Department of Finance and Personnel
www.dfpni.gov.uk

Creating small area income deprivation estimates for Northern Ireland: spatial microsimulation modelling

August 2008

NISRA Occasional Paper No.27

institute for social and technical research

Northern Ireland Statistics & Research Agency

University of Essex

ISTR RESEARCH REPORT
Small Area Income Deprivation Estimates for Northern Ireland

institute for social and technical research

Creating small area income deprivation estimates for Northern Ireland: spatial microsimulation modelling

DCLG SimInc WP2 (Northern Ireland) Final Report (D2.2)

Dr Ben Anderson

Technology and Social Change Research Centre

Institute for Social and Technical Research (ISTR), University of Essex

benander@essex.ac.uk

This report describes results from a preliminary investigation of the value of a spatial microsimulation technique in the estimation, for each SOA in Northern Ireland, of the incidence of income poverty as measured by the proportion of households whose income is below 60% of the UK median household income (%HHBMI). In this report we present firstly the results for 2005 using the FRS 2004/5 and the UK Census 2001 for unequivalised gross income and secondly results for 2003-5 using the pooled 2003-4 and 2004-5 FRS for equivalised net household income before housing costs. We discuss the results of validation against the source FRS and against other sources of small area income indicators including the NIMDM 2005 income domain score, and the Experian 2005 median income estimates. The report concludes with a summary of the findings and recommendations for further work.

University of Essex

http://istr.essex.ac.uk/tasc

Institute for Social and Technical Research

The work reported in this paper is part of the scientific programme of the Institute for Social and Technical Research (ISTR) at the University of Essex.

ISTR is a post-disciplinary institute employing social geographers, sociologists, political scientists, computer scientists, anthropologists, and psychologists specialising in 'socio-technical' research and consulting.

ISTR carries out research which combines the social and technological sciences to:

- generate insights into personal and social use of information and communication technologies,
- ground technological innovation in an understanding of people,
- provide analysis to support evidence-based 'information society' strategies and policies in the public and commercial domain.

We achieve this through a balanced programme of basic and applied research projects, consultancy and publication. For more information see http://istr.essex.ac.uk/tasc/

Contacting ISTR

Institute for Social and Technical Research
University of Essex
Wivenhoe Park
Colchester
Essex
UK

Web: http://istr.essex.ac.uk/tasc/

Citing This Paper

This report forms part of a research project funded by the Northern Ireland Statistics and Research Agency [NISRA].

Readers wishing to cite this paper are asked to use the following form of words:

Anderson, B (2008) Creating small area income deprivation estimates for Northern Ireland: spatial microsimulation modelling. Final report to the Northern Ireland Statistics and Research Agency [NISRA], Colchester: University of Essex.

Table of Contents

1 Introduction

As part of an investigation into potential alternative income domain indices that could contribute to future revisions of the Northern Ireland Multiple Deprivation Measures (NIMDM) we have developed a preliminary spatial microsimulation model to estimate the spatial distributions of income and income deprivation for each Northern Ireland Super Output Area (SOA) in 2005. This has been done by combining small area level data from the 2001 Census with the 2003-4 and 2004-5 Family Resources Survey (FRS). The study follows on from a literature and methodological review which culminated in a recommendation to explore the spatial microsimulation approach (Ballas, Dorling et al. 2006).

The initial income deprivation indicator selected was the proportion of households in each SOA whose gross household income was below 60% of the national (UK) median gross household income (%HHBMI) based on the FRS 2004-5. The 60% of median income threshold is currently considered a UK and EC standard poverty measure (Gordon and Townsend 2000; Eurostat 2007). Following discussion of initial results this was extended to the development of an indicator of households whose equivalised net income before housing costs (BHC) was less than 60% of the UK median equivalised net income BHC using the pooled 2003-4 and 2004-5 FRS surveys.

This report describes the implementation of the spatial microsimulation method and discusses the results of validating the estimates against the source FRS and against other sources of small area income indicators including the NIMDM 2005 income domain score and the Experian 2005 median income estimates. The report concludes with a summary of the findings and recommendations for further work.

2 Overview of Method

Following the review of methods described in Ballas et al (2006) and previous work at the LSOA level for England (Anderson 2007) this project has implemented a spatial microsimulation approach to combine Census and FRS data for appropriate years in order to estimate the indicator (%HHBMI). For the purposes of this report this requires the use of the FRS 2003/4/5 and the Census 2001.

The method requires the identification of a set of constraint variables which fulfil the following criteria:

1. They are common to both the FRS and the Census or at least can be derived from them;
2. They are available at the household level – as the indicator is at the household level;
3. They are known to be reasonable predictors of the indicator, or at least of income, at the small area level; and
4. They reasonably predict the indicator at the micro (household) level.

The constraints are then used to re-weight FRS households in each small area so that the aggregate statistics match those of the Census tables for that area.

An exhaustive review of Census and FRS data produces a list of variables that satisfy criteria 1 and 2. Criteria 3 involves filtering the results against the recommendations of the literature reviewed in phase I of the project whilst criteria 4 can be tested within the FRS using standard regression techniques.

The resulting variables are then used as input to a deterministic iterative proportional fitting method which allocates FRS households to SOAs using fractional weights in order to estimate the distributions of HHBMI.

2.1 Income definitions

In initial work (Phase I) the FRS ‘hhinc’ variable was used. This is the sum of all household income from all sources and includes gross earnings. In no cases was this value below zero.

In order to align this work with the standard DWP HHBAI definitions (HBAI07, Appendix 1) subsequent work (Phase II) defined income as the sum of all net household income from:

- Earnings & self employment (net of income tax and national insurance payments);
- Investments;
- Disability benefits;
- Retirement pensions plus any income support or pension credit;
- Working Tax Credit and/or Child Tax Credit received;
- Other pensions and benefits;
- Other/remaining sources.

The following expenditures are then removed to produce the net income before housing costs:

- Domestic rates / council tax;
- Contributions to occupational pension schemes (including all additional voluntary contributions to occupational pension schemes, and any contributions to stakeholder and personal pensions);
- Insurance premia payments made in case of sudden loss of earnings;
- All maintenance and child support payments, which are deducted from the income of the person making the payment and parental contributions to students living away from home;
- Student loan repayments.

It should be noted that households with negative net income are retained. However, households reporting negative BHC net income constitute only four (0.21%) Northern Irish households in the FRS 04-05 and therefore retaining households with negative income is unlikely to have any significant effect on the indicators as they will not substantially effect the median based calculations.

In the remainder of the report we discuss first the method and results for the initial gross income variable and secondly, using identical methods, the results for the equivalised net household income before housing costs.

3 Phase I: Gross household income

In this section we report the initial results using the unequivalised gross household income derived from the FRS 2004-5[1].

3.1 Constraint variable candidates

A review of the available household level Census 2001 tables for Northern Ireland on CASWEB (MIMAS) shows that whilst there appear to be many different small area tables, they are essentially different combinations of the following:

- Household Response Person (HRP) characteristics: Age, gender, marital status, NS-SEC, ethnicity, employment status, community background
- Household characteristics: accommodation type, tenure, number cars/vans, number rooms, composition, presence of long term limiting illness, presence of children, number of persons, number of children

As discussed in the Phase I review report, Williamson and Voas (Williamson and Voas 2000; Williamson 2005) have shown that at the small area level the variables shown in Table 1 are reasonable predictors of household income.

Table 1: Known predictors of small area income levels

Variable	Source
PEARNERS Proportion of households containing persons in employment 2001	Census
PHHSOC12 Proportion of economically active heads of household who were in social classes 1 or 2 (NS-SEC 1/2)	Census
PUSLRES Average number of residents per household	Census
PHOHETHM Proportion of heads of household from a non-white ethnic group	Census
PLLI Proportion of households containing adults suffering from a long term limiting illness	Census
PHOHCBUK Proportion of heads of households whose country of birth was the UK	Census
PCENHEAT Proportion of households with central heating in all or some rooms	Census
PROOM13 Proportion of households living in dwellings with 1 – 3 rooms	Census
FC Proportion of adults claiming family credit	DWP
JSA Proportion of adults claiming job seekers allowance	DWP

Following discussions with the NIMDM team at Oxford during the English phase of this work (Anderson 2007) we have chosen to discount the benefits derived data for the following reasons:

- To maintain independence between the experimental spatial microsimulation approach and the benefits-data derived NIMDM approach which it may eventually replace and against which it would be validated;
- The data is not at household level but at benefit unit level – although in most cases these are the same unit;
- The tax systems have changed and there is no certainty that their replacements will be equally good predictors even though they are means tested (Working Family Tax Credit, Child Tax Credit).

This leaves the Census variables described in Table 1. The list is further reduced because the FRS does not capture country of birth of the HRP, the number cars/vehicles nor the presence of central heating.

This produces the relatively short list of candidate constraint variables shown in Table 2. It should be noted that the spatial microsimulation approach requires household counts on the part of the Census data and thus a discrete set of categories on the part of the FRS data. Thus the definitions provided in Table 1 must be slightly amended and we can also introduce others that may potentially be of use.

[1] According to the FRS 2004-5, the (rounded) UK median household income was £21,800

Table 2: Constraint variable candidates

Variable	Definition	Notes
Number of earners in household	0,1,2,3+	
Employment status	NS-SEC 1, NS-SEC 2, NS-SEC 3, retired, economically inactive	This combination gives greater granularity.
Number of persons	1,2,3,4,5+	
HRP non-white	White (0), Non-white (1)	Could potentially be disaggregated but small sample size of FRS makes this problematic.
Limiting long term illness	0 (no) / 1 (yes)	Presence of at least one person with LLI
Number of rooms	1,2,3,4+	Number of rooms
Tenure	Own, rent from council, other social rent, private rent/rent free	
Sex of HRP	Male (0) / Female (1)	
Region	English Government office regions, Wales, Scotland and Northern Ireland	Used to provide regional weighting
Age	16-24, 25-34 … 85+	10 year age groupings of HRP
Composition	Couple, single parent, single person, other	
Number of children	0,1,2+	
Accommodation	Detached, semi-detached, terrace, flat, other	
Community background	Catholic, Protestant and other Christian, Other, None	Specific to Northern Ireland

3.2 Testing FRS 2004-5 constraint candidates

The first key test of the potential utility of these variables is the extent to which they predict income and, in particular the extent to which they predict the income deprivation indicators at the micro (household) level. This is relatively easy to assess using standard regression techniques and using the R-squared (or Pseudo R-squared) value as an indicator of the value of the constraint variables (Chin and Harding 2006). Whilst Chin and Harding report the use of repeated bi-variate regressions to test each variable independently (Chin and Harding 2006), this project used a stepwise or nested multivariate method. The multivariate approach means that correlations between constraint variables are taken into account and thus the 'pure' effects of each constraint can be revealed whilst the use of the stepwise technique automatically includes only those variables which have a statistically significant effect on the model and orders the resulting indicators in decreasing order of their effects which is critical to the performance of the simulation as will be discussed below. The overall model R-squared score is then an indicator of how well the included constraints predict the outcome variables (in this case HHBMI) at the household level and thus a confidence indicator for the robustness of the eventual results.

