin Lakeland	Corr coeff	0.5195	0.1716	0.4559	-0.5792	*.123*
	Sig	0.0003	0.2596	0.0016	0.0000	*ns*
N of trusts where ns		1.0000	4.0000	1.0000	0.0000	*3*

Table 25b Single Trust correlations with uncosted dependent

		GKIS	GPROPNST	GSOCENV	GKOWNOCC
6 South & East Belfast	Corr coeff	0.895	0.685	0.635	-0.787
	Sig	0.000	0.000	0.000	0.000
13 North & West Belfast	Corr coeff	0.832	0.669	0.477	-0.713
	Sig	0.000	0.000	0.012	0.000
14 Down & Lisburn	Corr coeff	0.769	0.536	0.580	-0.621
	Sig	0.000	0.000	0.000	0.000
15 North Down & Ards	Corr coeff	0.761	0.688	0.629	-0.729
	Sig	0.000	0.000	0.000	0.000
21 Causeway	Corr coeff	0.579	0.176	0.687	-0.739
	Sig	0.001	0.335	0.000	0.000
23 Homefirst	Corr coeff	0.591	0.400	0.597	-0.616
	Sig	0.000	0.000	0.000	0.000
31 Armagh & Dungannon	Corr coeff	0.402	0.033	0.332	-0.552
	Sig	0.011	0.843	0.039	0.000
33 Newry & Mourne	Corr coeff	0.437	0.303	0.429	-0.626
	Sig	0.014	0.098	0.016	0.000
34 Craigavon & Banbridge	Corr coeff	0.809	0.542	0.759	-0.784
	Sig	0.000	0.000	0.000	0.000
46 Foyle	Corr coeff	0.473	0.461	0.369	-0.567
	Sig	0.000	0.001	0.008	0.000
48 Sperrin Lakeland	Corr coeff	0.507	0.246	0.477	-0.602
	Sig	0.000	0.104	0.001	0.000
N of trusts where ns @ 5%		0	3	0	0

Table 26 Comparative properties of models with census measure of percentage aged 16-18 not in full-time education and Noble education component.

	Unstand Coeff	Signif	Unstand Coeff	Signif
GKIS	0.394	0.000	0.367	.000
GPROPNST	0.459	0.000	-	
GSOCENV	0.606	0.000	0.559	.000
GKOWNOCC	-0.500	0.005	-0.584	.001
GEDSC	-		0.519	.001

8.1.2 Tests with enhanced set of dummies

A further set of tests repeated much of the modelling 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 its local authority areas. The effect was negligible for both models. For the recommended log-log model the explanatory power of the dummies and supply variable increased from 0.131 to 0.159 and the value for the overall model increased slightly from 0.497 to 0.503. (The changes in the coefficients are shown in Table 27 and the impact on allocations in Table 28. The results shown here are based on the costed dependent in an earlier version of the data set, but the results are similar for the latest versions of both the costed and uncosted dependents.

Table 27 Changes in Unstandardized coefficients due to further subdivision of the Eastern Board dummies

	Coeffs in core model	Coeffs with enhanced dummies
GKIS	0.388	0.399
GPROPNST	0.380	0.364
GSOCENV	0.781	0.759
GKOWNOCC	-0.606	-0.569
Constant	-0.00754	0.003656

Table 28 Allocational implications of model created by further subdivision of the Eastern Board dummies

	Basic 11 dummies		Including additional dummies	
	Cost wted Popln	Percent of total	Cost wted Popln	Percent of total
EHSSB	15274	47.86%	14936	47.41%
NHSSB	5728	17.95%	5699	18.09%
SHSSB	4906	15.37%	4898	15.55%
WHSSB	6009	18.83%	5973	18.96%
	31916		31506	

8.1.3 Tests excluding clients in residential care

Although they represent a small proportion of the total client numbers, there is a suspicion that the high cost of clients in residential care and their relatively uneven distribution across Trusts may influence the modelling of the cost weighted data.

