CENTRE FOR RESEARCH IN ECONOMIC SOCIOLOGY AND INNOVATION
WORKING PAPER SERIES: ISSN 1759-2860





WELSH SMALL AREA ESTIMATES OF INCOME DEPRIVATION

CRESI WORKING PAPER NUMBER: 2009-04

Ben Anderson benander@essex.ac.uk

Department of Sociology
University of Essex

Abstract:

This paper describes the results of applying a spatial microsimulation approach to the estimation of small area (LSOA) income deprivation for Wales for 2005 using the Census 2001 and the FRS 2003/4 and 2004/5. The indicators used are the proportion of households in each LSOA whose equivalised household income was below 60% of the Welsh median equivalised household income in the appropriate year. The adjusted OECD scale was used to calculate equivalised income before (BHC) and after (AHC) housing costs and the results for these two variants are presented together with estimates of the % of children who are living in 'poor' households in each LSOA.

Keywords/tags:

Wales, small area, income deprivation, child poverty, spatial microsimulation

Citation:

Anderson, B. (2009) *Welsh Small Area Estimates of Income Deprivation*, Centre for Research in Economic Sociology and Innovation (CRESI) Working Paper 2009-04, University of Essex: Colchester

About CRESI:

Based in the UK's leading Sociology Department, the **Centre for Research in Economic Sociology and Innovation** (CRESI) is the first UK centre for research in economic sociology. With a clear focus on innovation, our research programmes highlight contemporary and historical processes of socio-economic transformation. You can read about our research and join our conversation.

Table of Contents

1 INTRODUCTION

As part of an investigation into potential alternative income domain indices that could contribute to future revisions of the income component of the Welsh Indices of Deprivation (WIMD) we have developed a preliminary spatial microsimulation model to estimate the spatial distributions of income and income deprivation for each Welsh Lower Layer Super Output Area (LSOA) in 2005. This has been done by combining small area level data from the 2001 census with the appropriate Family Resources Survey (FRS).

The work follows on from a literature and methodological review which culminated in a recommendation to explore the spatial microsimulation approach (Ballas, Dorling et al. 2006).

This paper describes the use of the method to estimate income deprivation at the LSOA level for 2005 for Wales using the adjusted OECD equivalence scale and for income both before and after housing costs. Thus the full income deprivation indicator specifications are:

1. **% HHBMI (BHC) = % of households in each LSOA whose equivalised net household income before housing costs (BHC) is less than 60% of the median equivalised net household income before housing costs for Wales**.
2. **% HHBMI (AHC) = % of households in each LSOA whose equivalised net household income after housing costs (AHC) is less than 60% of the median equivalised net household income after housing costs for Wales**.

In addition to the production of the indicators at LSOA level the paper also describes work to estimate the number and proportion of children living in 'poor' households as defined by these indicators in each LSOA.

2 OVERVIEW OF APPROACH

Following the review of methods described in Ballas et al (2006) and previous work at the LSOA level for England (Anderson 2007) we have implemented a spatial microsimulation approach to combine census and FRS data for appropriate years in order to estimate the indicators (HHBMI) for 2005. For the purposes of this preliminary work this requires the pooling of the FRS 2003/4 and 2004/5 surveys to produce a larger Welsh sample and the use of Census 2001 small area data.

The spatial microsimulation 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;

4 And they reasonably predict the indicator at the micro (household) level.

Variables fitting these criteria are identified using multivariate regression techniques and are then used to iteratively re-weight the survey data to fit each zone (LSOA) to the relevant Census 2001 tables using a deterministic iterative proportional fitting method that is discussed in more detail below. At the end of this process each household in the FRS is characterised by n fractional weights where n is the number of zones under study. It is then a relatively straightforward matter to use the weighted households to estimate the proportion of households whose income is below a given poverty line or the proportion of children in each LSOA who are living in a 'poor' household.

2.1 Spatial Data

The small area data used in this work was the Census 2001 LSOA small area tables accessed

via CASWEB[1].

A review of the available household level Census 2001 tables for Wales on CASWEB 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 elsewhere (Ballas, Dorling et al. 2006), Williamson and Voas (2000; 2005) have shown that the variables shown in Table 1 are reasonable predictors of household income at the small area level.

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 IMD 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 WIMD 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 produces the relatively short list of candidate constraint variables shown in Table 2. 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. 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] http://casweb.mimas.ac.uk/

Table 2: Constraint variable candidates

Variable	Definition	Notes
Number of earners in household	0,1,2,3+	
Employment status of HRP	NS-SEC 1, NS-SEC 2, NS-SEC 3, inactive, retired	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	
Gender of HRP	Male (0) / Female (1)	
Region	English Government office regions, Wales, Scotland and Northern Ireland	Used to provide regional weighting
Age of HRP	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	
Number of cars	0,1,2+	Not collected in FRS
Presence of central heating	0 (no) / 1 (yes)	Not collected in FRS
HRP born in UK	0 (no) / 1 (yes)	Not collected in FRS

2.2 Income data

The income survey data used is the Welsh subsample of the FRS 2003/4 and FRS 2004/5. The base income variable used is 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
- Other benefits
- Other/remaining sources

In order to align the income values with the DWP HHBAI definitions (HBAI07, Appendix 1) 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 (AVCs) 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;
- Parental contributions to students living away from home;
- Student loan repayments.

To calculate after housing costs (AHC) income the FRS variable 'gbhscost' is subtracted from the BHC income value. The variable 'gbhcost' is defined as '*the total amount spent on water and sewerage rates, rent, mortgage interest, household rent, structural insurance (adjusted for combined cases to be consistent with HBAI) and service charges*'.

3 EQUIVALISATION AND MEDIAN CALCULATIONS

The modified OECD equivalisation scale (Table 3) was used to control for household composition and to produce an equivalised measure of household income before housing costs.

Table 3: Modified OECD equivalisation scale (see Table A2 1.0, Appendix 2, HBAI07)

Composition	Scale value (BHC)	Scale value (AHC)
1st Adult	0.67	0.58
Spouse	0.33	0.42
Other 2nd adult	0.33	0.42
3rd adult	0.33	0.42
Subsequent adults	0.33	0.42
Children aged < 14	0.20	0.20
Children aged 14+	0.33	0.42

Thus the two equivalised income indicators were calculated by dividing the relevant (BHC/AHC) income by the aggregated household composition based weight.

This produces two final income variables – equivalised income before and after housing costs. These are then used as the basis for the calculations of the Welsh BHC/AHC medians and thence the allocation of households to the two indicator groups – above or below 60% of the relevant Welsh median. It should be noted that these calculations were done before the data were then pooled to provide a larger Welsh sample covering 2003/4 and 2004/5.

The final step is to create two variables which represent the number of children if the household is defined as income deprived or if they are not.

Table 4: FRS BHC/AHC equivalised income results for Wales

		BHC	AHC
2003/4			
	N	1278	1278
	Mean	£17,725.85	£16,277.85
	SE Mean	£422.08	£425.39
	Median	£14,961.59	£13,599.41
	60% median	£8,976.95	£8,159.65
2004/5			
	N	1239	1239
	Mean	£18,640.04	£16,991.67
	SE Mean	£342.20	£349.07
	Median	£16,172.00	£14,564.74
	60% median	£9,703.20	£8,738.84

It should be noted that households with negative income are retained. Households reporting negative BHC income constitute 0.55% of Welsh households in 2003-4 (0.81% in 2004/5) whilst 0.86% (1.37%) report negative AHC income. It is not expected that retaining negative income will therefore have any significant effect on the indicators as they will not substantially effect the median derived calculations.

4 CONSTRAINT SELECTION

The first key test of the potential utility of the constraint variables is the extent to which they predict income and, in particular the extent to which they predict the income deprivation indicator at the micro (household) level. This is relatively easy to assess using standard regression techniques and using the r square 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, this work 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 affects 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 22 and Table 23 in Annex A.1 report the results of six stepwise multivariate logistic regression models for Wales, two (BHC and AHC) for 2003-4 and similarly for 2004-5 and for the pooled 2003-5 FRS. Table 5 summarises these results and shows that we can be justified in pooling the 2003-4 and 2004-5 FRS data since the 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) there are additional significant constraint variables - HRP gender in the case of BHC and HRP age in the case of AHC.

Table 5: Significant constraints (in decreasing order of explanatory power)

2003-4		2004-5		2003-05 pooled	
BHC	AHC	BHC	AHC	BHC	AHC
Employment Status	Employment Status	Employment Status	Employment Status	Employment Status	Employment Status
Number of Earners	Tenure	Number of Earners	Number of Earners	Number of Earners	Tenure
Tenure	Number of Earners		Tenure	Tenure	Number of Earners
				Gender of HRP	HRP age

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

5 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 LSOAs in the sense that the weights represent the 'fractional existence' of the corresponding household in the corresponding LSOA. Conceptually the results can be thought of as a matrix of LSOAs (rows) and households (columns) where each cell contains the weight for that household in that LSOA.