Table 30 in Annex A.1 reports the results of two multivariate logistic regression models[2] for Northern Ireland which uses the constraint variables to predict the probability of having a household income below 60% median income. Model 1 ignores any cases where employment status or, in particular, community background are unknown. This reduces the sample to 1857 households. Model 2 includes all households by coding those ignored in Model 1 as missing in the relevant category. This enables us to see if the 'missing' households are unusual in any way and to increase the sample to its maximum of 1926 households.

The results of the models do not differ in any significant way. In both cases number of earners and rooms, one or two age categories, composition and employment status are the only statistically significant variables. In neither case is Community background significant when these other variables are taken into account. Interestingly the 'missing' employment status code proves significant with households in this category being over twice as likely to be HHBMI than NS-SEC 1 coded households. Cross-tabulation of the constraints does not indicate however that these 'missing' are concentrated in any one constraint group and it is therefore difficult to characterise them in any way.

[2] Using the form $p = 1/(1+\exp(-(B_0+\Sigma B_i*X_i)))$ where B_0 is a constant and B_i are coefficients of predictor variables X_i..

Table 31 in Annex A.1 reports the results of applying the stepwise method to both models. In this case the method removes those variables that do not significantly improve the performance of the model and orders the remainder in terms of explanatory power. Table 3 summarises the increment to the R-squared score for each variable block in Model 2 and provides an indicator of the relative value of each.

Table 3: Ordered change in pseudo R-squared score for nested logistic regressions for FRS 2004-5 for Northern Ireland

Variable	Incremental R-sq	Additional R- sq
Number of earners	33.33%	
Composition	41.56%	8.23%
Employment status	43.98%	2.42%
Number of persons	45.11%	1.13%
Number of rooms	45.48%	0.37%
Age	45.51%	0.03%

Notes:
Variable blocks are added incrementally

This suggests that for the FRS 2004-5 in Northern Ireland an initial set of constraint variables, in decreasing order of importance should be:

1. Number of earners
2. Composition
3. Employment status
4. Number of persons
5. Number of rooms
6. Age

3.3 Transforming Census 2001 variables

Having identified the constraint variables it is then necessary to derive them from the Census data in terms of household counts at the SOA level with as close a match to the definitions described in Table 2 as possible. The Census source tables are described in Table 4.

Table 4: Availability of Census 2001 sources for spatial microsimulation constraints at OA and SOA level for Northern Ireland

Variable	Table	Title	Coding Notes	OA level	SOA level
Number of earners	cs118	Number of Employed Persons and method of Travel…	Used to derive 'Number of earners' as 0,1,2,3+	Y	N
Composition	ks020	Household Composition	Used to derive composition as couple, single parent, single person, other	Y	Y
HRP NS-SEC/employment status	cs046 cs013	NS-SEC of Household Reference Person (HRP) by Tenure Age of HRP and Tenure by Employment status	Used to derive NS-SEC of HRP in 3 category form[3]. Retired are computed as HRPs aged 74+ (no work data available from Census) plus those aged 'pensionable age' to 74 who are inactive[4]. Inactive are computed as HRPS who are inactive aged under pensionable age. Inactive includes students, unemployed, other non-working.	Y	N
Number of persons	uv051	Number of persons	Used to derive number of persons as 1,2,3,4,5+	Y	N
Number of rooms	uv057	Number of rooms	Used to derive number of rooms as 1,2,3,4+	Y	N

[3] http://www.statistics.gov.uk/methods_quality/ns_sec/class_collapse.asp .

[4] Because NS-SEC is coded for those who are unemployed, inactive or retired in some circumstances the NS-SEC counts have been adjusted proportionately downwards to ensure that the categories sum to the correct number of households.

Age of HRP	cs003	Age of Household Reference Person (HRP) by Sex and Marital Status (Headship)	Coded into: under 24, 25 to 29, 30 to 44, 45 to 59, 60 to 64, 65 to 74, 75 to 84, 85 and over to match Census	Y	N
Tenure	cs327	Tenure	Coded into: Owned, social rent, private rent	Y	N
Presence of limiting long term illness (LLI)	ks021	Households with Dependent Children and Households with Limiting Long-term Illness		Y	Y
Ethnicity of HRP	uv114	Ethnic group of household reference person	Coded into: white, non-white	Y	N
Number of children	uv006	Number and age of dependent children	Coded into 0, 1, 2+	Y	N
Accommodation	ks016	Dwellings, Household Spaces and Accommodation Type	Coded into: Detached, Semi-detached, Terrace, Flat/maisonette, Other	Y	Y
Sex of HRP	CS003	Age of Household Reference Person (HRP) by Sex and Marital Status (Headship)	Male = 0, Female = 1	Y	N
Community Background of HRP	CS366	Tenure by Community Background (Religion or Religion Brought up in) of Household Reference Person (HRP)		Y	N

As we can see whilst Northern Ireland SOA level data for a range of tables has been released, many of the relevant tables are not yet available through the academic source used by this project – MIMAS' CASWEB. In the English work the project aggregated OA level data however for Northern Ireland NISRA were able to supply appropriate SOA level tables.

Table 5 shows a simplified FRS 2004-5 dataset containing 4 hypothetical households from Northern Ireland with example constraint variables and income indicators. In both the FRS and Census input files all constraint categories start at 0. Thus in Table 5 household 26115 was recorded as being aged 30-44, having 4+ rooms, one person, HRP = NSSEC 2, composition = single person and one earner, a yearly income of £ 6,448 and household income below 60% of the median. Whilst the overall UK sample size is 28,023, the sample size for Northern Ireland is only 1,926 and as we will see it is these who form the basis for the spatial microsimulation process.

Table 5: A hypothetical FRS 2004-5 dataset of four households and five constraints.

case	region	age	rooms	persons	nssec	comp	nearners	hhincpy	hhbmi
26115	12	2	3	0	1	2	1	6448	100
26116	12	2	2	0	3	2	0	6708	100
26117	12	2	3	4	0	0	2	72176	0
26118	12	4	3	0	0	2	1	19084	0

Table 6: A hypothetical Census 2001 dataset of four zones.

SOAcode	Region	N_hh	rooms_0	rooms_1	…	rooms _3	persons_0	persons_1	…
95AA01S1	12	530	3	5	…	457	6	153	…
95AA01S2	12	456	0	0	…	413	93	122	…
95AA01S3	12	376	0	3	…	314	98	117	…
95AA02W1	12	783	6	13	…	537	307	251	…

Table 6 shows a simplified partial Census 2001 dataset containing part of 2 constraint variables (number of rooms, number of persons).

With the Census counts appropriately re-coded and with the FRS 2004-5 household survey data to hand we then turn to the spatial microsimulation process.

3.4 Spatial microsimulation method

The methodology used here is an adapted form of the deterministic reweighting approach described by Ballas et al (2005). The objective is to produce a set of weights linking all eligible households to all SOAs, in the sense that the weights represent the 'fractional existence' of the corresponding household in the corresponding SOA. Conceptually the results can be thought of as a matrix of SOAs (rows) and households (columns) where each cell contains the weight for that household in that SOA.

To do this two sets of tables are required for each constraint for each SOA: the Census 2001 small area tables for the constraints (e.g. Table 7) and the analogous small area tables constructed from the FRS households for the region in which the zone is found (e.g. Table 8).

Table 7: Small Area Table for number of rooms derived from Census 2001 for the first SOA in Northern Ireland

SOAcode	Number of households	Number of rooms = 1	Number of rooms = 2	Number of rooms = 3	Number of rooms = 4+
95AA01S1	530	3	5	65	457

Table 8: 'Small Area Table' for number of persons derived from the FRS 2004-5 for Northern Ireland

Number of households	Number of rooms = 1	Number of persons = 2	Number of persons = 3	Number of persons = 4
1926	2	5	345	1574

Starting with SOA 1 all household weights (w_i) are initially set to 1. As discussed above and following Ballas et al (ibid) we implemented a regional weighting scheme so that only households belonging to the same region as the particular zone are allocated to it. In this case this means that only FRS households from Northern Ireland are placed into Northern Ireland SOAs. This avoids filling, for example, Belfast with Londoners. This was achieved at the weights' initialisation step, where the weights of households that do not belong to the same region as the area in question were set to 0 rather than w_i.

Table 9: Zone 1 households with initial weights

case	region	age	rooms	persons	nssec	comp	nearners	w_i
26115	12	2	3	0	1	2	1	1
26116	12	2	2	0	3	2	0	1
26117	12	2	3	4	0	0	2	1
26118	12	4	3	0	0	2	1	1

Then, for each constraint in turn, the weights are adjusted using the formula:

$$Nw_h = w_{ih} * c_{hj}/s_{hj}$$

where Nw_h is the new household weight for household h, w_{ih} is the initial weight for household h, c_{hj} is element hj of the Census data table (Table 7) and s_{hj} is element hj of the FRS 2004-5 statistical table (Table 8).

Putting this into practice for the four example households above, there are only two categories for the first constraint (rooms) – 3 room and 4+ room households thus producing two different weights. As Table 10 shows the effect so far is to weight down the FRS 4+ room households less than the FRS 3 room households for this zone so that the FRS sample fits the Census distributions.

Table 10: Zone 1 households with weights after fitting to constraint 1

case	region	Rooms	w_i	W_1
26115	12	3	530/1926 = 0.275	= 1 * (457/1574) = 0.29
26116	12	2	0.275	= 1 * (65/345) = 0.188
26117	12	3	0.275	= 1 * (457/1574) = 0.29
26118	12	3	0.275	= 1 * (457/1574) = 0.29

Having adjusted the weights for the first constraint the process then moves sequentially through each constraint variable multiplying each new weight by that produced by the previous step. Since the last constraint to be fitted will necessarily be fitted perfectly, it is necessary to order the variables in 'r sq

contribution' order (see Table 3) so that the last to be fitted is the one that accounts for the most variation in the outcome variable of interest (HHBMI in this case).

Having passed over all constraints once the process then loops back to constraint one and repeats the re-weighting starting from the weight produced in the last step (by the last constraint). Ballas et al (ibid) found that iterating the procedure between 5 and 10 times produced weights that reduced the error in fitting households to areas to a point where it no longer declined. Our experimentation (described in (Anderson 2007)) suggested that 20 iterations were sufficient to achieve a stable indicator value. Thus after iterating over the re-weighting procedure 20 times the simulation then moves on to the next zone and repeats the process.

In order to select whole numbers of households for each zone, Ballas et al (ibid) report using a process of integerisation once the weights had been calculated to select the 'best fit' households for a given area. This process turned the weights into whole numbers (integers) in order to select the top n where n is the number of households required for the ward. Ballas et al report that this integerisation produced some extremely poor results when tested against the Census distributions and described a swapping algorithm to swap households between their 1991 wards in order to reduce errors and produce a better fit.

Since it is likely that the integerisation process will inevitably reduce within-zone variation and for our purposes it is not necessary that each small area is allocated a whole number of households, we have not implemented the integeristion process. Instead our simplified method allows the final household weights for each small area to remain fractional so that all possible survey households are retained. Our experience is that this simplified method produces distributions that perform at least as well as Ballas et al's more complex combination of integerisation and household swapping.

The spatial microsimulation process has been implemented as a java-based tool which produces an output file summarising the input variables of interest (in this case the percentage of households below 60% median income) for each zone (in this case SOAs).