We have recomputed the models with standardised client costs excluding those in residential care – again these tests were carried out before a few minor changes to produce the final data set. The model remains well specified and none of the coefficients change sign. The impact on the value of the coefficients is negligible, except for the proportion of 16-18 year olds not in education. None of the Board level allocational implications exceed 1% (see Tables 29 and 30)

Table 29 Coefficients of models based on cost weighted dependent excluding clients in residential care

	Unstandardized coefficients			
	All clients costed	Excl. residential gp	All clients costed	Excl. residential gp
Constant			-0.00754	+0.00589
	0.394	0.392		
GKIS			0.388	0.388
GPROPNST	0.459	0.324	0.380	0.266
GSOCENV	0.606	0.635	0.781	0.762
GKOWNOCC	-0.500	-0.437	-0.606	-0.514

Table 30 Allocational implications of models based on cost weighted dependent excluding clients in residential care

Board or trust	Trust code	Model for all clients		Model excl residential gp	
	code	Cost wted Popln	Percent of total	Cost wted Popln	Percent of total
EHSSB		15274	47.86%	16277	46.34%
NHSSB		5728	17.95%	6503	18.51%
SHSSB		4906	15.37%	5602	15.95%
WHSSB		6009	18.83%	6744	19.20%
		31916		35127	
Down&Lisburn	14	3679	11.53%	3991	11.36%
N&W Belfast	13	6783	21.25%	7055	20.08%
NDown&Ards	15	1403	4.40%	1603	4.56%
S&E Belfast	6	3410	10.68%	3628	10.33%
Causeway	21	1387	4.35%	1597	4.55%
Homefirst	23	4341	13.60%	4907	13.97%
Armagh&Dungannon	31	1438	4.51%	1675	4.77%
Craigavon&Banbridge	34	1891	5.92%	2117	6.03%
Newry&Mourne	33	1576	4.94%	1809	5.15%
Foyle	46	4448	13.94%	4925	14.02%
Sperrin Lakeland	48	1561	4.89%	1819	5.18%
		31916		35127	

8.2 Unmet need tests and adjustments

In all the previous sections of this report, we have carried-out the modelling with conventional methods to establish how models based on new activity and census data compare with the model in current use. Now we turn to two sets of issues that may lead us to revise the above models: whether there may be unmet need due to problems of access and supply in, firstly, rural areas and, secondly, areas of high deprivation. We consider the former by examining role of population density in the modelling – and the second by adopting the shortfall approach described in the Deloitte and Touche report.

8.3 Population density.

There has been extensive debate on the role of measures of population density and sparsity in resource allocation formulae. On one level they can simply be acting as a proxy for accessibility measures on the presumption that services will be more thinly provided in sparse areas. Indeed, 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, particularly those that require the travel of service personnel, to sparsely distributed populations. However, even where there is clear evidence of these extra costs, our argument is that they should be addressed by specific weightings not by trying 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
- 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.

Despite these concerns we have attempted to force both a simple measure of population density variable into the model and a version of the density measure that only applies at low densities. The results are reported at the end of Annex 1. In both test the variables were either insignificant or very unstable and we conclude there is no solid evidence on which to base a population density correction.

8.4 Shortfall method for testing for unmet need

To test for unmet need effects, the project was asked to apply the methods recommended in the Sutton and Locke paper "Inequalities in Health and Social Care Use: the Implications for Resource Allocation in the HPSS". These are the same methods that 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 morbidity 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 Locke 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 Locke)

- 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 Locke article 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. The interpretation of these tests is further complicated by the use of logged variables.

We have carried out extensive tests. The results are described in Annex 1. The results of these tests lead us to conclude that there is no case for applying the unmet need corrections based on "surplus" deprivation and morbidity, as described the Sutton and Locke and AREA report, to the models recommended here.

9. Conclusions and Recommendations

Much of the work on this project has been related to the successive development of the dependent variables. These developments have either followed the release of improved ward coding, additional activity data, or address (ward code) details for the next of kin of those being looked after in care. There have also been several iterations in the production of cost estimates, until finance directors agreed the figures presented in this report.

Not achieving good specification has been a problem throughout. In the early runs, the costed dependents tended to perform slightly better than their uncosted counterparts, but in these early analyses we failed to achieve models that approach specification. The breakthrough occurred when we decided to develop new dependent variables that compensate for missing ward codes directly, rather than presume these would be covered by the Trust dummies.

The final set of dependent variables provides much better specified models. When these were initially explored with a LOG-LIN form we noticed that the type of misspecification was consistent with incorrect functional form. As there is a case (based on their distributional properties) for log transforming a significant subset of the dependent variables, we re-ran the modelling with LOG-LOG transformations and arrived at well-specified models.