To do this two sets of tables are required for each constraint for each LSOA: the Census 2001 small area tables for the constraints (e.g. Table 6) and the analogous small area tables constructed from the FRS households for the region in which the zone is found (e.g. Table 7)

Table 6: Small Area Table for number of earners derived from Census 2001 for the first LSOA in Wales

Zonecode	Number of households	Number of earners = 0	Number of earners = 1	Number of earners = 2	Number of earners = 3+
W01000001	517	294	132	85	6

Table 7: 'Small Area Table' for number of earners derived from the FRS 2003/4/5 for Wales

Number of households	Number of earners = 0	Number of earners = 1	Number of earners = 2	Number of earners = 3+
1308	608	333	320	47

Starting with LSOA 1 all household weights (w_i) are initially set to 1. Following Ballas et al we implemented a regional weighting scheme so that only households belonging to the same UK Government Office region as the particular zone are allocated to it. In this case this means that only FRS households from Wales are placed into Welsh LSOAs. This avoids filling, for example, Cardiff 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 8: First four Zone 1 households with initial weights

case	region	age	rooms	persons	nssec	comp	nearners	w_i
26115	10	2	3	0	1	2	1	1
26116	10	2	2	0	3	2	0	1
26117	10	2	3	4	0	0	2	1
26118	10	4	3	0	0	2	1	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 6) and s_{hj} is element hj of the FRS statistical table (Table 7).

As an example, using the number of earners constraint Table 9 shows the calculations of the first weights for the first four households so that the FRS sample fits the Census distributions on this one dimension.

Table 9: First four Zone 1 households with weights after fitting to constraint 1

case	region	nearners	W_1
26115	10	1	= 1 * (132/333) = 0.396
26116	10	0	= 1 * (294/608) = 0.484
26117	10	2	= 1 * (85/320) = 0.266
26118	10	1	= 1 * (132/333) = 0.396
..	..	..	..

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 (cf. Table 5) so that the last to be fitted is the one which 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 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 10 iterations were sufficient to achieve a stable indicator value. Thus after iterating over the re-weighting

procedure 10 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 report using a process of integerisation once the weights had been calculated to select the 'best fit' households for a given area (Ballas, Clarke et al. 2005). 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 % households below 60% median income) for each zone (in this case LSOAs).

Table 10: Example simulation output file (partial)

AREA	area_reg	HH_id	WEIGHT	hhincyearly	Hhbmi_bhc
W01000001	10	10246	.21	13884	0
W01000001	10	10247	.18	3692	100
W01000001	10	10248	.04	31564	0
W01000001	10	10249	.23	16432	0

Table Note: hhbmi_bhc = 100 if net equivalised income is less than 60% of the median, 0 otherwise.

Table 10 shows the first four rows of an example output file. For each area (e.g. LSOA) 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 (BHC) indicator as calculated in the source FRS data. 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 % 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.

6 RESULTS: HOUSEHOLDS BELOW MEDIAN INCOME

The results of the spatial microsimulation process are summarised in Table 11 and Table 12.

Table 11 shows the 5 'poorest' and 5 'wealthiest' LSOAs in Wales according to the BHC indicator whilst Table 12 shows the same but for AHC. The table also shows in which decile of the Index of Deprivation income domain score the LSOA is found. Thus the LSOAs with the highest %HHBMI are also generally in the 8th or 9th (i.e. most deprived) decile of the income domain.

Table 11: The 5 least and most deprived LSOAs in Wales (2003-5 pooled) by simulated % HHBMI (equivalised BHC)

	LSOA code	District name	Equivalised HHBMI (BHC)	Equivalised HHBMI (BHC) decile	Income domain score	Income domain score decile
1	W01001793	Cardiff	5.96%	0	1.1	0
2	W01001427	Caerphilly	5.99%	0	0.3	0
3	W01000282	Flintshire	6.10%	0	0.7	0
4	W01001636	Newport	6.37%	0	0.1	0
5	W01001841	Cardiff	6.41%	0	0.6	0
1	W01000220	Denbighshire	18.19%	9	27.9	7
2	W01001274	Rhondda, Cynon, Taff	17.63%	9	80.6	9
3	W01001275	Rhondda, Cynon, Taff	17.14%	9	44.4	8
4	W01001280	Rhondda, Cynon, Taff	16.91%	9	58.2	9
5	W01000226	Denbighshire	16.90%	9	40.9	8

Table 12: The 5 least and most deprived LSOAs in Wales (2003-5 pooled) by simulated % HHBMI (equivalised AHC)

	LSOA code	District name	Equivalised HHBMI (AHC)	Equivalised HHBMI (AHC) decile	Income domain score	Income domain score decile
1	W01001427	Caerphilly	7.19%	0	0.3	0
2	W01001636	Newport	7.21%	0	0.1	0
3	W01001841	Cardiff	7.23%	0	0.6	0
4	W01001224	Rhondda, Cynon, Taff	7.39%	0	0	0
5	W01000282	Flintshire	7.55%	0	0.7	0
1	W01001884	Cardiff	41.97%	9	98.2	9
2	W01000830	Swansea	40.37%	9	99.1	9
3	W01000862	Swansea	35.00%	9	96.6	9
4	W01001345	Caerphilly	34.61%	9	87.5	9
5	W01000570	Pembrokeshire	34.34%	9	65.6	9

As we can see the two indicators produce different results for the most deprived areas although rather similar results for the least deprived and this is confirmed by a Spearman rank correlation of only 0.679 and by Figure 1.

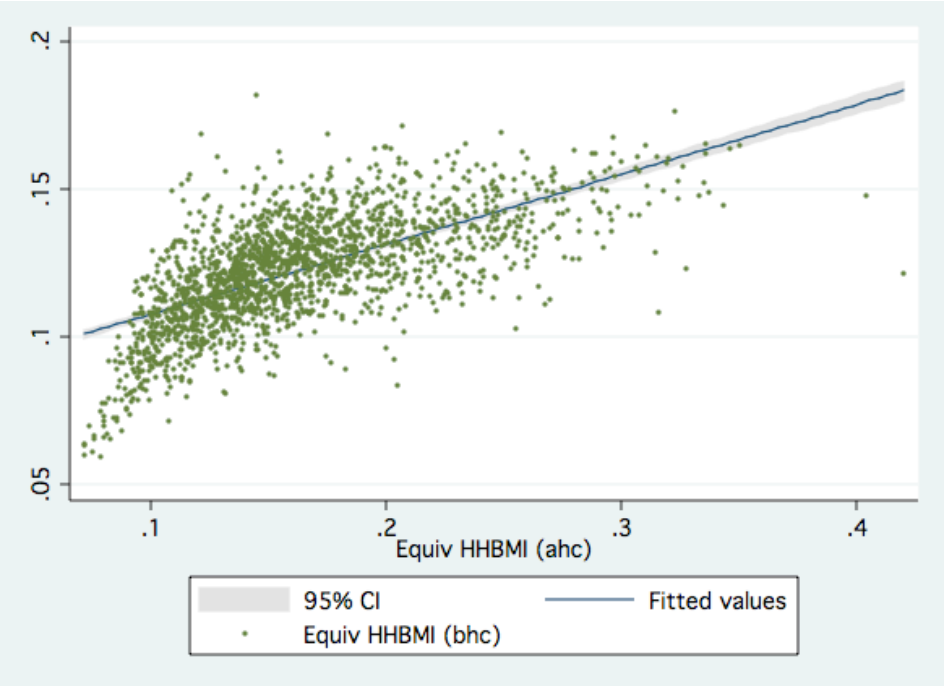

Figure 1: Comparison of BHC and AHC HHBMI indicators at LSOA for Wales (Census 2001, FRS 2003-05 pooled)

The spatial distributions of the BHC indicator (Figure 5) suggest a higher concentration of poor households in the former mining areas of South Wales and in the coastal areas on the Pembrokeshire/Ceredigion border as well as in other pockets in specific urban areas (see also Figure 2 and Figure 6). In contrast the areas with the highest % HHBMI according to the AHC indicator are more markedly concentrated in the Valleys and South Wales urban areas. Overall however the AHC indicator is considerably more diffuse in its distribution and thus may be a better 'relative poverty' indicator in comparison to the rather tighter BHC distribution which supports less differentiation (Figure 4).

It should therefore be apparent that the utility of each indicator in a revised Index of Multiple Deprivation will depend on the political objective and outcome desired since they reveal slightly different patterns of poverty.