Table 11: Example simulation output file (partial)

AREA	area_reg	HH_id	WEIGHT	hh_reg	hhincyearly_04_05	hhbmi_uk04_05
95AA01S1	12	26115	0.01	12	6448	100
95AA01S1	12	26116	0	12	6708	100
95AA01S1	12	26117	0.61	12	72176	0
95AA01S1	12	26118	0.02	12	19084	0

Table 11 shows the first four rows of an example output file. For each area (e.g. SOA) there is an FRS household with a specific weight. Notice that this weight can be zero. In addition there is the FRS household's yearly income and HHBMI indicator as calculated in the source FRS data (see Table 5). In addition any number of other FRS variables can be included provided that we can be confident that they are predicted by the chosen constraint variables.

Calculating the percentage of HHBMI is thus a straightforward matter of summing the weighted hhbmi indicator (i.e. the sum of weight * hhbmi) for each area and dividing by the number of households in that area. Similarly any other statistic can be calculated – such as the median household income or the variance for each area. Finally it is a relatively straightforward matter to change the chosen indicator. It simply requires the new indicator to be calculated in the source FRS data and the process of constraint variable testing to be repeated before re-running the simulation process.

3.5 Results

The results of this process are summarised in Table 12 which shows the 5 'poorest' and 5 'wealthiest' SOAs in Northern Ireland according to their simulated %HHBMI.

Table 12: The 5 'poorest' and 5 'wealthiest' SOAs in Northern Ireland by simulated %HHBMI

	SOA code	SOA Name	Local Government District	%HHBMI
1	95GG40S2	Shankill_2	Belfast	58.97
2	95GG19S2	Crumlin_2	Belfast	57.92
3	95MM27S1	Strand_1	Derry	56.66
4	95GG04S3	Ballymacarrett_3	Belfast	56.32
5	95GG35S2	New Lodge_2	Belfast	56.18
1	95AA01S1	Aldergrove_1	Antrim	6.72
2	95WW20S3	Mallusk_3	Newtownabbey	7.05
3	95WW19S2	Jordanstown_2	Newtownabbey	8.09
4	95XX01S2	Ballycrochan_2	North Down	8.36
5	95HH02S2	Bluefield_2	Carrickfergus	10.09

Figure 1 shows the spatial distribution of the simulated %HHBMI across Northern Ireland at the SOA level and indicates for the main part that the areas with highest income deprivation are the urban areas of (especially) Belfast and Londonderry with a small number of other areas to the north and west visible.

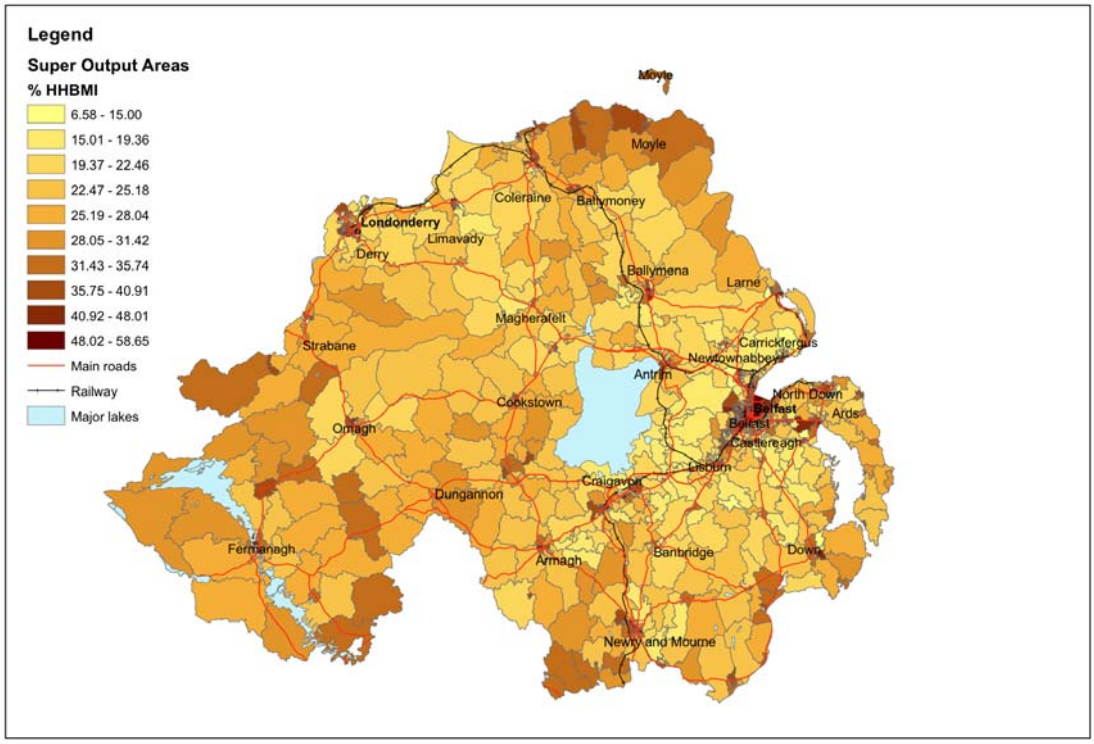

Figure 1: Spatial Distribution of simulated %HHBMI at SOA level (Cities and District Councils labelled)

To give an indication of the degree of local heterogeneity, Figure 2 shows the same results but for the area immediately surrounding Belfast city centre. Here we can see the concentration of estimated income deprivation in particular areas although we should be aware that the visually dominant Sydenham_1, Duncairn_1 and Island_1 areas have relatively few households for their apparent size, dominated as they are by industrial, dock and airport land-use.

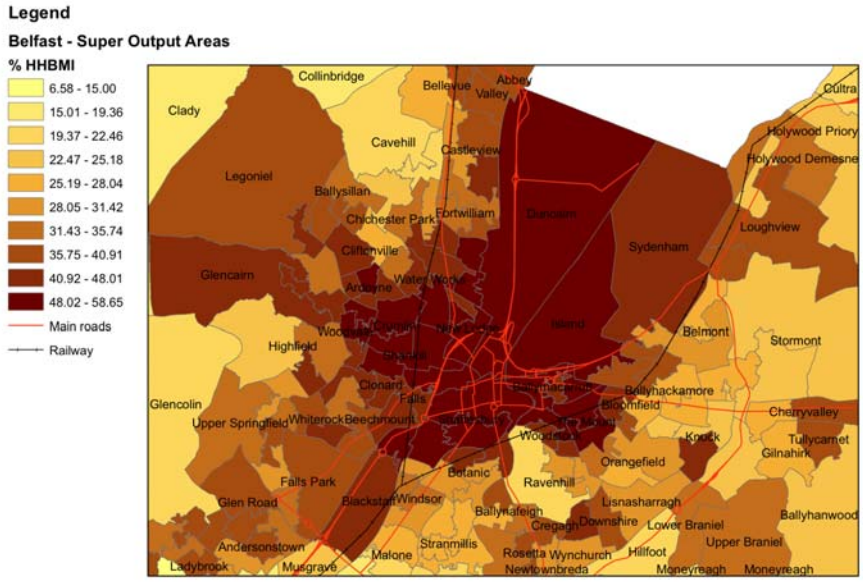

Figure 2: Spatial Distribution of simulated %HHBMI at SOA level (Belfast, electoral wards labelled)

3.6 Validation

In order to test the validity of the estimated distributions of HHBMI we can make three kinds of comparisons:

- Compare estimated results with initial source survey results (i.e. at regional or country level) to check internal validity and that the process accurately re-creates inter-regional or inter-country variation. In this case we use the FRS;
- Compare estimated constraint counts with initial constraint counts to check internal validity. This is the analysis of Total Absolute Error (TAE) discussed in Ballas et al (Ballas, Clarke et al. 1999; Ballas, Dorling et al. 2006);
- Compare estimated SOA level results with other known small area estimates. In this case we use the SOA level Northern Ireland NIMDM 2005 income domain scores and the Experian 2005 median household income estimates as a comparison.

3.6.1 Comparison with source and other survey data

Table 13 shows the HHBMI indicator as calculated from the FRS 2004-5 and as estimated from the spatial microsimulation process. In general we would expect the microsimulation result to lie within the 95% confidence interval of the survey estimate and as can be seen, the spatially microsimulated estimate using all weighted households is well within the overall Northern Ireland HHBMI confidence intervals for 2004-5.

Table 13: Comparison of simulated mean regional HHBMI results with source (FRS 2004-5) results

	FRS 2004-5 (Source)			Sim 2004-5
	%HHBMI	Lower 95% CI	Upper 95% CI	%HHBMI
Northern Ireland	31.724	29.643	33.804	30.382

3.6.2 Comparison of initial and estimated constraint counts

By entering the constraint counts as variables to be estimated it is possible to compare the initial 'true' constraint household counts with the estimated counts following the spatial microsimulation procedure. Whilst the objective is to minimise the difference between the 'true' and estimated counts (the TAE), it is not yet clear in the literature what values of error are acceptable. However the analysis provides a useful indicator of potential problems with the results.

Table 14 shows these results and we can see immediately how the order of the variables entered into the process is vital. It can also be seen that the mean difference between the actual (Census) and simulated constraint values is very low with the most errors to be seen in the age constraint - the largest error being 8.7 households in the Age (75-84) category. As the mean proportion column shows these errors are extremely small when compared to the overall SOA household numbers however in future work it may be worth experimenting with the order of the variables to try to reduce the error further.

Table 14: TAE for all SOAs as mean household absolute difference and absolute difference as a proportion of the SOA level household count.

Variable	Mean absolute difference	Mean absolute proportion (%)
age 16-24	0.594	0.08%
age 25 to 29	1.206	0.17%
age 30 to 44	5.897	0.84%
age 45 to 59	5.394	0.77%
age 60 to 64	1.728	0.24%
age 65 to 74	4.072	0.59%
age 75 to 84	8.672	1.24%
age 85 or over	4.561	0.65%
Rooms (1)	0.031	0.00%
Rooms (2)	0.249	0.04%
Rooms (3)	1.696	0.25%
Rooms (4+)	1.915	0.28%
Persons (1)	1.964	0.29%
Persons (2)	2.189	0.32%
Persons (3)	1.609	0.24%
Persons (4)	1.907	0.28%
Persons (5+)	1.478	0.22%
Employment (NSSEC 1)	1.087	0.16%
Employment (NSSEC 2)	0.778	0.11%
Employment (NSSEC 3)	0.828	0.12%
Employment (Retired)	0.862	0.13%
Employment (Inactive)	2.146	0.31%
Composition (couples)	1.678	0.27%
Composition (lone parents)	0.626	0.10%
Composition (single persons)	1.964	0.29%
Composition (other)	3.394	0.54%
Number of earners (0)	0.000	0.00%
Number of earners (1)	0.000	0.00%
Number of earners (2)	0.000	0.00%
Number of earners (3+)	0.000	0.00%

3.6.3 Comparison with other known small area estimates

We have only two sources of small area comparison – the NIMDM 2005 income domain score[5] and the Experian 2005 median household income estimates. In the case of the latter the spatial microsimulation process was repeated using gross household income and, as a result, a different set of constraints[6].

Table 15: Rank order correlations comparing simulated results with NIMDM 2005 income domain score and Experian 2005 median household income at SOA level.