The selection of variables had to be achieved using two-stage least squares methods because the supply variable, the minimum distance to a children's social services facility, was found to be endogenous with the most likely set of need drivers for both the costed and uncosted dependents. The two stage selection process identified the same need drivers in both the costed and uncosted modelling. The log-log form of the 2SLS models with these variables are well-specified. When the variables are re-modelled using OLS without the supply variable, in order to compute the coefficients for the allocation formulae, the results are as in Table 31.

Table 31 Coefficients when variables selected by 2SLS are re-run in an OLS model without the supply variable.

Var name	Variable short definition	Coefficients for the costed and uncosted models	
		Uncosted	Costed
GKIS	Children in income support households	0.3127	0.371
GPROPNST	Prop of 16-18 yr olds not in full time education	0.1985	0.389
GSOCENV	Social environment score	0.1361	0.760
GKOWNOCC	Children in owner occupation	-0.1511	-0.626

We have applied a wide range of sensitivity tests and followed the Deloitte and Touche recommendations for unmet need. In all we have investigated:

- Robustness at Trust level
- The impact of excluding Boards from the modelling
- The effects of using an enhanced set of dummy variables
- The consequences of excluding the clients in residential care
- Possible unmet need effects relating to population density, deprivation and non-linearities associated with variables already in the model

As none of the unmet need tests gave strong or unambiguous results we do not think there is a case for amending the coefficients in the core model.

In an attempt to improve the robustness of the model, we have tried to substitute similar variables for the less robust components of the model and have repeated the entire stepwise procedure without these variables. In both cases we have been unable to produce a more robust model and are therefore happy to recommend the core models (as shown in Table 31) without any corrections. Of the two models, the version with the uncosted dependent is slightly more robust. Both are well-specified and have reasonable explanatory power – as judged by the values of r-squared in the 2SLS modelling.

The variables in both models are associated (either positively or negatively) to the types of factors that are cited as creating a need for social services support for families and children. The variables cover a good range of such factors: i.e. poverty, housing conditions, social environment and educational involvement. The statistical properties of the models with the costed and uncosted dependent are similar in many respects. As they were derived by 2SLS techniques, it is not possible to make a simple comparison of their explanatory powers from these stages of the analyses, but we get some indication from the values of r-squared for the final OLS runs. The OLS run for the costed dependent has an overall adjusted r-squared of 47.8% of which 5.2% is due to the Trust dummies. The corresponding figures for the uncosted dependent are 60.4% and 12.4%. On this basis, the uncosted model looks to have approximately 5.5% greater explanatory power, which combined with its greater robustness, gives it better overall statistical characteristics than the costed version.

The other issues to consider when choosing between the two are the nature and derivation of the cost weights. The argument for using a cost weighting is that it goes some way to taking account of the severity of individual cases, but this advantage largely relies on the validity of the weightings. While these were the best that could be done with the resources to hand, they do rely on social worker activity data collected in an earlier survey and the CFRG may want to consider whether the resulting weights will fully reflect the current distribution of social worker time between different client groups.

As both models use the same variables, the issue of how regularly these can be updated applies equally to both. In the past, not all these variables would have been capable of regular updating, but we suspect that it should be possible to develop acceptable approximations that will be revised more regularly than the decennial censuses.

- The income support variable should be regularly updateable, although in common with other claimant counts it is susceptible to changes in benefit regulations and levels of up-take.

- The proportion of 16-18 year-olds not in full-time education is derived from the census, (and performs better than the substitute in the modelling, the Noble education score). However, it may be possible to estimate the numbers not in education on a regular basis and not rely on decennial censuses.
- The Noble social environment score may still contain values from the 1991 census, and has been converted from older ward boundaries for the purposes of this exercise. In England the equivalent suite of deprivation indices has just been revised to take account of the new census data and ward boundaries, so we hope that the same will be true for the Northern Ireland version.
- The proportion of children in owner-occupied properties derives from the census, but as with the education variable, there may be some opportunity to regularly update this from trends in housing tenure.

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.

Functional form

The types of multivariate regression used in this modelling assumes 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. Translations 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 translations 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 repressors 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.

In this project, the prevalence study found very difference rates of learning disability in different age and sex groups in the population. When we want to construct a ward level measure of the numbers of people with LD we then 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 LD rates are computed for each age-sex group fop each ward, then these rates are applied to the national average ward composition giving a corrected number of people with LD. 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 characterisistics, 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 he 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 LD and the proportion over people without qualifications.