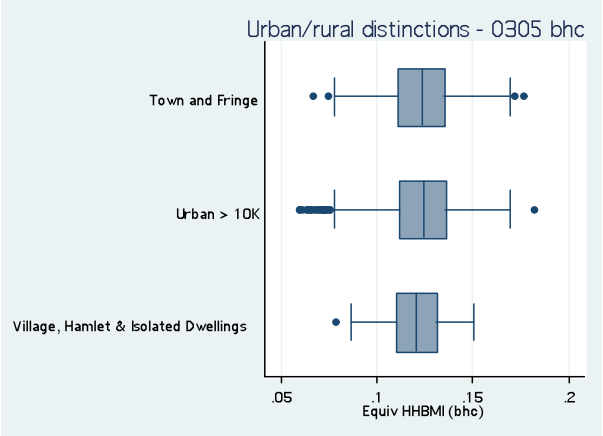

Figure 2: Comparison of BHC HHBMI indicators at LSOA for Wales across urban/rural classification types (Census 2001, FRS 2003-05 pooled)

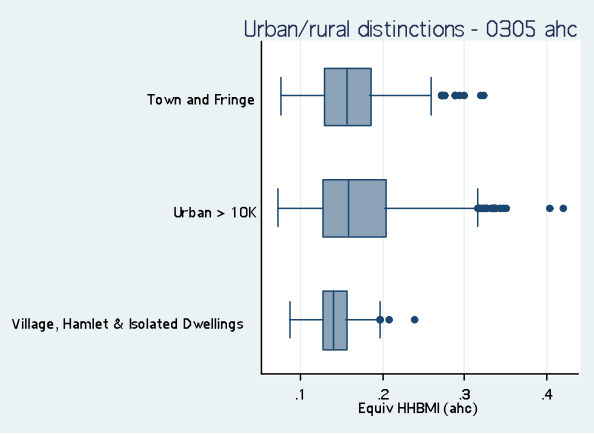

Figure 3: Comparison of AHC HHBMI indicators at LSOA for Wales across urban/rural classification types (Census 2001, FRS 2003-05 pooled))

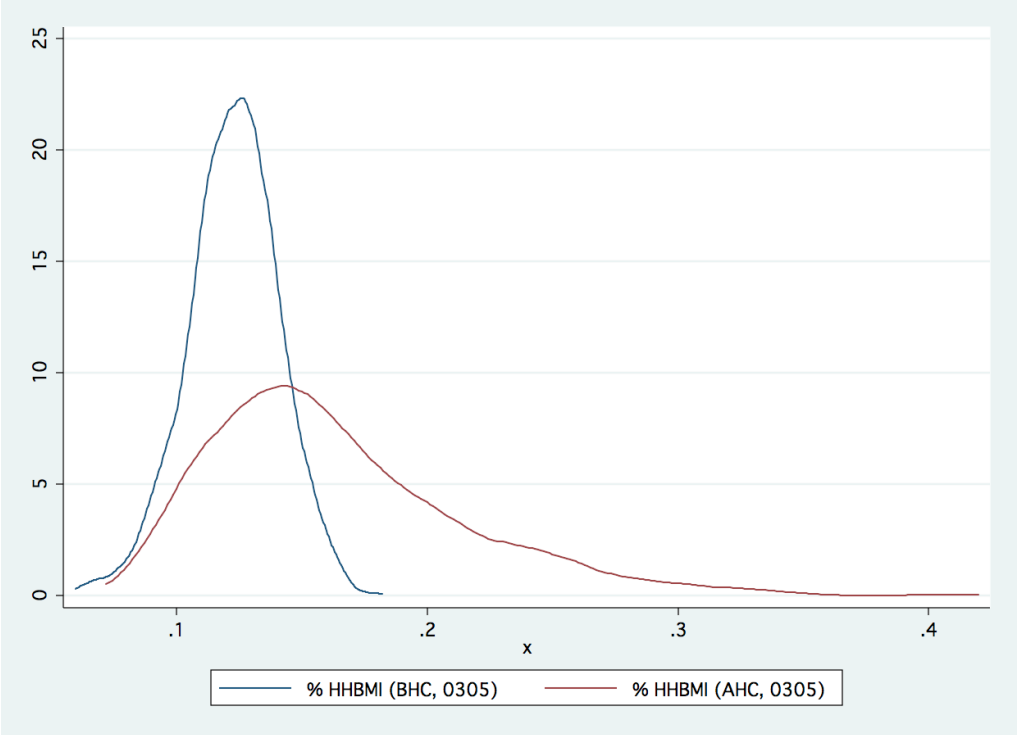

Figure 4: Kdensity distributions of BHC and AHC HHBMI indicators at LSOA level for Wales (Census 2001, FRS 2003-05 pooled)

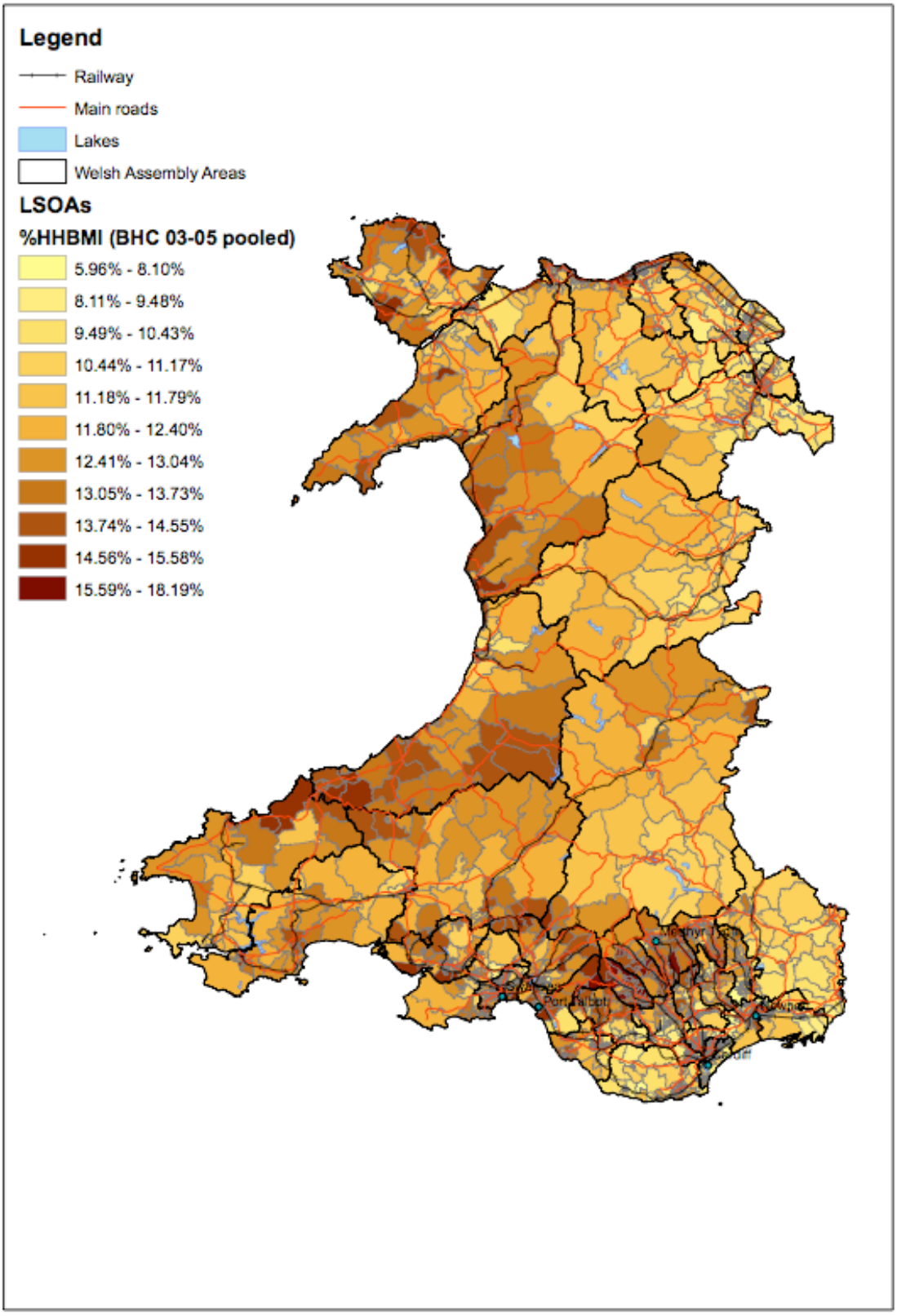

Figure 5: Spatial distribution of %HHBMI in Wales (BHC, LSOA level, Census 2001, FRS 2003-05 pooled)