Region	Sim HHBMI 2005 vs NIMDM 2005 income domain score (SOAs, Spearman)	Sim Median Income 2005 vs Experian 2005 Median Income (SOAs, Spearman)
Northern Ireland	0.754 (p<0.001)	0.821 (p<0.001)

[5] As this is defined as the percentage of the relevant population in receipt of selected benefits, it is still not entirely comparable.

[6] In decreasing order of importance: Number of earners, Employment status, Number of persons, Number of children, Accommodation.

Figure 3 shows the fit between the NIMDM 2005 income domain score and the simulated HHBMI at SOA level whilst Table 15 shows the rank order correlations. As we would expect there is a strong rank order correlation between HHBMI and the NIMDM income domain score.

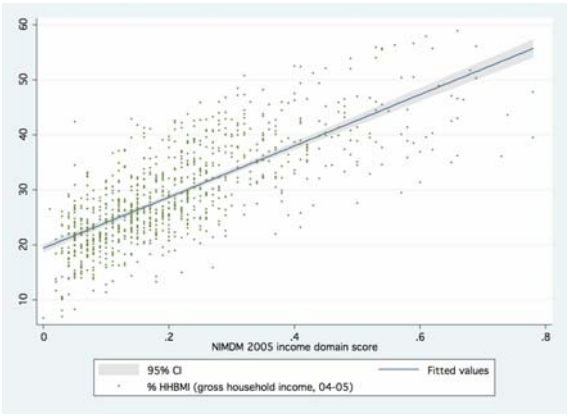

Figure 3: NIMDM 2005 income domain score vs Simulated HHBMI (All Northern Ireland SOAs)

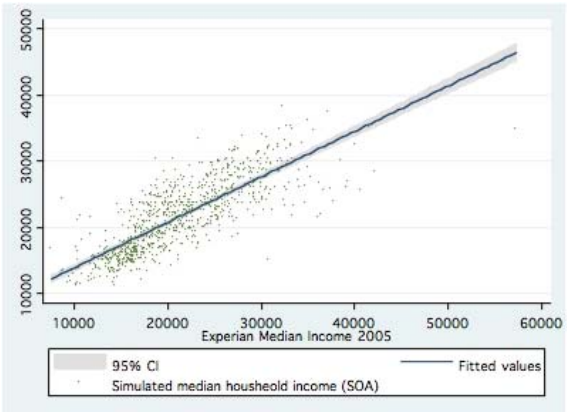

Figure 4: Experian 2005 median income vs Simulated Median Income 2005 (All NI SOAs)

Figure 4 shows the fit between the simulated median income at SOA level and the Experian 2005 estimated median income. Again we can see a reasonable degree of fit which is supported by the strong rank order correlation (see Table 15) although, as with the earlier English results the Experian data shows a much larger range (£7,574-£57,267) compared to the simulated median (£11,336-£38,532) although as Figure 4 illustrates this is mostly due to a single Experian outlier at the top end of the distribution. The goodness of fit to the SOA level Experian income estimates also suggests that the Census 2001 could be used as an effective surrogate for a non-existent Census 2005[7].

[7] The extent to which the Experian income model relies on Census 2001 data as opposed to their own updated small area data is unclear from the documentation. If it is considerable then these correlation results are to be expected.

4 Phase II: Equivalised net household income before housing costs

In this section of the report we describe the results from the use of the equivalised net household income before housing costs as the basis for the HHBMI indicator and using the pooled[8] 2003-4 and 2004-5 FRS data[9]. Since the constraint selection and spatial microsimulation methods are identical to those used above we do not re-describe these in detail but simply report the appropriate results and decision points.

4.1 Equivalisation and median calculations

The modified OECD equivalisation scale (Table 16) was used to control for household composition and to produce an equivalised measure of household income before housing costs.

Table 16: Modified OECD equivalisation scale (see Table A2 1.0, Appendix 2, HBAI07)

Composition	Scale value (BHC)
1st Adult	0.67
Subsequent adults	0.33
Children aged < 14	0.20
Children aged 14-18	0.33

Thus the equivalised income indicator was calculated by dividing the net BHC income by the aggregated household composition based weight.

This produces the final income variable – equivalised income before housing costs. These are then used as the basis for the calculations of the UK BHC median and thence the allocation of households to the two indicator groups – above or below 60% of the relevant UK median.

4.2 Identification of constraint variables

As before stepwise logistic regression is used to identify the most useful constraint variables from amongst the set of candidates. Table 2 shows the candidate constraint variables, Table 32 in Annex A.2 shows the results for a logistic regression model entering all constraints[10] and Table 17 summarises the results for the separate and pooled FRS surveys whilst Table 18 shows the contribution of each constraint to the overall R-squared score when added incrementally in the order suggested. These results suggest that we can be justified in pooling the 2003-4 and 2004-5 FRS data because the most powerful predictors of each indicator at the household level are essentially identical although it is interesting to note that with the larger pooled sample (03-05), the number of children becomes significant even though income has been equivalised but that HRP age (03-04), number of persons and HRP ethnicity (04-05) are not significant and therefore may be spurious results in those years due to small sample sizes. It is also noticeable that neither community background nor rural/urban location contribute significantly to the models although they may of course have indirect effects on those constraints that are significant. The notably lower R-squared scores compared to that for gross income (Table 3) suggest that the use of net income (HHBAI definition) and the conversion to equivalised income reduces the ability of the constraint variables to predict income deprivation because it removes much of the variation that is predicted by, for example, the number of persons, number of children and household composition. In the case of the 2003-4 data we can see that these three variables are not statistically significant although number of persons and composition are in 2004-5 albeit with very small effects. The borderline effects of these variables are further suggested by the selection of number of children and composition in the pooled 2003-5 sample.

[8] Note that the 2003-4 data are not inflated; the two data sets are simply pooled.

[9] For comparison the rounded UK median equivalised net household income was £16,530 in 2003-4, £17,370 in 2004-05) and £16,950 for the pooled 2003-05 sample. The threshold used for the 2003-05 sample was 60% of the median of the pooled sample (£16,950).

[10] Including the rural/urban indicator for the 2004-5 FRS sample supplied by NISRA.

Table 17: Significant constraints (in decreasing order of explanatory power)

	2003-4	2004-5	2003-05 pooled
	Number of Earners	Number of Earners	Employment Status
	Employment Status	Employment Status	Number of Earners
	Presence of LLI	Tenure	Presence of LLI
		Presence of LLI	Number of children
		Number of persons	Composition
		Composition	Tenure
		HRP non-white	
Pseudo R sq	0.179	0.180	0.178
N	1895	1913	3812

Table 18: Overall and additional pseudo R-squared values (constraints added incrementally)

2003-4		2004-5		2003-5 pooled	
Overall pseudo R sq	Additional Pseudo R sq	Overall pseudo R sq	Additional Pseudo R sq	Overall pseudo R sq	Additional Pseudo R sq
Number of Earners 0.1236		Number of Earners 0.1213		Employment Status 0.1266	
Employment Status 0.1746	5.10%	Employment Status 0.1532	3.19%	Number of Earners 0.1629	3.63%
Presence of LLI 0.1787	0.41%	Tenure 0.1601	0.69%	Presence of LLI 0.1673	0.44%
		Presence of LLI 0.1642	0.41%	Number of children 0.1703	0.30%
		Number of persons 0.1686	0.44%	Composition 0.1751	0.48%
		Composition 0.1770	0.84%	Tenure 0.1775	0.24%
		HRP non-white 0.1799	0.29%		

As previously discussed these constraints are then used at the small area (SOA) level to iteratively re-weight the FRS to fit each SOA and so produce an estimate of the %HHBMI for each SOA for each indicator and also an estimate of the proportion of children living in 'poor' households in each SOA for each indicator. Whilst results for 2003-4 and 2004-5 have been generated separately we report only those for the pooled 2003-5 data using the constraints identified above.

4.3 Equivalised net income results

The results of this spatial microsimulation process are summarised in Table 19 to Table 21 and the complete data is supplied in the associated excel workbook.

Table 19 to Table 21 show the 5 'poorest' and 5 'wealthiest' SOAs in Northern Ireland according to the equivalised net income BHC indicator for each year of data and the pooled 2003-5 FRS respectively. In all cases the SOAs with the highest %HHBMI are also in the 9th (i.e. most deprived) decile of the income domain score (not shown) and in all but two cases those with the lowest %HHBMI score are in the least deprived income domain score decile. The two exceptions are the 2003-4 and 2003-5 results for Beechill_1 (95II02S1), which is in the 3rd least deprived income domain decile.

Table 19: HHBMI of the 5 poorest and wealthiest SOAs in Northern Ireland (2003-4, equivalised BHC)

	SOA code	SOA name	Local Government District	Equivalised %HHBMI (BHC)
1	95WW20S3	Mallusk_3	Newtownabbey	6.84
2	95AA01S1	Aldergrove_1	Antrim	7.29
3	95II03S1	Cairnshill_1	Castlereagh	7.42
4	95WW19S2	Jordanstown_2	Newtownabbey	7.58
5	95II02S1	Beechill_1	Castlereagh	7.87
1	95SS06S2	Collin Glen_2	Lisburn	34.66
2	95MM12S2	Crevagh_2	Derry	32.95
3	95MM10S1	Creggan Central_1	Derry	31.76
4	95MM25S3	Shantallow West_3	Derry	30.02
5	95GG35S2	New Lodge_2	Belfast	29.57

Table 20: HHBMI of the 5 poorest and wealthiest SOAs in Northern Ireland (2004-5, equivalised BHC)

	SOA code	SOA name	Local Government District	Equivalised %HHBMI (BHC)
1	95WW20S3	Mallusk_3	Newtownabbey	7.84
2	95WW19S2	Jordanstown_2	Newtownabbey	8.25
3	95XX01S2	Ballycrochan_2	North Down	8.61
4	95II03S1	Cairnshill_1	Castlereagh	8.96
5	95SS02S2	Ballymacash_2	Lisburn	9.31
1	95SS06S2	Collin Glen_2	Lisburn	31.28
2	95MM12S2	Crevagh_2	Derry	29.80
3	95MM25S3	Shantallow West_3	Derry	27.91
4	95MM10S1	Creggan Central_1	Derry	27.44
5	95MM25S1	Shantallow West_1	Derry	26.48

Table 21: HHBMI of the 5 poorest and wealthiest SOAs in Northern Ireland (2003-5, equivalised BHC)

	SOA code	SOA name	Local Government District	Equivalised %HHBMI (BHC)
1	95WW20S3	Mallusk_3	Newtownabbey	6.97
2	95WW19S2	Jordanstown_2	Newtownabbey	7.83
3	95II03S1	Cairnshill_1	Castlereagh	8.18
4	95XX01S2	Ballycrochan_2	North Down	8.33
5	95II02S1	Beechill_1	Castlereagh	8.62
1	95SS06S2	Collin Glen_2	Lisburn	35.48
2	95MM12S2	Crevagh_2	Derry	33.20
3	95MM10S1	Creggan Central_1	Derry	30.93
4	95MM25S3	Shantallow West_3	Derry	30.86
5	95MM25S1	Shantallow West_1	Derry	29.32

These tables suggest that we can be justified in pooling the 2003-4 and 2004-5 data because the SOAs at the upper and lower ends of the deprivation dimensions are broadly similar. This is confirmed by Spearman rank correlations between the three different versions (Table 22) and also by Figure 5.