The base model includes the proportion of people without qualifications 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 with no qualifications. (this is the 90% spline point). In the top 10% of wards the variable is given the value of the proportion without education 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 w 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 on 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 Unmet need tests and adjustments

To test for unmet need effects, the project was asked to apply the methods recommended in the Sutton and Locke paper “Inequalities in Health and Social Care Use: the Implications for Resource Allocation in the HPSS”. The same methods that 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 morbidity 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 Locke 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 Locke)

- 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.

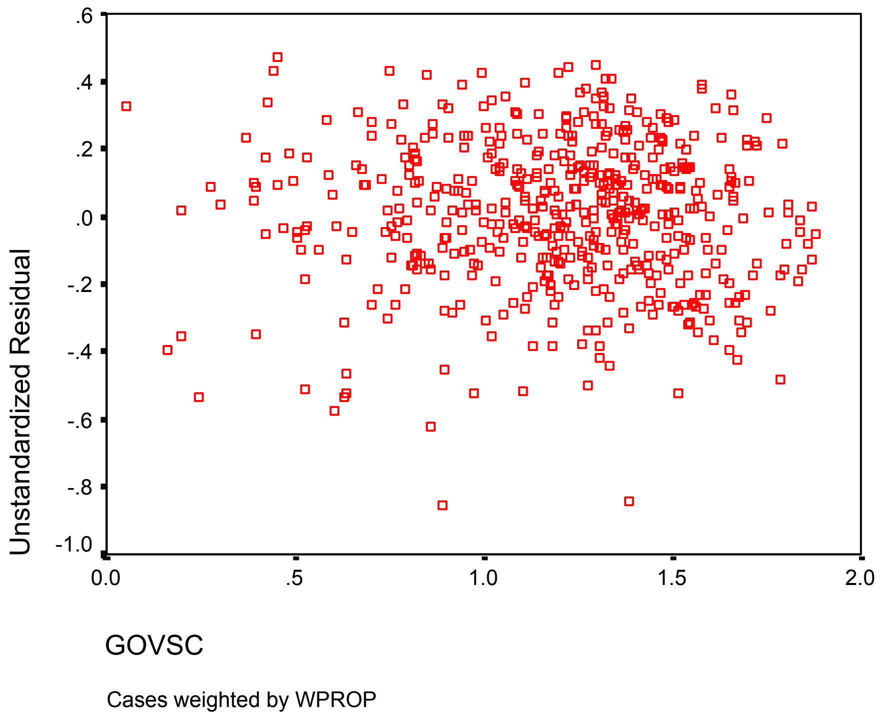
At the outset it has to be noted that the application of spline and other regression approaches to unmet need when models are derived via 2SLS is relatively unexplored territory; and is not described in either the Sutton and Locke article nor the AREA report. Indeed the latter work subsumes all supply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

In the following discussion we both test 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 A1.1 Residual from recommended four 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 four-variable model. A scattergram of this relation is shown in Figure 1. The figure shows the results for the uncosted model; the figure for the costed model is very similar. It demonstrates that there is so little evidence of a basic relation between the Noble score and the residuals from the model (Uncosted model: r=-0.046; signif=0.293 Costed model: r=-0.025; signif=0.568) that we would not expect to find spline points or significant relationships with the overall Noble score unless these were caused by interactions with variables already in the formula.

Despite the unpromising plot of the residuals from the core model against the Noble index, we have included splined variables based on the log of the Noble score in the final OLS runs with the four need drivers. One splined variable is added per run. The values of the standardised coefficients of these variables and the significance of their t values are shown in Table A.1.2 .

Table A.1.2 Coefficients and significance of coefficients when splined versions of the Noble overall score are added to the basic resource allocation models.

Range of Noble covered by spline variable	Uncosted dep		Costed dep	
	Std coeff	Signif of t value	Std coeff	Signif of t value
Highest 5%	-.059	.110	-.04	.35
Highest 10%	-.115	.008	-.109	.021
Highest 20%	-.144	.001	-.142	.004
Highest 40%	-.132	.006	-.129	.020
Highest 50%	-.175	.000	-.187	.000
Highest 60%	-.132	.006	-.129	.020
Highest 70%	-.132	.008	-.108	.059
All values	-.175	.010	-.109	.163

For almost all of the spline ranges tested, the coefficient of the log of the overall Noble score is significant and negative (see Table 2). For the uncosted model the full (unsplined) Noble variable is also significant, though this is not true for the costed model.