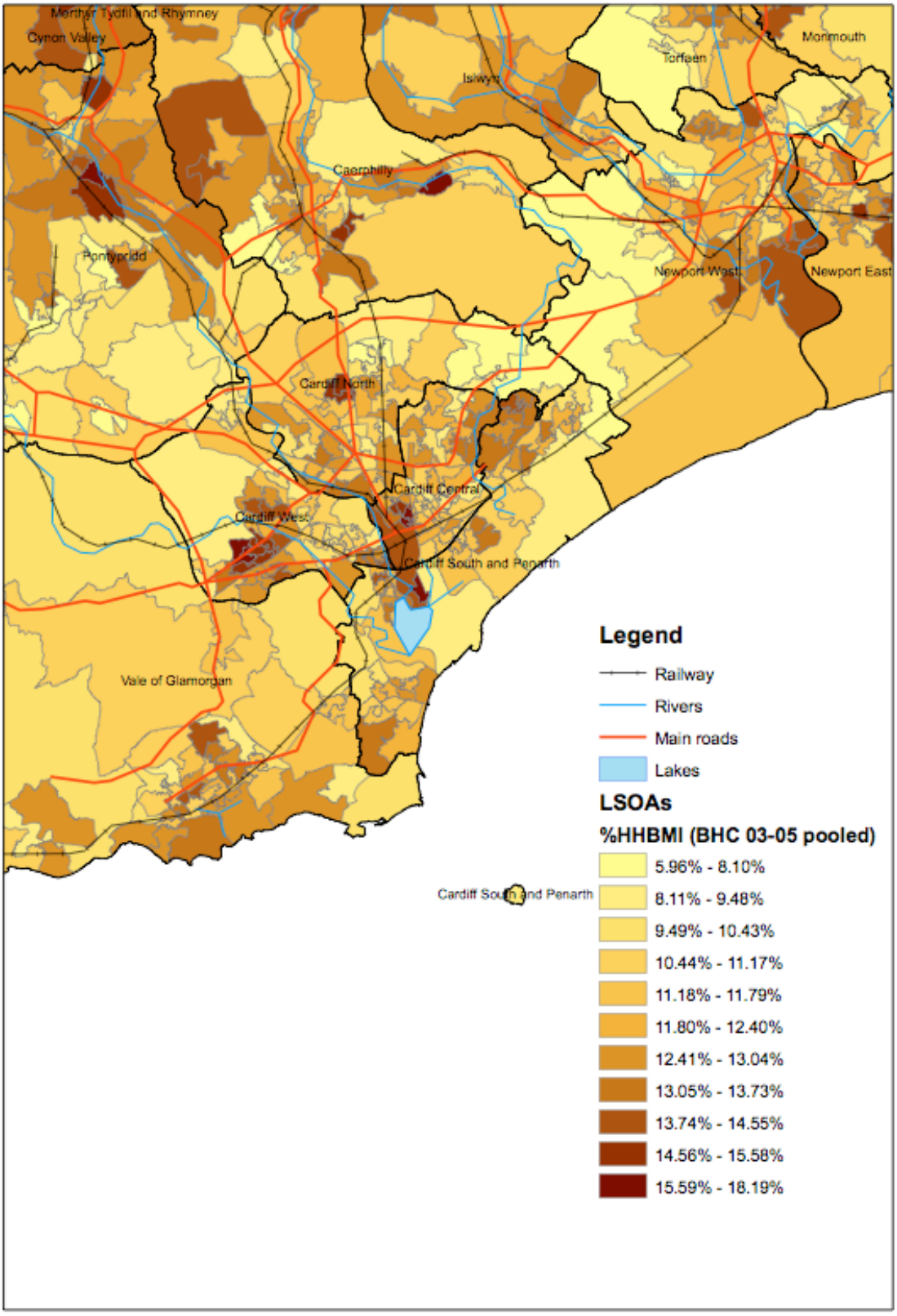

Figure 6: Spatial distribution of %HHBMI in the Cardiff area (BHC, LSOA level, Census 2001, FRS 2003-05 pooled)

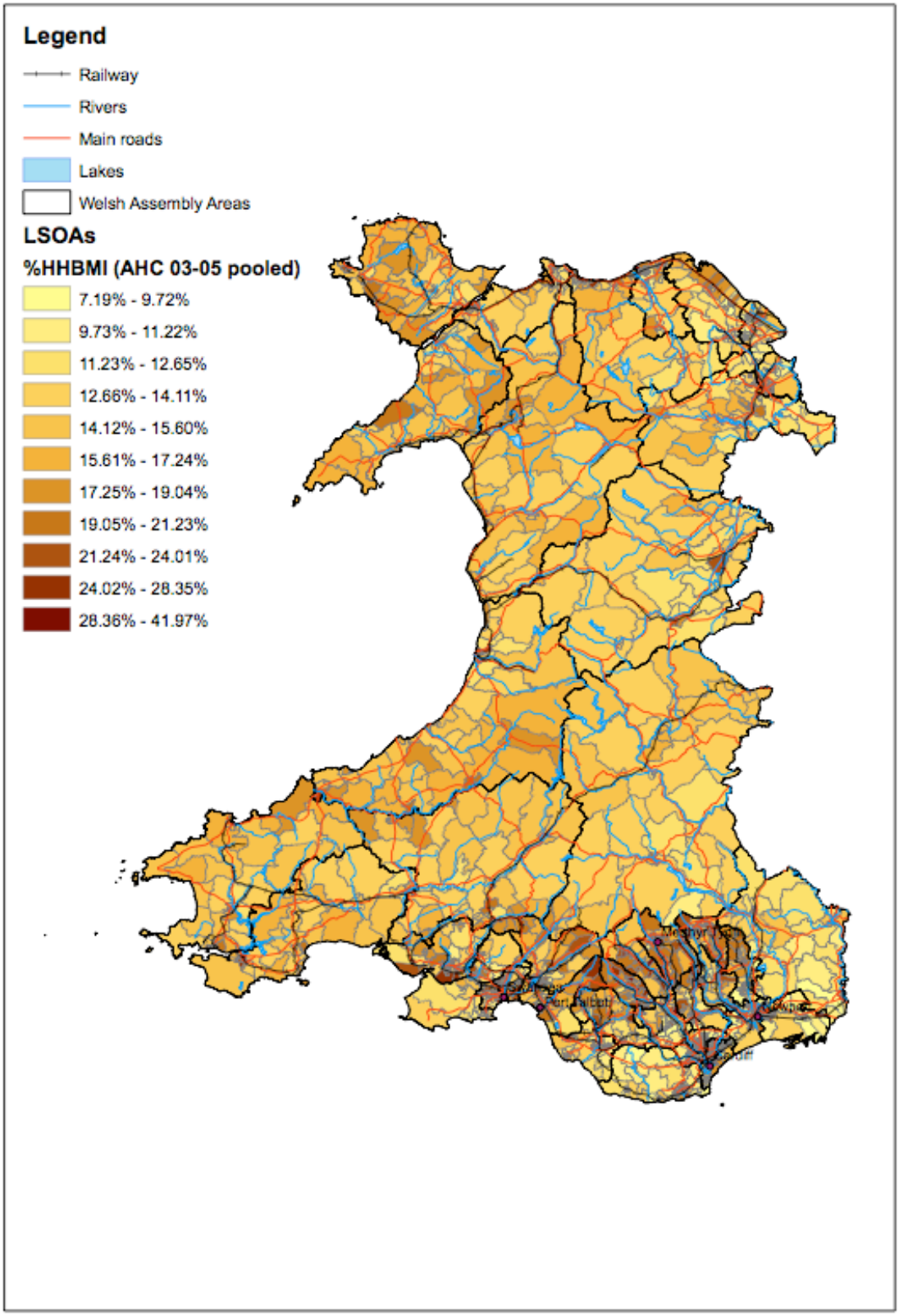

Figure 7: Spatial distribution of %HHBMI in Wales (AHC, LSOA level, Census 2001, FRS 2003-05 pooled)