Table 22: Rank correlations of indicator versions

	FRS 2003-4	FRS2004-5	FRS 2003-5 pooled
FRS 2004-5	0.928		
FRS 2003-5 pooled	0.970	0.969	
Original gross unequivalised 2004-5 FRS	0.875	0.755	0.821

We can see that whilst the original unequivalised gross income indicator shows a relatively good correlation with the new equivalised net income indicator the scatter charts in Figure 5 demonstrate a non-linear relationship and thus rather different distributions.

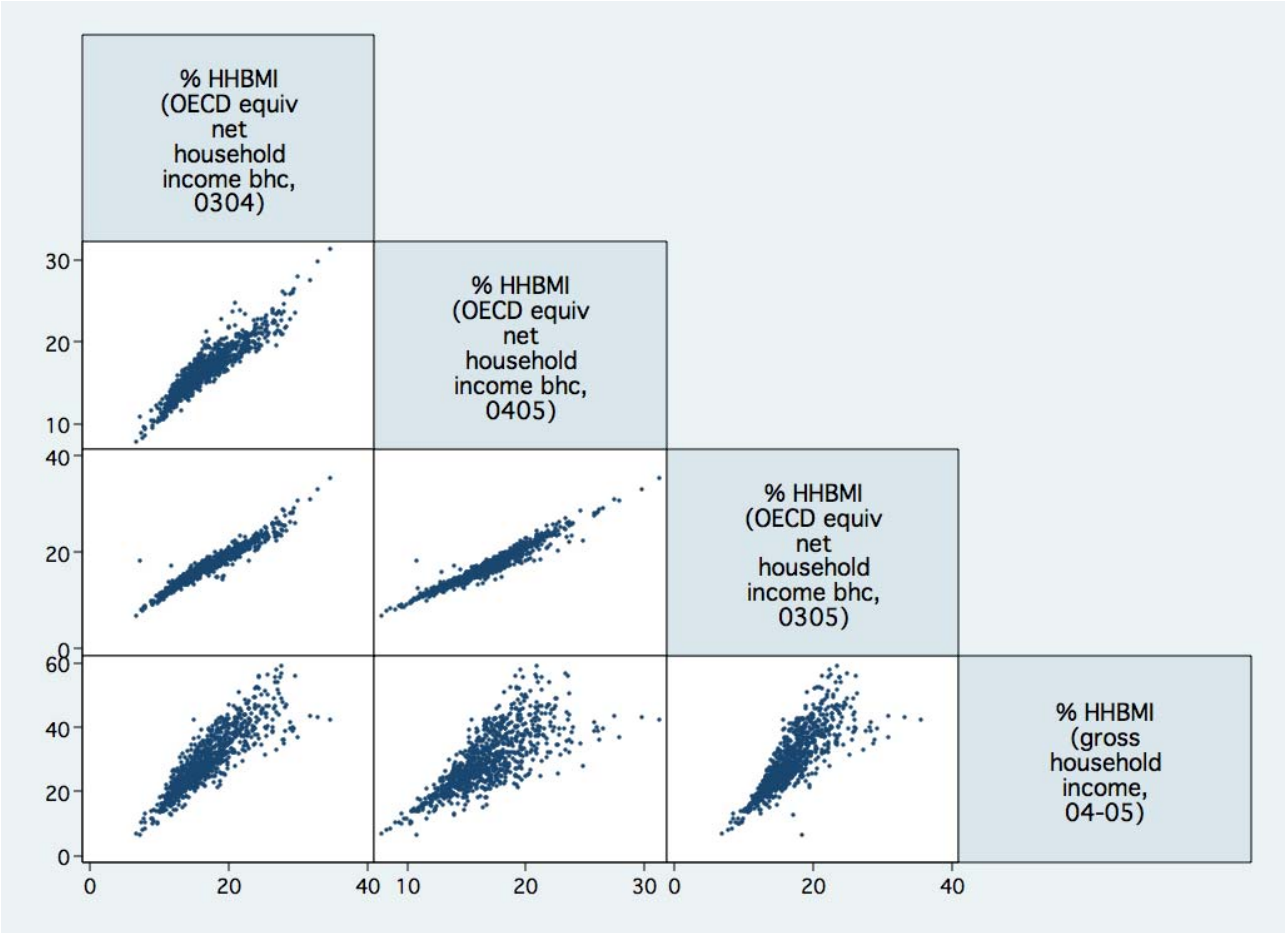

Figure 5: Comparison of BHC equivalised HHBMI indicators at SOA for Northern Ireland (2003-4, 2004-5 and 2003-05 pooled) with the initial unequivalised gross income results

4.4 Equivalised net income results validation

As before a number of validation procedures can be implemented and in this section we report the comparison with the source FRS data, comparisons with the NIMDM 2005 and the TAE.

4.4.1 Comparison with source and other survey data

Table 23 shows the HHBMI indicators (and 95% confidence interval) as calculated from the relevant source FRS data and as estimated from the spatial microsimulation process at the Northern Ireland level. Overall there appears to be a tendency to slightly underestimate HHBMI compared to the source FRS results. In general we would expect the microsimulation result to lie within the 95% confidence interval of the survey estimate and as can be seen none of the spatially microsimulated estimates lie inside these boundaries. Interestingly the pooled 2003-5 FRS %HHBMI estimates are marginally closer to the observed results than the 2003-4 or 2004-5 estimates suggesting that the size of the FRS sample used effects the robustness of the results.

Table 23: Comparison of simulated equivalised net household income BHC based %HHBMI results with source results

	FRS (Source)			Spatial simulation
	%HHBMI	Lower 95% CI	Upper 95 % CI	%HHBMI
FRS 2003-4	19.457	17.684	21.231	17.106
FRS 2004-5	19.263	17.500	21.026	16.716
FRS 2003-5 pooled	19.256	18.009	20.503	17.040

Table note:
Simulation estimate based on all weighted households in Northern Ireland

The extent to which this is an issue for the reliability of the small area results depends on the extent to which the FRS surveys are representative samples, the extent to which there has been social change since the 2001 Census (and particularly between 2003-4 and 2004-5) and the degree of fit error produced by the

spatial microsimulation method. We deal with the latter in the next section but it is worth re-iterating that because the spatial microsimulation method is in effect a multi-dimensional re-weighting scheme based on Census counts, it may be that it produces more robust income estimates than the source FRS when using FRS data collected close to Census years. Conversely, if FRS data is used which has been collected further from Census years and if there has also been significant socio-demographic change that influences (relative) income distributions then re-weighting to fit Census distributions is likely to lead to erroneous estimates. Of course this has little to say about the reliability of small area estimates although the same principles apply.

4.4.2 Comparison of initial and estimated constraint counts

Table 24 to Table 26 show the TAE results for the new indicators. As before the mean difference between the actual (Census) and simulated constraint values is very low with the most errors to be seen in the Employment status (retired) constraint in the 2003-4 and 2004-5 data and Composition (other) in the 2003-5 data. As the mean proportion columns show these errors are small when compared to the overall SOA household numbers.

Table 24: TAE for all SOAs as mean household count absolute difference and the same figure as a proportion of the SOA level households (2003-4).

Variable	Mean difference	Mean proportion
Presence of LLI (no)	1.247	0.19%
Presence of LLI (yes)	0.873	0.13%
NSSEC 1	0.768	0.11%
NSSEC 2	0.600	0.09%
NSSEC 3	0.590	0.08%
Inactive	0.595	0.09%
Retired	1.363	0.20%
Number of earners (0)	0.000	0.00%
Number of earners (1)	0.000	0.00%
Number of earners (2)	0.000	0.00%
Number of earners (3+)	0.000	0.00%

Table 25: TAE for all SOAs as mean household count absolute difference and the same figure as a proportion of the SOA level households (2004-5).

Variable	Mean difference	Mean proportion (%)
HRP non-white (no)	2.534	0.38%
HRP non-white (yes)	0.014	0.00%
Composition (couples)	2.321	0.37%
Composition (lone parents)	1.240	0.18%
Composition (single persons)	1.095	0.16%
Composition (other)	4.229	0.65%
Persons (1)	1.095	0.16%
Persons (2)	0.902	0.14%
Persons (3)	1.075	0.16%
Persons (4)	0.568	0.09%
Persons (5+)	1.407	0.22%
Presence of LLI (no)	1.278	0.20%
Presence of LLI (yes)	1.156	0.18%
Tenure: Own	2.194	0.33%
Tenure: social rent	0.705	0.10%
Tenure: private rent	0.233	0.03%
NSSEC 1	2.604	0.37%
NSSEC 2	2.620	0.38%
NSSEC 3	2.595	0.36%
Inactive	2.542	0.36%
Retired	5.237	0.75%

Number of earners (0)	0.000	0.00%
Number of earners (1)	0.000	0.00%
Number of earners (2)	0.000	0.00%
Number of earners (3+)	0.000	0.00%

Table 26: TAE for all SOAs as mean household count absolute difference and the same figure as a proportion of the SOA level households (2003-5 pooled).

Variable	Mean difference	Mean proportion (%)
Tenure: Own	1.504	0.23%
Tenure: social rent	0.423	0.06%
Tenure: private rent	0.227	0.03%
Composition (couples)	1.739	0.28%
Composition (lone parents)	1.112	0.16%
Composition (single persons)	1.431	0.21%
Composition (other)	3.701	0.58%
Children: 0	0.383	0.06%
Children: 1	0.310	0.05%
Children: 2+	0.501	0.08%
Presence of LLI (no)	1.294	0.20%
Presence of LLI (yes)	1.200	0.18%
Number of earners (0)	3.306	0.47%
Number of earners (1)	1.253	0.17%
Number of earners (2)	1.526	0.22%
Number of earners (3+)	0.527	0.08%
NSSEC 1	0.003	0.00%
NSSEC 2	0.002	0.00%
NSSEC 3	0.002	0.00%
Inactive	0.000	0.00%
Retired	0.000	0.00%

Table note:
Non-zero TAE in the employment status constraint is due to non-rounding of the original imputed NS-SEC values (see Section 3.3)

4.4.3 Comparison with other known small area estimates

As before we compare the results with the income domain score from the NIMDM 2005 using a rank order correlation and also visually. In general all of the new indicators show a stronger correlation with the NIMDM 2005 income domain score than the original unequivalised gross household income indicator. Given that the NIMDM 2005 income domain is based on 2003 benefits data the higher correlation using the 2003-4 data is to be expected.

Table 27: Rank order correlations comparing simulated results with NIMDM 2005 income domain score at SOA level.