Superficially these negative values might be taken as evidence of unmet need. However, from Fig A.1.1 and the low correlations with the residuals, it is clear that the Noble score is significant not because it complements the variables already in the model, but because of its correlations (multicollinearity) with those variables. There are highly significant bivariate correlations between the log of the Noble score and all four logged variables in the model, with: GKIS (.881), GPROPNST (.605), GSOCENVSC (.553), and with GKOWNOCC (-0.685).

If we concentrate on the very strong correlation with GKIS and remove GKIS from the model and replace it with GOVSC, the coefficient of GOVSC becomes positive and not quite significant at 5% (prob of t=0.56). This supports the interpretation that the logged Noble score (GOVSC) is competing with GKIS but is not as strong a predictor of activity. This is confirmed when we re-examine the audit trail of the 2SLS modelling. At that stage, GOVSC is rejected in favour of the four recommended variables (see Table A.1.3 for the uncosted 2SLS analysis and Table A.1.4 for the costed version).

Table A.1.3 2SLS variable selection run including Noble score and 3 morbidity variables in addition to the 4 recommended variables (uncosted dependent)

```
Dependent variable.. GVN1

Listwise Deletion of Missing Data

Multiple R           .79488
R Square             .63184
Adjusted R Square    .61791
Standard Error       .20855

            Analysis of Variance:

                DF   Sum of Squares      Mean Square

Regression      19        37.481857        1.9727293
Residuals      502        21.839718         .0434918

F =      45.35866       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B         SE B        Beta         T  Sig T

FCCSMIN        -.014599     .003336   -.274055    -4.376  .0000
TRDUMA1         .138885     .044037    .101859     3.154  .0017
TRDUMA2        -.109806     .046738   -.075175    -2.349  .0192
TRDUMA3        -.085125     .040474   -.065602    -2.103  .0359
TRDUMA4        -.145397     .036230   -.132405    -4.013  .0001
TRDUMA5        -.010262     .038318   -.009503     -.268  .7890
TRDUMA7        -.179943     .042205   -.155150    -4.264  .0000
TRDUMA8         .060108     .039915    .047855     1.506  .1327
TRDUMA9        -.216850     .047565   -.152059    -4.559  .0000
TRDUMA10       -.085154     .040219   -.075963    -2.117  .0347
TRDUMA11        .274345     .046958    .214873     5.842  .0000
GKIS            .340129     .064164    .374298     5.301  .0000
GPROPNST        .320716     .077799    .181607     4.122  .0000
GSOCENV         .285080     .111389    .124997     2.559  .0108
GKOWNOCC       -.407975     .126252   -.162582    -3.231  .0013
GOVSC           .045851     .087476    .048052      .524  .6004
GGENHLTH       -.323854     .239901   -.156311    -1.350  .1776
GCIRCDIS        .100368     .107546    .055858      .933  .3511
GGENMH          .223346     .227993    .083904      .980  .3277
(Constant)      .259920     .141411                1.838  .0666
```

Table A.1.4 2SLS variable selection run including Noble score and 3 morbidity variables in addition to the 4 recommended variables (costed dependent)

```
Dependent variable.. GVC1

Listwise Deletion of Missing Data

Multiple R           .71242
R Square             .50754
Adjusted R Square    .48891
Standard Error       .32374

            Analysis of Variance:

                DF   Sum of Squares      Mean Square

Regression      19        54.240379        2.8547568
Residuals      502        52.628965         .1048058

F =      27.23855       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

FCCSMIN        -.017982     .005179  -.250825    -3.472  .0006
TRDUMA1         .059946     .068360   .032668      .877  .3810
TRDUMA2        -.069706     .072553  -.035459     -.961  .3371
TRDUMA3        -.115034     .062830  -.065871    -1.831  .0677
TRDUMA4        -.112387     .056241  -.076046    -1.998  .0462
TRDUMA5        -.091574     .059482  -.063011    -1.540  .1243
TRDUMA7        -.212461     .065516  -.136115    -3.243  .0013
TRDUMA8         .089588     .061962   .052997     1.446  .1488
TRDUMA9        -.171739     .073838  -.089481    -2.326  .0204
TRDUMA10       -.048407     .062434  -.032086     -.775  .4385
TRDUMA11        .193833     .072895   .112803     2.659  .0081
GKIS            .357051     .099604   .291954     3.585  .0004
GPROPNST        .470493     .120771   .197960     3.896  .0001
GSOCENV         .416789     .172915   .135787     2.410  .0163
GKOWNOCC       -.470304     .195987  -.139261    -2.400  .0168
GOVSC           .115471     .135793   .089917      .850  .3955
GGENHLTH       -.506235     .372409  -.181552    -1.359  .1746
GCIRCDIS        .093873     .166948   .038819      .562  .5742
GGENMH          .474595     .353925   .132476     1.341  .1805
(Constant)      .166608     .219519                .759  .4482
```