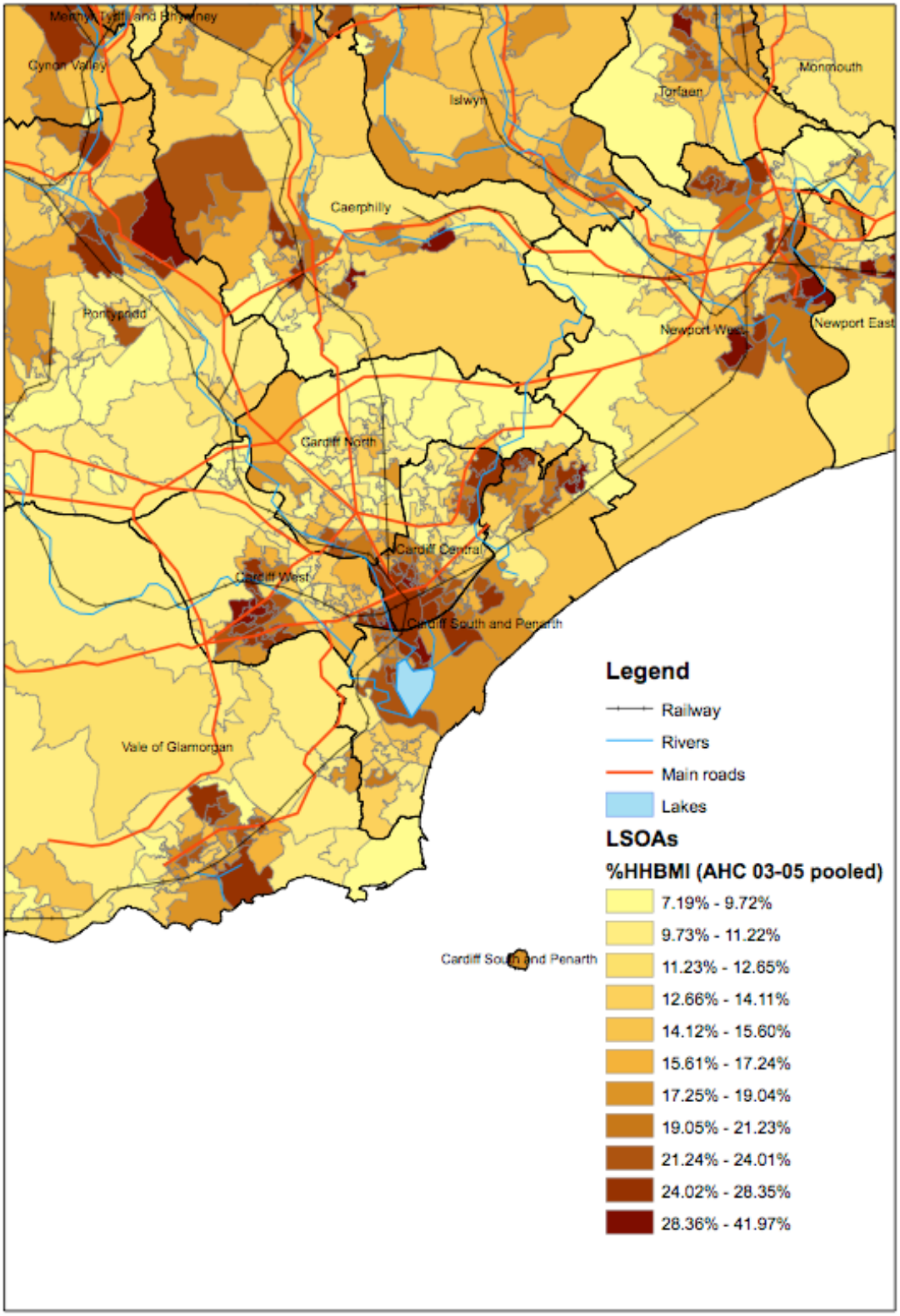

Figure 8: Spatial distribution of %HHBMI in the Cardiff area (AHC, LSOA level, Census 2001, FRS 2003-05 pooled)

6.1 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 census-derived 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 LSOA level results with other known small area estimates. In this case no equivalent small area data are available but we can compare the results to the income domain score of the Welsh IMD 2005;

6.1.1 Comparison with source and other survey data

Table 13 shows the HHBMI indicators (and 95% confidence interval) as calculated from the relevant source FRS data and as estimated from the spatial microsimulation process. Overall there appears to be a tendency to slightly underestimate HHBMI compared to the source FRS results. In general however 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 estimates provide a reasonable fit since they lie within these boundaries.

Table 13: Comparison of simulated mean regional HHBMI results with source (FRS 2003-4-5 pooled) results. Any simulated estimates lying outside the FRS 2003-4-5 observed means are shown in red.

	Source FRS 2003-4-5 pooled	SE mean	95% CI (+/-)	Spatial simulation (FRS 2003-4-5 pooled)	Difference
Equiv HHBMI (BHC)	12.992%	0.670	1.314	12.259%	-0.733%
Equiv HHBMI (AHC)	17.322%	0.754	1.479	16.294%	-1.028%

6.1.2 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' census constraint household counts with the estimated counts following the spatial microsimulation procedure. The total absolute error (TAE) is the difference for each constraint category for each area summed over all areas whilst the standardised absolute error (SAE) is TAE divided by the number of units (in this case households).

Whilst minimising the difference between the 'true' and estimated counts is the objective it is not yet clear in the literature what values of error are acceptable although Smith et al suggest that an SAE of less than 20% and ideally less than 10% in 90% of the areas is desirable especially where the prevalence rate of the phenomenon of interest is low (Smith, Clarke et al. 2009).

As Table 14 and Table 15 shows the HBMI models for Wales perform substantially better than this and we have disaggregated the SAE to reveal the constraints that show the poorest fit. We can see that the levels of error are relatively low for both indicators with the largest error being for households with no earners (11.9% in each case) and in all cases 90% of areas had SAE rates of less than 5%. We can also see how the order of the constraints means that the last category to be fitted (HRP Employment) fits perfectly.

Table 14: BHC indicator minimum, mean, maximum and 90% measures for SAE for each constraint for all Welsh LSOAs (2003-4-5).

Variable	Minimum	Mean	Maximum	90%
Tenure: Own	0.01%	1.83%	7.13%	3.51%
Tenure: Council rent	0.00%	0.36%	3.60%	0.99%
Tenure: Social rent	0.00%	0.11%	2.99%	0.29%
Tenure: Private rent	0.00%	0.29%	4.26%	0.60%
N earners: 0	0.00%	1.67%	11.89%	3.98%
N earners: 1	0.00%	0.79%	8.76%	1.99%
N earners: 2	0.00%	0.70%	3.97%	1.65%
N earners: 3+	0.00%	0.17%	1.42%	0.38%
Employment (HRP): NS-SEC 1	0.00%	0.00%	0.00%	0.00%
Employment (HRP): NS-SEC 2	0.00%	0.00%	0.00%	0.00%
Employment (HRP): NS-SEC 3	0.00%	0.00%	0.00%	0.00%
Employment (HRP): Inactive	0.00%	0.00%	0.00%	0.00%
Employment (HRP): Retired	0.00%	0.00%	0.00%	0.00%

Table 15: AHC indicator minimum, mean, maximum and 90% measures for SAE for each constraint for all Welsh LSOAs (2003-4-5).

Variable	Minimum	Mean	Maximum	90%
Age (HRP): 16-24	0.00%	0.09%	2.17%	0.19%
Age (HRP): 25-34	0.00%	0.33%	3.58%	0.76%
Age (HRP): 35-44	0.00%	0.43%	2.99%	0.91%
Age (HRP): 45-54	0.00%	0.43%	2.56%	0.92%
Age (HRP): 55-64	0.00%	0.42%	2.33%	0.84%
Age (HRP): 65-74	0.00%	0.58%	2.00%	1.06%
Age (HRP): 75-84	0.00%	0.36%	1.51%	0.69%
Age (HRP): 85+	0.00%	0.12%	0.70%	0.24%
N earners: 0	0.00%	1.65%	11.87%	3.93%
N earners: 1	0.00%	0.78%	8.75%	1.93%
N earners: 2	0.00%	0.70%	3.98%	1.62%
N earners: 3+	0.00%	0.17%	1.42%	0.38%
Tenure: Own	0.00%	2.13%	7.18%	3.75%
Tenure: Council rent	0.00%	0.29%	3.74%	0.78%
Tenure: Social rent	0.00%	0.08%	1.77%	0.22%
Tenure: Private rent	0.00%	0.25%	3.98%	0.54%
Employment (HRP): NS-SEC 1	0.00%	0.00%	0.00%	0.00%
Employment (HRP): NS-SEC 2	0.00%	0.00%	0.00%	0.00%
Employment (HRP): NS-SEC 3	0.00%	0.00%	0.00%	0.00%
Employment (HRP): Inactive	0.00%	0.00%	0.00%	0.00%
Employment (HRP): Retired	0.00%	0.00%	0.00%	0.00%

Analysing the specific characteristics of those LSOAs where fit is poor and developing strategies to resolve this is outside the scope of the current project but could be an avenue for future work.

6.1.3 Comparison with other known small area estimates

In this section we compare the results with the income domain score of the Welsh IMD 2005. Figure 9 shows the fit between the Welsh IMD 2005 income domain score and the simulated HHBMI indicators at LSOA level whilst Table 16 shows the rank order correlations. As we would expect there is a strong rank order correlation between HHBMI using the equivalised indicator and the WIMD income domain score and this is especially the case for the after housing costs indicator.

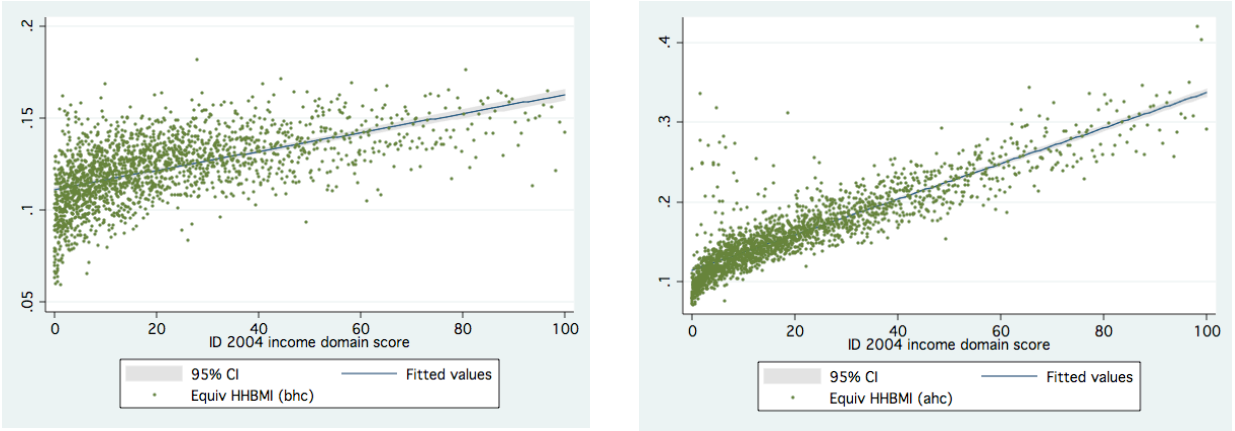

Figure 9: WIMD 2005 income domain score vs spatial microsimulation results for each indicator

Table 16: Rank order correlations comparing simulated 2005 results with WIMD 2005 income domain score at LSOA level within each region.