Sample	%HHBMI vs. NIMDM 2005 income domain score (SOAs, Spearman)
FRS 2003-4	0.915
FRS 2004-5	0.899
FRS 2003-5 pooled	0.912
Original Gross Income indicator 2004-05 (for comparison)	0.754

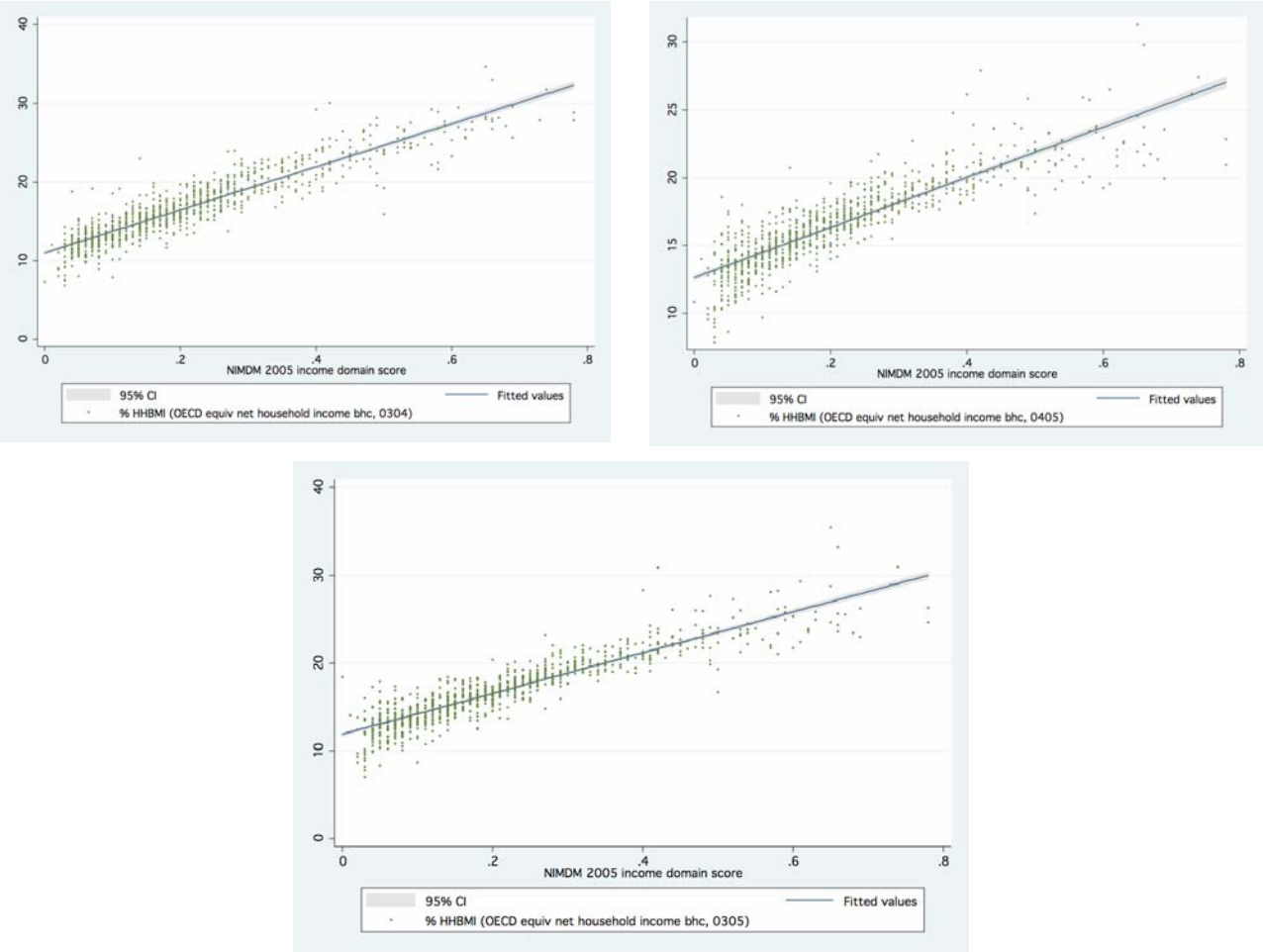

Figure 6: NIMDM 2005 income domain score vs. Simulated HHBMI indicators (All Northern Ireland SOAs)

4.5 Small area inequality results

Whilst the overall level of income deprivation at SOA level is one policy interest another may be the degree of income inequality within SOAs given empirical and theoretical suggestions that local inequalities can be the source of community tensions. A unique feature of the spatial microsimulation method is that it can create microdata for each SOA in the form of weighted FRS cases. Once generated this data can then be used to calculate measures of inequality *within* each SOA.

In order to explore this possibility we used the methods described above to model equivalised net household income before housing costs using the 2003-05 pooled FRS sample. The same process of multivariate regression was used to identify the number of rooms, composition, accommodation, number of children, number of earners and employment status as significant constraints (see Annex A.3).

Following spatial microsimulation of household income the weighted microdata was used to calculate the Gini coefficient for each SOA[11]. As used in standard economic analysis the Gini coefficient represents the inequality of income distribution within a population. The closer the coefficient is to 0 the more equal the distribution whilst the closer to 1 (100%) the more unequal. As far as we are aware this measure of inequality has never been used at the small area level.

[11] This was done by running the stata command 'ineqdeco' (Jenkins, S. (2006). ineqdeco/0: Inequality indices, with optional decomposition by subgroup. Colchester, Essex, University of Essex.) on the weighted cases in each SOA separately. It should be noted that ineqdeco ignores cases where values are less than or equal to 0.

Table 28: Comparison of source and spatial microsimulation results for equivalised net household income and Gini coeffient

		FRS 2003-5 (Source)		Spatial simulation
		Lower 95% CI	Upper 95% CI	
Mean equivalised net household income	£18,312	£17,950	£18,673	£18,922
Gini coefficient	0.287	.280	.296	0.291

Notes:
Confidence intervals for Gini coefficient calculated using the bias corrected bootstrapping method (stata command 'ineqerr')

Overall the estimated Gini for Northern Ireland using all weighted households was 0.291 compared to 0.287 for the FRS Northern Ireland sample and thus fell within the 95% confidence limits for the FRS source although the estimated mean equivalised net household income lay just outside these limits (see Table 28). At the SOA level the minimum Gini was 0.227 and the maximum 0.350. Figure 7 shows the distribution of the coefficient across the SOAs and illustrates the concentration towards the upper end of the range.

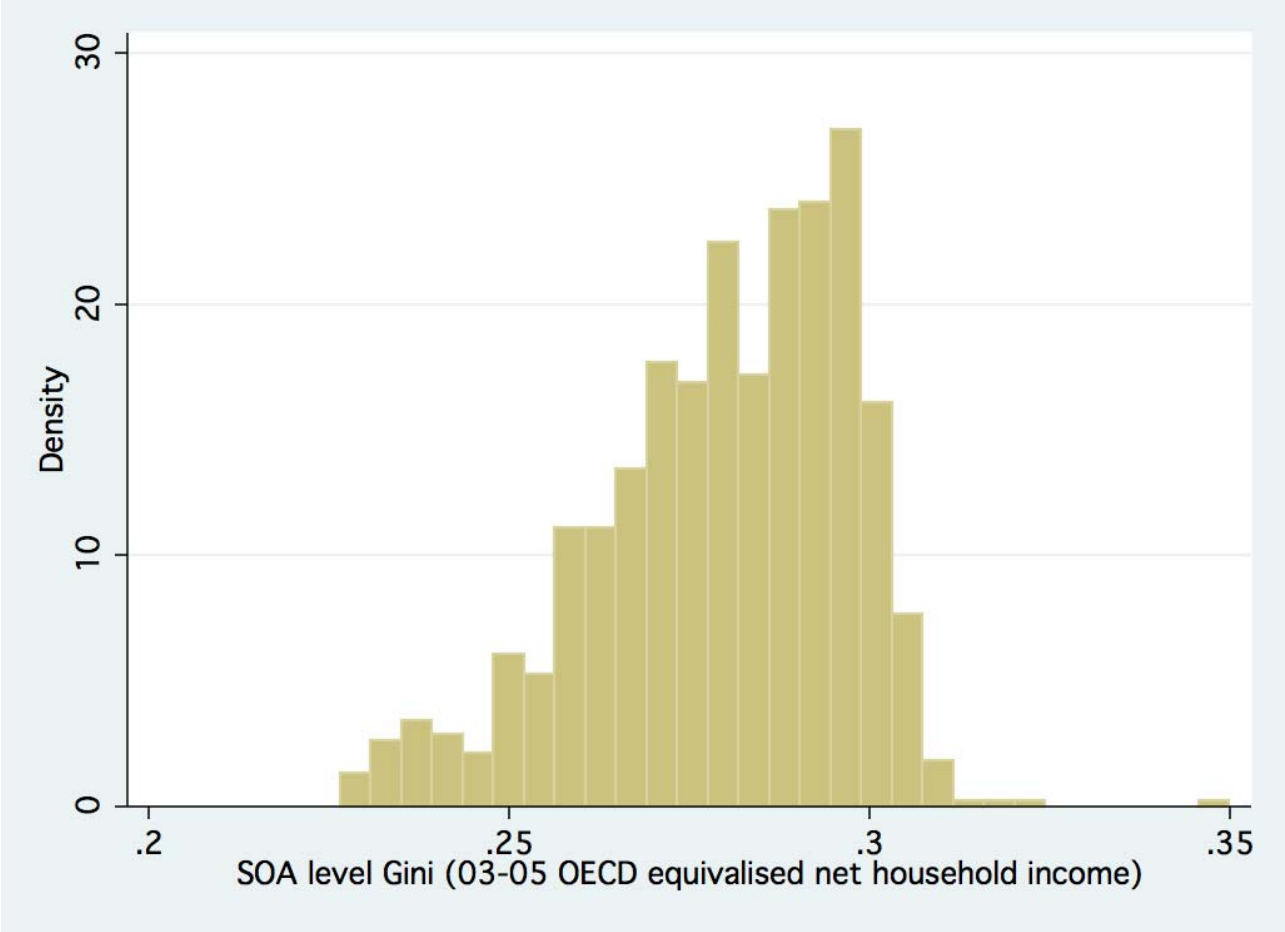

Figure 7: SOA level Gini coefficient distribution

Table 29 shows the five SOAs with the lowest and highest Gini coefficients respectively and, for comparison, their estimated equivalised net income based %HHBMI. Those SOAs with lowest income inequality (smallest Gini) are all in Belfast and perhaps therefore somewhat homogeneous urban SOAs whilst those with highest Gini coefficients are more distributed and perhaps more heterogeneous (see also Figure 8 and Figure 9).