The Sutton and Locke paper and the AREA report make the case, especially when allocating resources to health care, for testing the effects of adding several measures of morbidity into the allocation formulae, in much the same way as we have explored the effects of adding the Noble score. The three most relevant variables in set we recieved are ward measures of:

overall morbidity

morbidity due to circulatory disorders, and

morbidity relating to mental health problems.

On the face of it, the non-statistical case for including any of these three is relatively weak as they do no refer specifically to children or teenagers and only the mental health variable has much bearing on the types of need met by the majority of personal social services. Tables A.1.5 shows the effects of forcing these variables into the basic allocation models. In both the costed and uncosted versions the coefficients of the three morbidity variables are not significant. Re-examining Tables A.1.3 and A.1.4 we also see that these variables would have been rejected in the 2SLS selection process.

Table A.1.5 Effects on coefficients of forcing the three morbidity variables into the final model

Variable	Std Coeff	t	Sig. T
Uncosted vsn			
GKIS	0.3954	6.3345	0.0000
GPROPNST	0.1310	3.0306	0.0026
GSOCENV	0.2276	5.4700	0.0000
GKOWNOCC	-0.2234	-4.6853	0.0000
GGENHLTH	-0.1578	-1.4161	0.1574
GCIRCDIS	0.0695	1.1465	0.2521
GGENMH	0.0657	0.7587	0.4484
Costed vsn			
GKIS	0.3285	4.5783	0.0000
GPROPNST	0.1574	3.1701	0.0016
GSOCENV	0.2326	4.8643	0.0000
GKOWNOCC	-0.1984	-3.6207	0.0003
GGENHLTH	-0.1630	-1.2730	0.2036
GCIRCDIS	0.0488	0.6999	0.4843
GGENMH	0.1129	1.1350	0.2569

On the basis of the above, we conclude that there is no case for applying the unmet need corrections based on "surplus" deprivation and morbidity, as described the Sutton and Locke and AREA report, to the models recommended here.

Additional tests in relation to population density.

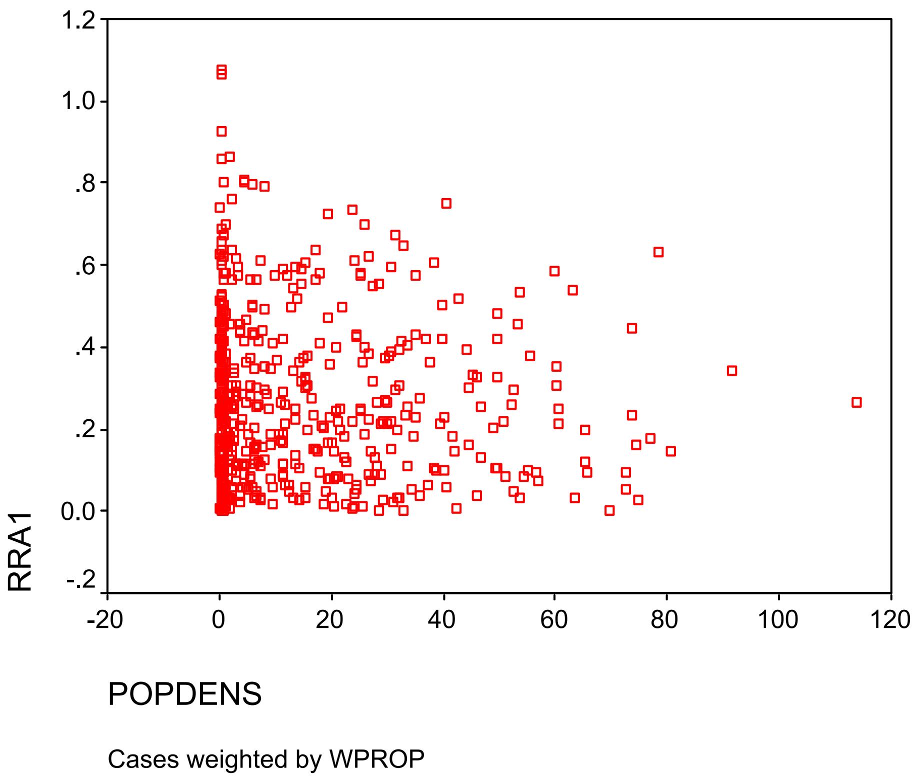
There has been extensive debate on the role of measures of population density and sparsity in resource allocation formulae. On one level they can simply be acting as a proxy for accessibility measures on the presumption that services will be more thinly provided in sparse areas. Indeed, the pressure 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, particularly those that require the travel of service personnel, to sparsely distributed populations. However, even where there is clear evidence of these extra costs, our argument is that they should be addressed by specific weightings not by trying 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
- 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.