Indicator	Spearman rho
BHC	0.6041
AHC	0.8834

Figure 9 suggests the presence of a number of outliers which are low on the WIMD 2005 income score but relatively high on the simulated % HHBMI and this is particularly noticeable in the case of the AHC indicator.

Deeper analysis (Table 17) reveals that the two LSOAs which were in both the top 10% of HHBMI (AHC) and the bottom 10% of the IMD 2005 income score[2] have high or relatively high proportions of students (who cannot claim relevant benefits) according to the 2001 Census. This suggests that one of the main differences between the Indices of Deprivation income domain results and the HHBMI (AHC) results will be the inclusion of low income student households. LSOAs with higher proportions of students are therefore likely to appear to be 'more deprived' using the HHBMI indicator than would be the case for the WIMD income domain score.

Table 17: Outlier analysis

LSOA code	LA name	%HHBMI (equivalised,BHC)	%HHBMI (equivalised,AHC) decile	WIMD 2005 income domain score	WIMD 2005 income domain score decile	% students in LSOA (of all persons aged 16-74)	% students rank (Wales)
W01000092	Gwynedd	24.77%	9	1.9	0	69.93%	5
W01001725	Cardiff	33.58%	9	1.6	0	78.07%	1

7 RESULTS: CHILDREN IN POVERTY

Besides the level of household income deprivation in each LSOA, the small area distribution of child poverty is also of policy interest. In this section of the paper we describe preliminary results of using the spatial microsimulation method to estimate the proportion of children in each LSOA who live in a household which is 'poor' – that is it has an equivalised income that is

[2] Somewhat counter intuitively this means their WIMD 2005 income score would be low (i.e. not deprived) as the table shows.

less than 60% of the Welsh household median.

7.1 Method

The construction of the indicator followed the same process as that described above for the 2003-05 pooled sample with the following amendments:

- The total number of children in each FRS household (aged 0-16 or in full time education) were allocated to one of two variables: poor_kids and rich_kids depending on the income level of the household.
- The same simulation constraints were used as in the HHBMI indicator model but in addition the number of children (0,1,2+) was added to ensure a reasonable LSOA level fit to the number of children recorded by the 2001 census. It should be noted that the coarse granularity of this additional constraint means that there will not be an exact match between the total estimated number of children and the total recorded by the census. We analyse any differences below.
- The estimated variable was not %HHBMI but the % of children who lived in poor (and not poor) households in each area.

The model was repeated for the BHC and AHC measures.

7.2 Results

Figure 10 shows the estimated and Census recorded total number of children at LSOA level. As we can see the two simulation estimates are extremely close (r = 0.9998[3]) whilst both differ slightly from the Census (r = 0.9887 (BHC), r = 0.9881 (AHC). Table 18 shows that overall the mean difference for the BHC total was 1.8% and 2.2% for the AHC total with a maximum and minimum of c. 20%. Given that the indicator of interest is the proportion of children in each LSOA who are living in poor households the slight under or overestimation of the number of children will only bias the results if the model systematically over or underestimates either 'rich' or 'poor' children. In the absence of data that could be used to test this[4] we are forced to assume that this is not the case and that equal proportions of 'rich' and 'poor' children are under or overestimated in each LSOA.

Table 18: Difference between total number of children estimated and recorded as a proportion of those recorded (LSOA level)

Variable	Obs	Mean	Std. Dev.	Min	Max
BHC	1896	.018	.0386	-.187	.197
AHC	1896	.022	.0399	-.199	.209

[3] Using stata command: pwcorr
[4] Such as child tax credit data

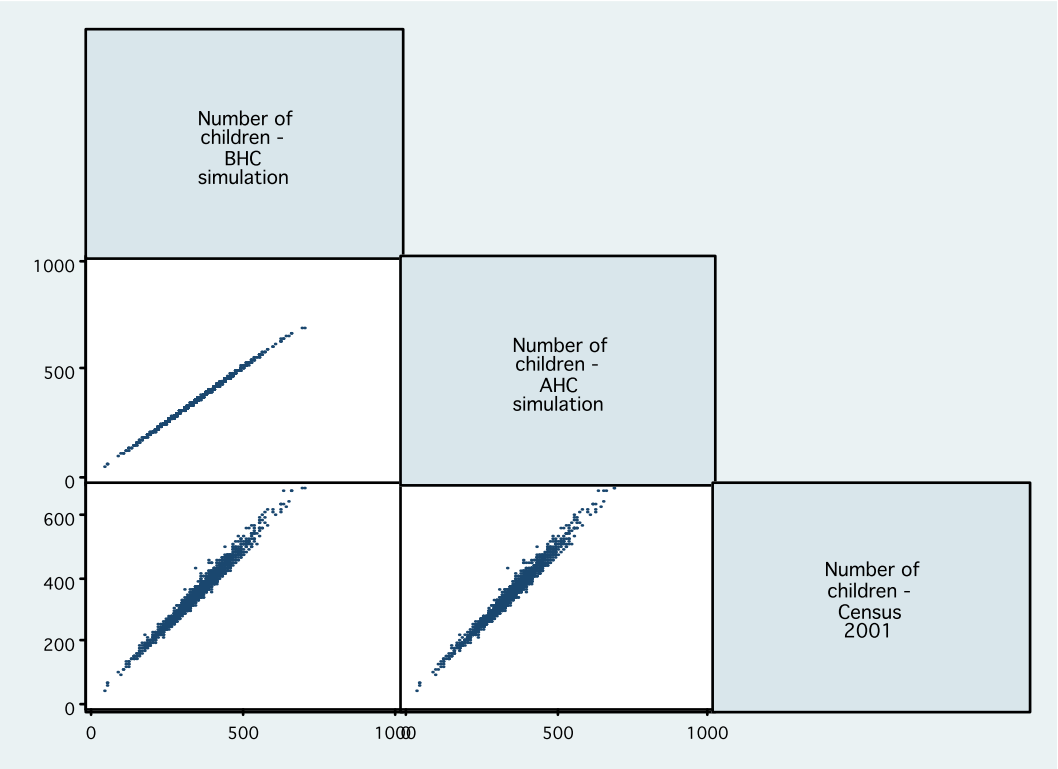

Figure 10: Comparison of BHC and AHC total number of children estimates (LSOA level, Census 2001, FRS 2003-05 pooled)

Figure 11 and Table 19 compare the results for the proportion of children in poor households with the original %HHBMI indicators and the IMD 2005 income domain score. As above the AHC indicators show the closest correlations.

Table 19: Rank correlations between child poverty, % HHBMI and income domain indicators (LSOA level, Census 2001, FRS 2003-05 pooled)

	%HHBMI (BHC)	% children in poor households (BHC)	%HHBMI (AHC)	% children in poor households (AHC)
% children in poor households (BHC)	0.819			
%HHBMI (AHC)	0.679	0.947		
% children in poor households (AHC)	0.630	0.925	0.989	
Welsh IMD 2005 Income Domain Score	0.604	0.855	0.883	0.893