Table 29: The five highest and lowest SOAs according to their estimated Gini coefficient

	SOA Code	SOA Name	Local Government District	Gini	%HHBMI (equiv, BHC)
1	95GG12S1	Botanic_1	Belfast	0.350	14.81
2	95GG12S4	Botanic_4	Belfast	0.322	18.37
3	95WW24S1	Valley_1	Newtownabbey	0.317	18.55
4	95VV08W1	Creggan	Newry and Mourne	0.312	22.23
5	95KK05W1	Killycolpy	Cookstown	0.311	20.41
1	95GG19S1	Crumlin_1	Belfast	0.227	21.25
2	95GG18S1	Clonard_1	Belfast	0.229	23.94
3	95GG10S1	Blackstaff_1	Belfast	0.230	19.81
4	95GG02S1	Ardoyne_1	Belfast	0.230	26.10
5	95GG51S2	Woodvale_2	Belfast	0.230	22.91

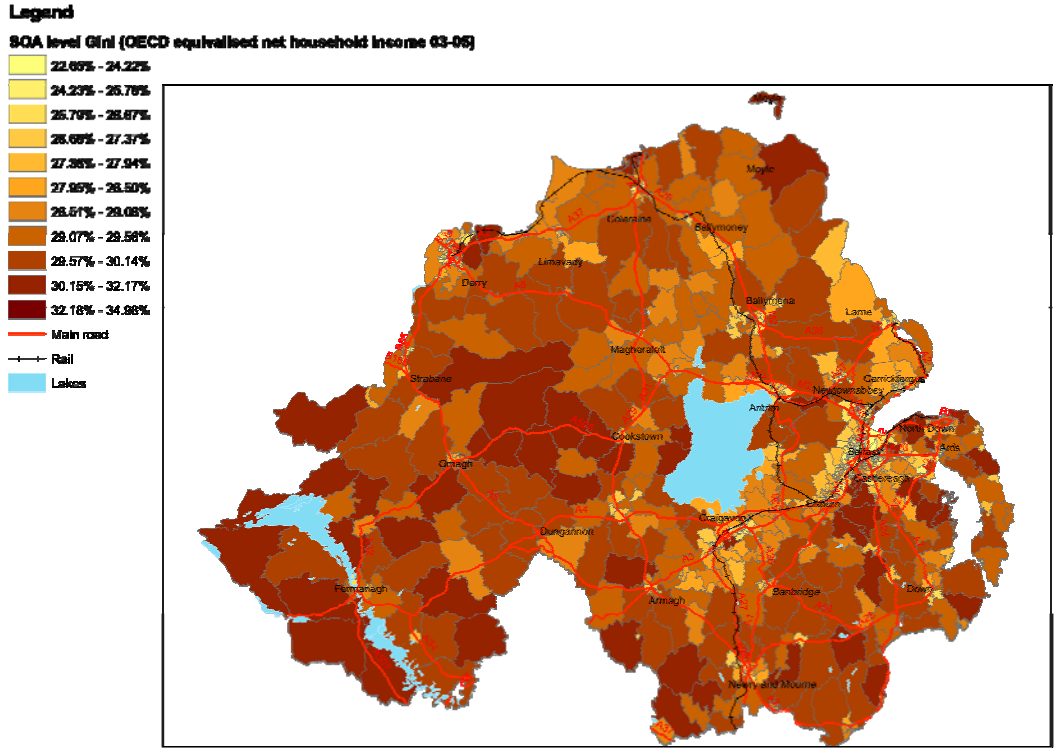

Figure 8: Spatial distributions of SOA level Gini coefficient (Northern Ireland, Local Government Districts labelled)

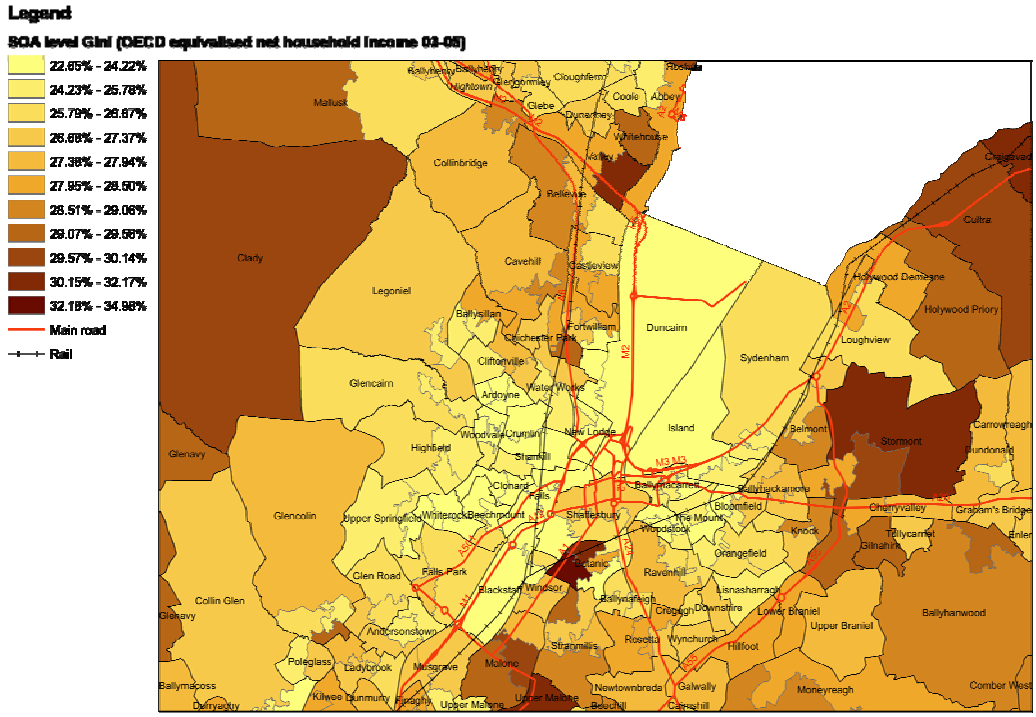

Figure 9: Spatial distributions of SOA level Gini coefficient (Belfast, Wards labelled)

As Figure 10 shows there is a negative although not very strong relationship between the SOA Gini and the %HHBMI (Spearman's rho = -0.323). Thus those SOAs with lower rates of income deprivation tend also to be those with slightly more income inequality – thus poorer SOAs tend to be more evenly poor.

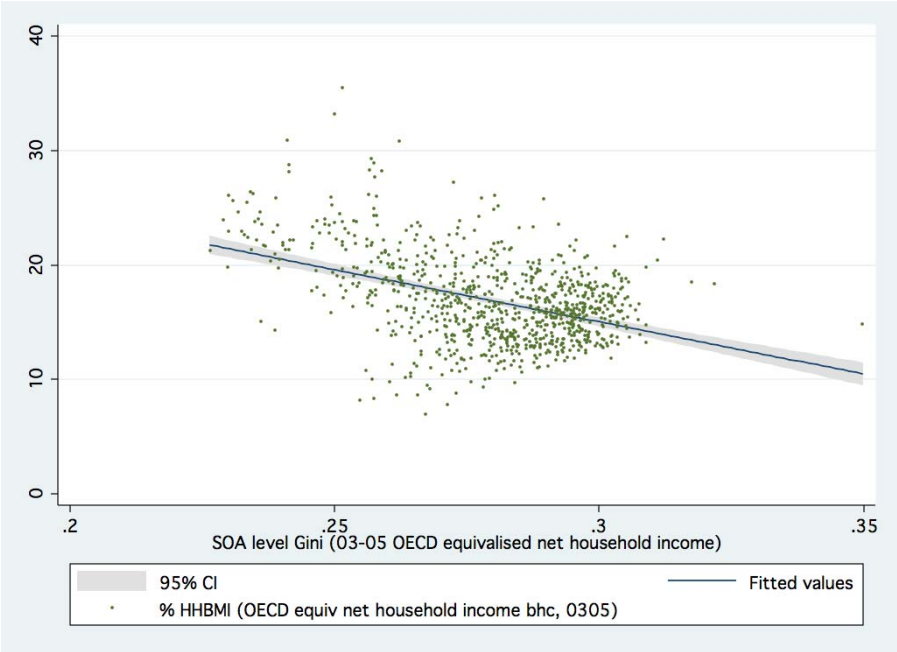

Figure 10: Relationship between %HHBMI (equivalised, BHC) and SOA level Gini coefficient

5 Implications and Recommendations

Overall the results of this preliminary work are encouraging.

The initial results (Phase I) for 2004-5 using a combination of Census 2001 and FRS 2004-5 to model the percentage of households whose gross household income was below 60% of the UK median provide a synthetic household dataset which is able to replicate the Northern Ireland FRS overall %HHBMI and which also produce a good fit to the NIMDM 2005 income domain score at the SOA level. In addition, the simulated median income shows a good fit to the Experian 2005 estimated median household income data again at the SOA level.

The subsequent results (Phase II) for 2003-4, 2004-5 and 2003-5 using the revised equivalised net income indicator confirm this optimistic conclusion. All three show strong correlations with the NIMDM 2005 income domain score although they are less well able to reproduce the aggregate indicator results. The lower R-squared values produced by the stepwise regression models (compare Table 3 and Table 17) suggest that this may be because the constraints are less well able to predict this indicator at the micro level perhaps because of the conflicting effects of household size since the more people there are in a household, potentially the more earners there might be but the lower the equivalised income.

Finally, the production of small-area inequality measures (Gini coefficients) proved successful in that plausible estimates were generated indicating the level of income inequality within each SOA. However, until other estimates can be generated from real or modelled micro-data from other sources the validity of these estimates remains uncertain.

It has also been suggested that an alternative approach would be to apply the regression coefficients derived from the stepwise logistic models to the Census households records themselves if such access were made possible. The extent to which this would produce different results is not currently clear but it would have the advantage of enabling the construction of statistical confidence intervals to the results. One drawback of this approach would be that the relationship between the constraint variables and income as defined by the regression coefficients would be assumed to be the same in each small area. This is, to some extent also true of the current spatial microsimulation approach since it may be that some constraints are better predictors of income in some small areas than others as has been suggested by the use of regional models within England (Anderson 2007). With the available FRS sample size of course there is no way to analyse the extent to which this is the case.

Overall, the results suggest that the spatial microsimulation method will be of value in estimating spatial distributions of income deprivation indicators when contemporaneous Census and survey data is available as will next be the case in 2011. It may also be of value in inter-censual years where a relatively small amount of SOA level socio-economic change can be assumed.

However, as should now be clear the main issue for producing spatially microsimulated results in inter-censual years that are some time after the Census data has been collected is that SOA level socio-economic change is likely to have occurred and the longer the time-frame the more change there may have been. We therefore recommend three further research activities.

- Comparison of the SOA 2005 simulated income estimates (Census 2001 + FRS) with a new NIMDM income domain score should it become available to investigate the plausibility of using decennial Census data as a basis for inter-censual estimates;
- Exploration of methods to produce robust historical linked Census counts at the SOA level as a basis for the projection of small area Census counts to inter-censual years;
- A review of the availability and reliability of updated constraint variables from other sources including commercial providers in inter-censual years.
- An analysis of the effects of potential under/over-reporting or non-response biases in the FRS.

Finally our experience with the spatial microsimulation method suggests that it could be used to estimate new indicators that are currently unavailable in the NIMDM but which could be of value to users. We therefore suggest a final additional research activity, to review potential new indicators that could be provided by the method. This should include systematic testing of the potential reliability of the indicators using the methods discussed in this report.

6 Acknowledgements

Census data for Northern Ireland were originally created and funded by the Northern Ireland Statistics and Research Agency (NISRA) and are distributed by the Census Dissemination Unit, MIMAS (University of Manchester). Output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

The FRS is collected and sponsored by the Department for Work and Pensions and distributed by the UK Data Archive, University of Essex, Colchester. FRS data are copyright and are reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

The NIMDM 2005 were constructed by the Social Disadvantage Research Centre at the Department of Social Policy and Social Research at the University of Oxford and distributed by the Northern Ireland Statistics and Research Agency (NISRA).

The Experian 2005 median income data were produced by Experian Ltd and are distributed by the Census Dissemination Unit, MIMAS (University of Manchester).

7 References

Anderson, B. (2007). Creating small area income estimates for England: spatial microsimulation modelling. A report to the Department of Communities and Local Government. Ipswich, Chimera, University of Essex.

Anderson, B. (2007). DCLG SimInc D1: Preliminary Results Report. DCLG SimInc Project Deliverable. Ipswich, Chimera, University of Essex.