From the relatively low value of the correlation coefficient between density and costed activity (Table 16 in main report), it is unlikely that density will be a significant predictor in the models. If we simply force density into the model with the four core variables, the standardised coefficient has a value of 0.012, with a t-value of 0.228 and a significance of 0.820. The issue here is more one of whether or not the relation between activity and density changes at low densities, or whether there is any evidence of a relation between the residuals from the model and density at low levels of density. A plot of the absolute values of the residuals from the models vs. population density shows a very wide range of residuals at low densities and no signs of the relevant trend.

Figure A.1.2 Plot of residual from core model vs. population density



We have used spline variables to check further for evidence of a relation with density in areas of low density. Three variables were used, that take the value for density in the wards with the 5, 10 and 20% lowest values and have zero values elsewhere. When these are, separately, added to the basic model we get the coefficients reported in Table A.1.6. The coefficients for the two variables relating to the areas of lowest density are both positive and significant. Superficially this is consistent with the argument that introducing these terms enable the model to better predict activity in areas of lowest density; though it might also be an effect relating to lower activity in areas of lower deprivation (density is negatively correlated with deprivation). However, if we think of this as correction factor – and the correction needs to be highest in the areas of lowest density - then we would want the spline variables to have negative coefficients. Moreover, the variable is not at all stable in its effects. When we look at the coefficient for the variable having values in the lowest 20%, this is now significant and negative; and, as previously noted, over the entire range it is not at all significant. There might be a case for using the 10% spline variables to adjust the coefficients in the model, but we are reluctant to do this because of the instability of the variable and its potentially ambiguous interpretation, especially in the context of logged variables.

Table A.1.6 Significance of splined (logged) density variables when added to core model

Range for density variable	Standardised coefficient	T test for coeff	Signif. of t
Lowest 5%	0.090	2.136	0.033
Lowest 10%	0.112	2.589	0.010
Lowest 20%	-0.111	-2.475	0.014
All values	0.012	0.228	0.820

ANNEX 2 Candidate Needs Drivers in Full Data Set

The modelling started with a set of variables relating to the socio-economic circumstances of children, supplemented by a subset of the variables relating to adults shown in the second table in this Annex.

Family and Child Care POC		Census or other source
Income		
Unemployed	ADUNEMP	KS09
JSA claimants	KJAS	
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	

More general variables largely relating to adults		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)		KS09
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

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 3

NOTES ON AGREED STRATEGY FOR COMPUTING POC3 UNIT COSTS FOR MODELLING (FROM NOV 28TH MEETING)

1.Residential

1.1. NIO funded clients

Drop all NIO funded clients from analysis as these are entirely outwith Board and Trust funding and arguably tangential to POC3. Numbers are too few to have any impact on modelling.
Relevant institutions are Lisnevin, Rathgael (Lakewood/Rathgael) and those on remand. Any clients at with these codes on SOSCARE are excluded from the calculations of unit costs.

1.2. Trust and Board funded clients in residential care

Use sum of residential budgetlines on FR22. For UCHT remove total cost for Lakewood (£2,698,270)

For Board funded clients at Glenmona, Lakewood and Brindley House – add product of number of clients and average rates to residential totals on FR22s.

Cost per place at Lakewood	96367
Cost per place at Glenmona	76440
Cost per place at Brindley House	118000

Use client numbers supplied by Glenmona and Lakewood (in table below) rather than SOSCARE counts. Can use Brindley House numbers from SOSCARE as there are approx 8 clients on the database(4 Foyle, 1 SL, 1-2 NWBT) – corresponding to Western Board's estimate.