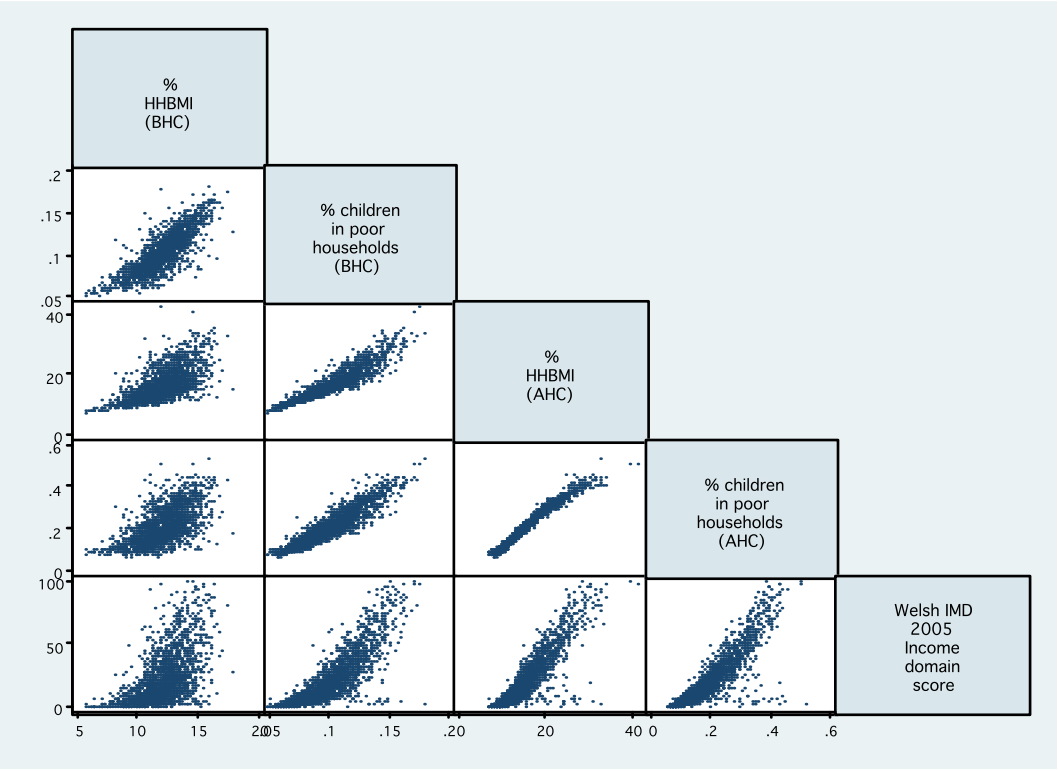

Figure 11: Comparison of % children in poor households (BHC and AHC) with %HHBMI indicators and WIMD 2005 income domain score (LSOA level, Census 2001, FRS 2003-05 pooled)

Table 20 and Table 21 show the lowest and highest 5 LSOAs for each indicator and for the most part confirm that the children in poverty estimates are similar to the %HHBMI and income domain indicators. However it is also clear that some LSOAs are estimated to have high rates of child poverty but low income domain scores – such as W01001725 (Cardiff, AHC) and W01000513 (Ceredigion, AHC). Both of these LSOAs have relatively few children (53 and 158 respectively which are in the lowest 5%) so that small number effects may be at work. However as reported above W01001725 has a high student population so that the %HHBMI indicator reports 'poor' households where the income domain score does not.

Table 20: Lowest and highest proportions of children in poor households (BHC, LSOA level, Census 2001, FRS 2003-05 pooled)

	LSOA	District	% children in poor households (BHC)	% HHBMI (BHC)	Welsh IMD 2005 Income Domain score decile
1	W01001725	Cardiff	18.13%	16.20%	0
2	W01001884	Cardiff	17.81%	12.18%	9
3	W01001274	Rhondda, Cynon, Taff	17.50%	17.63%	9
4	W01000830	Swansea	17.34%	14.77%	9
5	W01001699	Cardiff	17.32%	15.62%	9
1	W01001427	Caerphilly	5.02%	5.99%	0
2	W01001224	Rhondda, Cynon, Taff	5.17%	6.99%	0
3	W01001841	Cardiff	5.32%	6.41%	0
4	W01001636	Newport	5.33%	6.37%	0
5	W01001793	Cardiff	5.41%	5.96%	0

Table 21: Lowest and highest proportions of children in poor households (AHC, LSOA level, Census 2001, FRS 2003-05 pooled)

	LSOA	District	% children in poor households (AHC)	% HHBMI (AHC)	Welsh IMD 2005 Income Domain score decile
1	W01001725	Cardiff	53.06%	33.58%	0
2	W01001884	Cardiff	51.00%	41.97%	9
3	W01000830	Swansea	50.68%	40.37%	9
4	W01000743	Swansea	46.06%	25.66%	9
5	W01000513	Ceredigion	45.95%	31.79%	1
1	W01001847	Cardiff	6.39%	8.55%	0
2	W01001730	Cardiff	6.45%	9.36%	0
3	W01001778	Cardiff	6.67%	8.74%	0
4	W01001641	Newport	6.84%	9.12%	0
5	W01000397	Wrexham	7.01%	8.58%	0

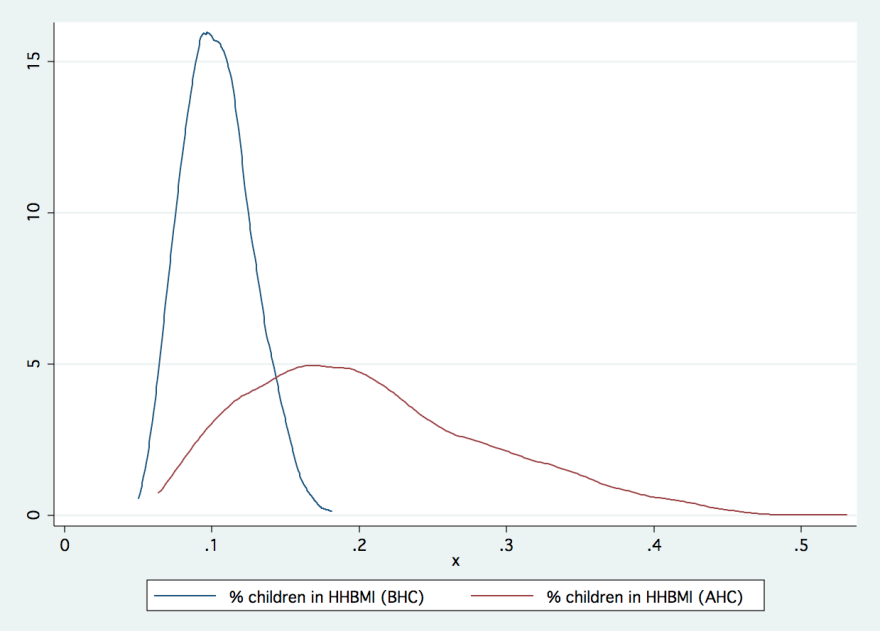

Figure 12: Kdensity distributions of BHC and AHC % children in HHBMI indicators at LSOA level for Wales (Census 2001, FRS 2003-05 pooled)

As with the %HHBMI the child poverty indicators vary spatially and the spatial patterns vary between the BHC and AHC forms (see Figure 13 to Figure 16). As before the AHC measure is

more obviously concentrated in the valleys and urban areas of south Wales and is also more diffuse (Figure 12) than the BHC indicator.

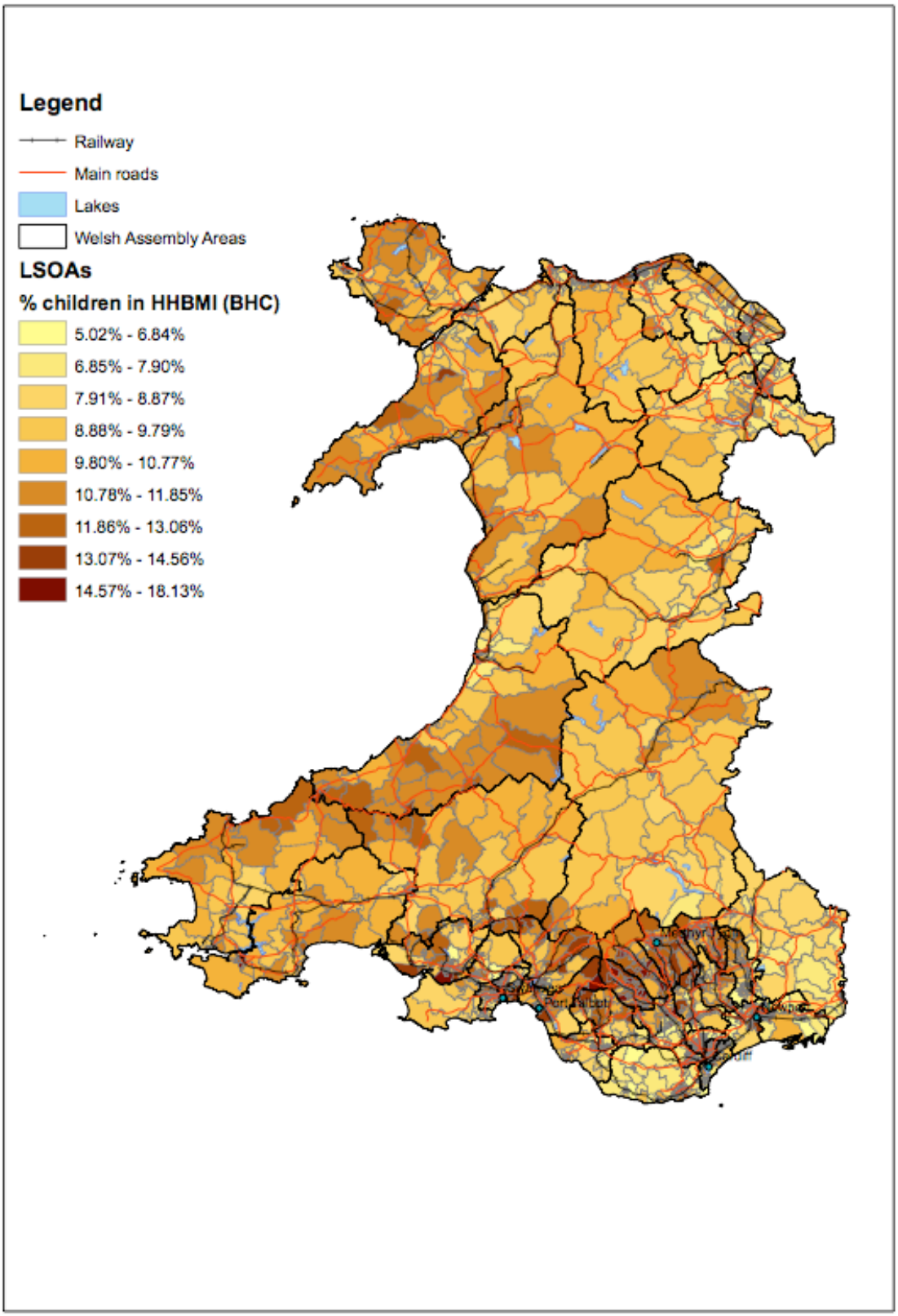

Figure 13: Spatial distribution of % children in HHBMI in Wales (BHC, LSOA level, Census 2001, FRS 2003-05 pooled)