Ballas, D., G. Clarke, et al. (2005). "SimBritain: A Spatial Microsimulation Approach to Population Dynamics." Population, Space and Place **11**: 13-34.

Ballas, D., G. Clarke, et al. (1999). Exploring Microsimulation Methodologies for the Estimation of Household Attributes. 4th International Conference on GeoComputation, Mary Washington College, Virginia, USA.

Ballas, D., D. Dorling, et al. (2006). Assessing the feasibility of producing small area income estimates: Phase I project report. Sheffield, Department of Geography, University of Sheffield.

Chin, S.-F. and A. Harding (2006). Regional Dimensions: Creating Synthetic Small-area Microdata and Spatial Microsimulation Models. NATSEM Technical Paper no. 33. Canberra, National Centre for Social and Economic Modelling, University of Canberra.

Eurostat. (2007). "Inequality of income distribution (S80/S20 income quintile share ratio), at-risk-of-poverty rate and at-persistent-risk-of-poverty rate." Key indicators on EU policy – Structural indicators – Social Cohesion – Living conditions Retrieved July 4th, 2008, from http://europa.eu.int/estatref/info/sdds/en/strind/socohe_di_base.htm.

Gordon, D. and P. Townsend (2000). Breadline Europe: The Measurement of Poverty. Bristol, Policy Press.

Jenkins, S. (2006). ineqdeco/0: Inequality indices, with optional decomposition by subgroup. Colchester, Essex, University of Essex.

Williamson, P. (2005). Income imputation for small areas. ESRC Grant Final Report (Award # H507255166). Liverpool, University of Liverpool.

Williamson, P. and D. Voas (2000). "Income Estimates for Small Areas: Lessons from the Census Rehearsal." BURISA **146**: 2-10.

Annex A Statistical Annex

A.1 2004-5 constraints testing for initial gross non-equivalised income indicator

Table 30: Results of logistic regression models predicting household income below 60% median income for FRS 2004-5 for all potential constraint variables for Northern Ireland

	Model 1	Model 2
Sex of HRP (Female = 1)	-0.034	-0.028
Presence of at least 1 person with limiting long-term illness	-0.165	-0.159
Number of earners	-1.380***	-1.368***
Number of persons	-0.574	-0.555
Number of rooms	-0.489*	-0.492*
Ethnicity of HRP (Non-white = 1)	1.418	1.657
Accommodation: Semi-Detached (Detached)	0.258	0.251
Terrace	0.164	0.174
Flat/maisonette	-0.328	-0.318
Other	-2.522	-2.499
HRP age 25 to 29 (16-24)	-0.690	-0.754
30 to 44	-0.838*	-0.903*
45 to 59	-0.648	-0.716
60 to 64	-0.923	-0.991*
65 to 74	0.008	0.024
75 to 84	-0.318	-0.193
85 or over	0.343	0.424
Composition: Single parent (Couple)	0.453	0.515
Single person	1.187**	1.244***
Other	-1.428*	-1.425*
Number of children: 1 (0)	0.748	0.682
2+	0.421	0.316
Tenure: Social rent (own)	0.259	0.259
Private rent	0.198	0.188
Employment status: NS-SEC 2 (NS-SEC 1)	2.051***	2.071***
NS-SEC 3	1.723***	1.780***
Inactive	2.671***	2.743***
Retired	1.555**	1.485**
Missing		2.611*
Community Background: Protestant and other Christian (Catholic)	0.023	0.023
Other	-1.257	-1.342
None	0.590	0.580
Missing		-0.352
Constant	-0.003	-0.026
Pseudo R-squared	0.467	0.470
N	1857	1926

Notes:
Values = regression coefficient
Model 1 = 'Missing/refused/unknown' for Employment status and Community background not included
Model 2 = 'Missing/refused/unknown' included
Category in parentheses = contrast
* = p < 0.05, ** = p < 0.01, *** = p < 0.001

Table 31: Results of stepwise logistic regression models predicting household income below 60% median income for FRS 2004-5 for all potential constraint variables for Northern Ireland

	Model 1	Model 2
Number of earners	-1.214***	-1.261***
Composition (Couple)		
Single parent	1.060***	0.964***
Single person	1.148***	1.229***
Other	-1.400*	-1.405*
Employment status (NS-SEC 1)		
NS-SEC2	2.076***	2.105***
NS-SEC3	1.798***	1.833***
Inactive	2.759***	2.846***
Retired	2.203***	1.687**
Missing		2.695**
Number of persons	-0.530***	-0.505***
Number of rooms	-0.457**	-0.437**
HRP age (16-24)		
25 to 29		-0.783
30 to 44		-0.921*
45 to 59		-0.847*
60 to 64		-1.152**
65 to 74		-0.182
75 to 84		-0.374
85 or over		0.277
Constant	-0.772	-0.073
Pseudo R-squared	0.453	0.462
N	1857	1926

Notes:
Values = regression coefficient
Model 1 = 'Missing/refused/unknown' for Employment status and Community background not included
Model 2 = 'Missing/refused/unknown' included
Category in parentheses = contrast
* = p < 0.05, ** = p < 0.01, *** = p < 0.001

A.2 2003-4 and 2004-5 constraints testing for equivalised income indicator

Table 32: Results of logistic regression models predicting equivalised net household income below 60% median income for all potential constraints for Northern Ireland

	FRS 2003-4	FRS 2004-5	FRS 2003-5
HRP Sex	0.080	-0.067	0.001
Presence of LLI	-0.509*	-0.598**	-0.510***
Number of earners	-1.317***	-1.545***	-1.362***
Number of persons	0.277	0.334	0.229
Number of rooms	-0.083	-0.170	-0.102
HRP non-white	-0.648	1.812*	0.991
Community background: Protestant and other Christian (Catholic)	-0.032	-0.159	-0.103
Other	-1.277	-0.538	-0.884
None	0.013	0.732*	0.238
Missing	-0.201	-0.274	-0.164
Accommodation: Semi-Detached	-0.518**	0.083	-0.265*
Terrace	-0.078	0.069	-0.081
Flat/maisonette	-0.165	0.303	-0.057
Other	-1.282	0.000	-1.353
HRP age 25 to 29 (16-24)	-0.686	-0.535	-0.535
30 to 44	0.184	-0.462	-0.057
45 to 59	0.209	-0.188	0.044
60 to 64	-0.103	-0.525	-0.230
65 to 74	0.967	-0.199	0.492
75 to 84	0.984	-0.114	0.478
85 or over	1.228	-0.226	0.543
Composition: Single parent (couple)	0.260	-0.492	-0.201
Single person	0.755**	0.635*	0.706***
Other	-0.162	-0.103	-0.233
Number of children: 1 (0)	0.248	0.567	0.505*
2+	-0.017	0.384	0.429
HRP Employment status: NS-SEC 2 (NS-SEC 1)	2.006***	1.433***	1.667***
NS-SEC3	1.511***	1.061**	1.202***
Inactive	2.099***	1.817***	1.935***
Retired	-0.053	0.401	0.174
Tenure: social rent (own)	-0.185	-0.647**	-0.424**
Private rent	0.195	-0.531*	-0.176
Urban/rural: Urban (BMUA)		0.057	
Rural		0.388*	
Constant	-2.171**	-1.180	-1.652**
Pseudo R-squared	0.202	0.193	0.186
N	1895	1913	3812

Notes:
Values = regression coefficient
'Missing/refused/unknown' for Community background included
Urban/rural indicator omitted from pooled 2003-5 sample as only available for 2004-5
Category in parentheses = contrast
* = p < 0.05, ** = p < 0.01, *** = p < 0.001

Table 33: Results of stepwise logistic regression models predicting equivalised net household income below 60% median income for all potential constraints for Northern Ireland

	FRS 2003-4		FRS 2004-5		FRS 2003-5
Number of earners	-1.353***	Number of earners	-1.432***	HRP: NS-SEC2 (NS-SEC1)	1.668***
HRP: NS-SEC2 (NS-SEC1)	2.091***	HRP: NS-SEC2 (NS-SEC1)	1.477***	HRP: NS-SEC3	1.187***
HRP: NS-SEC3	1.535***	HRP: NS-SEC3	1.140***	HRP: Inactive	1.907***
HRP: Inactive	1.966***	HRP: Inactive	1.890***	HRP: Retired	0.630*
HRP: Retired	0.564	HRP: Retired	0.557	Number of earners	-1.358***
Presence of LLI	-0.485**	Tenure: social rent (own)	-0.557**	Presence of LLI	-0.480***
		Tenure: Private rent	-0.405	I child (0)	0.685***
		Presence of LLI	-0.585**	2+ children	0.940***
		Number of persons	0.392***	Composition: Single parent (couple)	-0.443*
		Composition: Single parent (couple)	-0.303	Composition: Single person	0.472***
		Composition: Single person	0.684***	Composition: Other	0.040
		Composition: Other	-0.108	Tenure: social rent (own)	-0.397***
		HRP non-white	1.839*	Tenure: Private rent	-0.198
Constant	-1.704***		-1.992***		-1.992***
Pseudo R squared	0.179		0.180		0.179
N	1895	N	1913	N	3812

Notes:
Values = regression coefficient
Category in parentheses = contrast
* = p < 0.05, ** = p < 0.01, *** = p < 0.001

A.3 2003-5 constraints testing for equivalised net household income (for Gini calculation)

Table 34: Results of OLS regression models predicting equivalised net household for Northern Ireland (FRS 2003-05)

	b
HRP gender	-485
Presence of LLI	22
Number of earners	4532***
Number of persons	-368
Number of rooms	1248**
HRP non-white	-1764
HRP Community background: Protestant and other Christian (Catholic)	-4
Other	5025**
None	1687
Missing	290
Accommodation: Semi-Detached (Detached)	-1133**
Terrace	-2361***
Flat/maisonette	-470
Other	2525
HRP age 25 to 29 (16-24)	2017*
30 to 44	1396
45 to 59	691
60 to 64	1916
65 to 74	2899*
75 to 84	2515
85 or over	2773
Composition: Single parent (couple)	2278**
Single person	-1220
Other	-1611
Number of children: 1 child (0)	-3970***
2+ children	-5724***
HRP Employment status: NS-SEC2 (NS-SEC1)	-6054***
NS-SEC3	-7819***
Inactive	-8029***
Retired	-7620***
Tenure: social rent (own)	185
Private rent	-879
Constant	18667***
R-Squared	0.295
N	3812

Notes:
Values = regression coefficient
'Missing/refused/unknown' for Community background included
Category in parentheses = contrast
* = p < 0.05, ** = p < 0.01, *** = p < 0.001

Table 35: Results of Stepwise OLS regression models predicting equivalised net household for Northern Ireland (FRS 2003-05)

	b
Employment status (NS-SEC 1)	
NS-SEC2	-5999***
NS-SEC3	-7831***
Inactive	-8357***
Retired	-6433***
Number of earners	4345***
Number of children (0)	
1	-4357***
2+	-6466***
Accommodation (Detached)	
Semi-Detached	-1129**
Terrace	-2430***
Flat/maisonette	-618
Other	2197
Composition (Couple)	
Single parent	2101**
Single person	-1157**
Other	-2111***
Number of rooms	1207**
Constant	19788
R-squared	0.292
N	3812

Notes:
Values = regression coefficient
Category in parentheses = contrast
* = p < 0.05, ** = p < 0.01, *** = p < 0.001