	March 2002	September 2002	March 2002	September 2002
SEBT	1	1	4	6
NWBT	16	20	8	6
DLT	5	5	3	3
UCHT	0	0	2	1
Homefirst	4	0	4	4
Causeway	1	1	0	0
A & D	0	1	0	1
N & M	1	1	1	2
C & B	0	0	3	1
Foyle	4	4	3	3
SL	2	2	0	1
Total	34	35	28	28

Add social work and admin costs to residential totals.

2. Fostering

Total cost of fostering should be sum of budgetlines labelled family placements and all statutory payments.
At the Nov 28th meeting it was suggested that the age distribution of people in care should be taken into account (as rates vary with age – as in Table below). I've experimented with this and decided that we should not use these rates. The figure we want to compute is an average based on the total spend per Trust divided by the actual number of clients, not an age weighted number.

Age	Cost (£/week)	Ratio
0_4	71.47	1.00
5_10	81.41	1.14
11_15	101.36	1.42
16 and over	128.38	1.80

Add proportion of costs of social work and administration

3. Other children in care

Use estimates of relative social work input (from previous survey) to add social work costs, and then add proportion of admin and costs of remaining services

4. Child protection register (not in care)

Use estimates of relative social work input (from previous survey) to add social work costs, and then add proportion of admin and costs of remaining services

5. Other clients

Use estimates of relative social work input (from previous survey) to add social work costs, and then add proportion of admin and costs of remaining services

6. Which budgetlines to include in costs of other/remaining services?

Potentially everything that is not residential care, fostering, social work or administration should be included. These are mainly day centre and outreach services.

However

SureStart should be excluded. We have no information on which wards/clients receive the services; also directed at different client group from bulk of PoC3 monies. Should/can exclude when it is listed on FR22s, but when not explicitly mentioned need to check whether any Trusts have re-labelled or apportioned SURESTART income on the FR22s.

Payments appear for Down and Lisburn, Foyle and
C&B Trust - £155,864, but this is all under the community section (not Personal & Social Services) of the FR22 so no adjustment needed.

A&D Trust - £271,323, but this is all under the community section (not Personal & Social Services) of the FR22 so no adjustment needed.

N&M Trust - £394,947 is included under the heading Family Day Centres so if we are removing SURESTART funding an adjustment would need to be made for this.

Adoption may be funded at Board level everywhere except Eastern. In Eastern it appears on the South and East Belfast FR22. There are two entries, one labelled "adoption services" the other "Voluntary Organisation Support", and both may be figures for the entire Board. Could either ignore them – as similar figures may not be available for the other Boards, or apportion the S&E Belfast figure across the four Eastern Trusts.

Agreed that we would exclude costs of adoption services not least because these clients are unlikely to be on SOSCARE.

7. Board funds to voluntary agencies (not appearing on FR22s)

These should not be added/included as there is no readily available information on the geographical distribution of clients.

But there is a general (and irresolvable) problem of the costs of outreach services and other activities included on the FR22s but not specific to the clients on SOSCARE. E.g. comment from UCHT that in the more affluent areas there may be high social work costs for inspecting playgroup and child minding facilities. Can be argued that cost of these services should be excluded from calculations (as with SURESTART), as they are not geographically distributed in same way as SOSCARE clients. This has not been done as it is assumed that most of the "remaining services" listed on the FR22s will be supporting clients on SOSCARE who are not in residential placements.

8.Capital charges

Absence of separate budgetlines on most FR22s suggests these have been apportioned. However there are capital charge budgetlines on the FR22s for N&W and S&E Belfast. Paul Ballantine has supplied details of how these should be apportioned.

9.Unit costs resulting from above method of calculation

UNIT COSTS					
	Residential	**Foster**	**Other BLA**	**CPR not BLA**	**Other**
14.00 Down & Lisburn	88248	19169	5805	4805	3642
15.00 North Down & Ards	109087	18910	6507	5838	3560
13.00 North & West Belfast	81065	26117	5506	5027	3249
6.00 South & East Belfast	58375	21114	5559	4934	2889
21.00 Causeway	88189	22512	6423	5398	3218
23.00 Homefirst	91453	15771	7408	4836	2561
31.00 Armagh & Dungannon	134099	17087	5434	4170	1851
34.00 Craigavon & Banbridge	119747	18176	5912	4289	3358
33.00 Newry & Mourne	90092	23603	6478	5158	4797
46.00 Foyle	90979	18461	4272	3727	1944
48.00 Sperrin Lakeland	122697	19987	3404	4126	2168