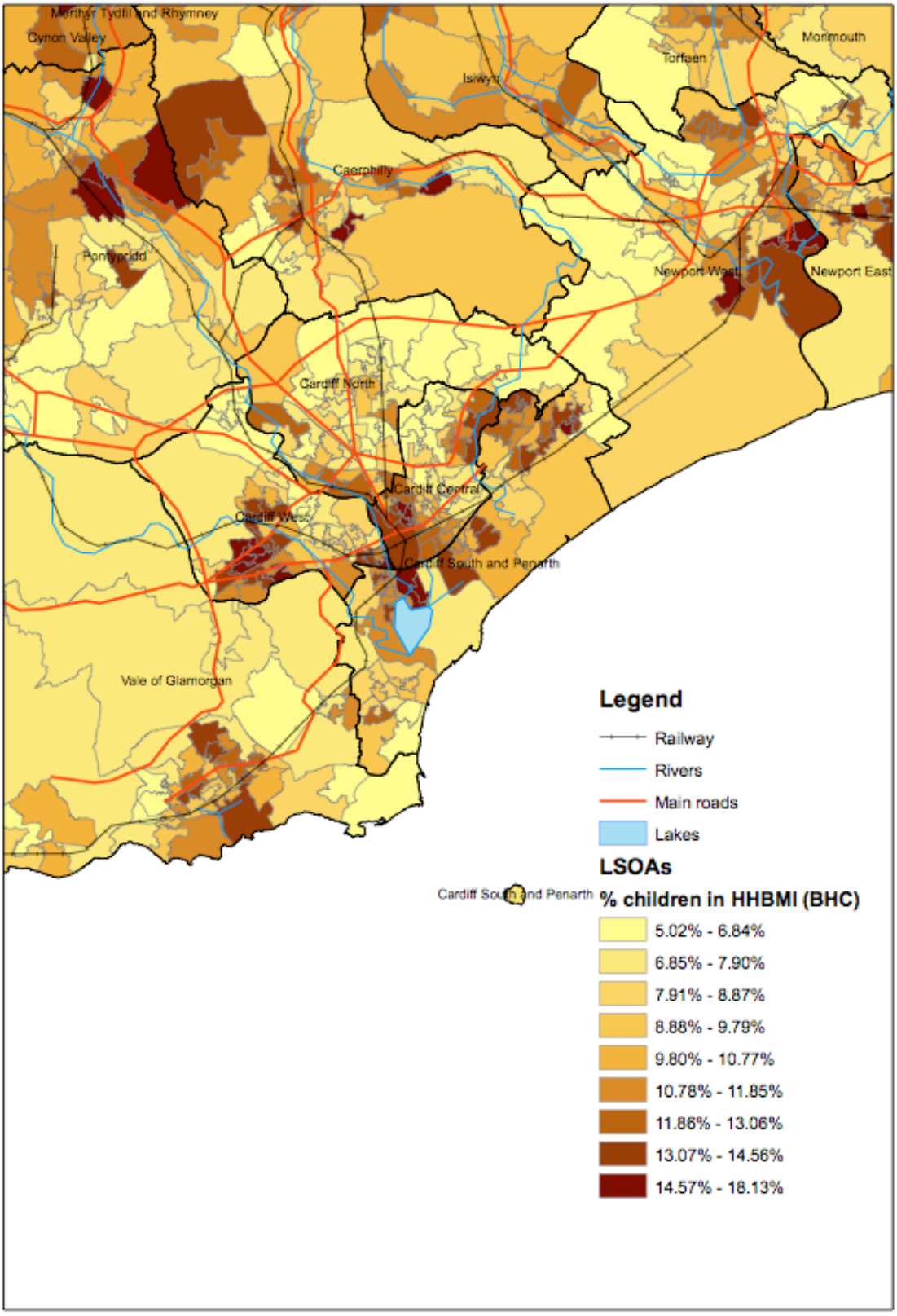

Figure 14: Spatial distribution of % children in HHBMI in the Cardiff area (BHC, LSOA level, Census 2001, FRS 2003-05 pooled)

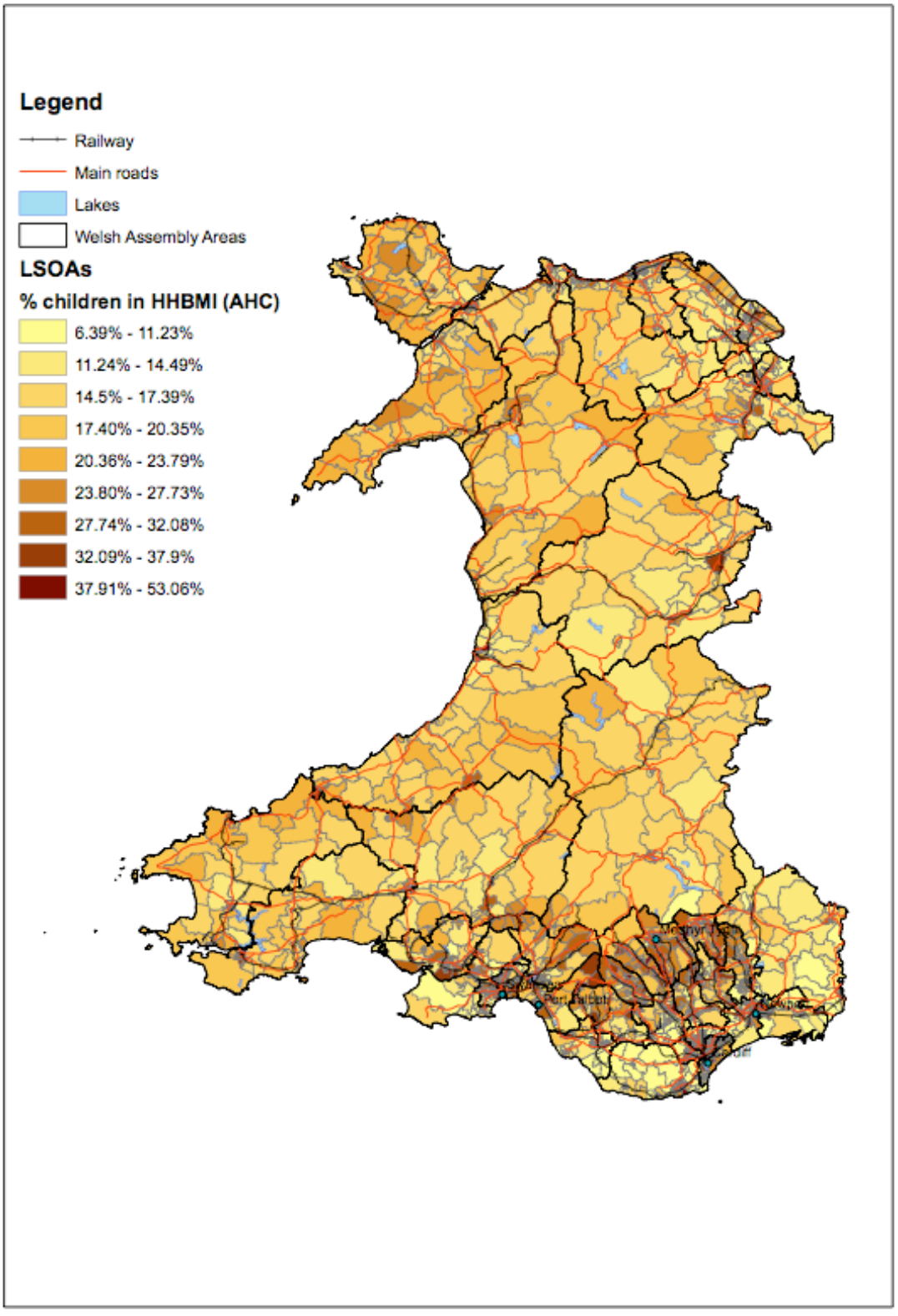

Figure 15: Spatial distribution of % children in HHBMI in Wales (AHC, LSOA level, Census 2001, FRS 2003-05 pooled)

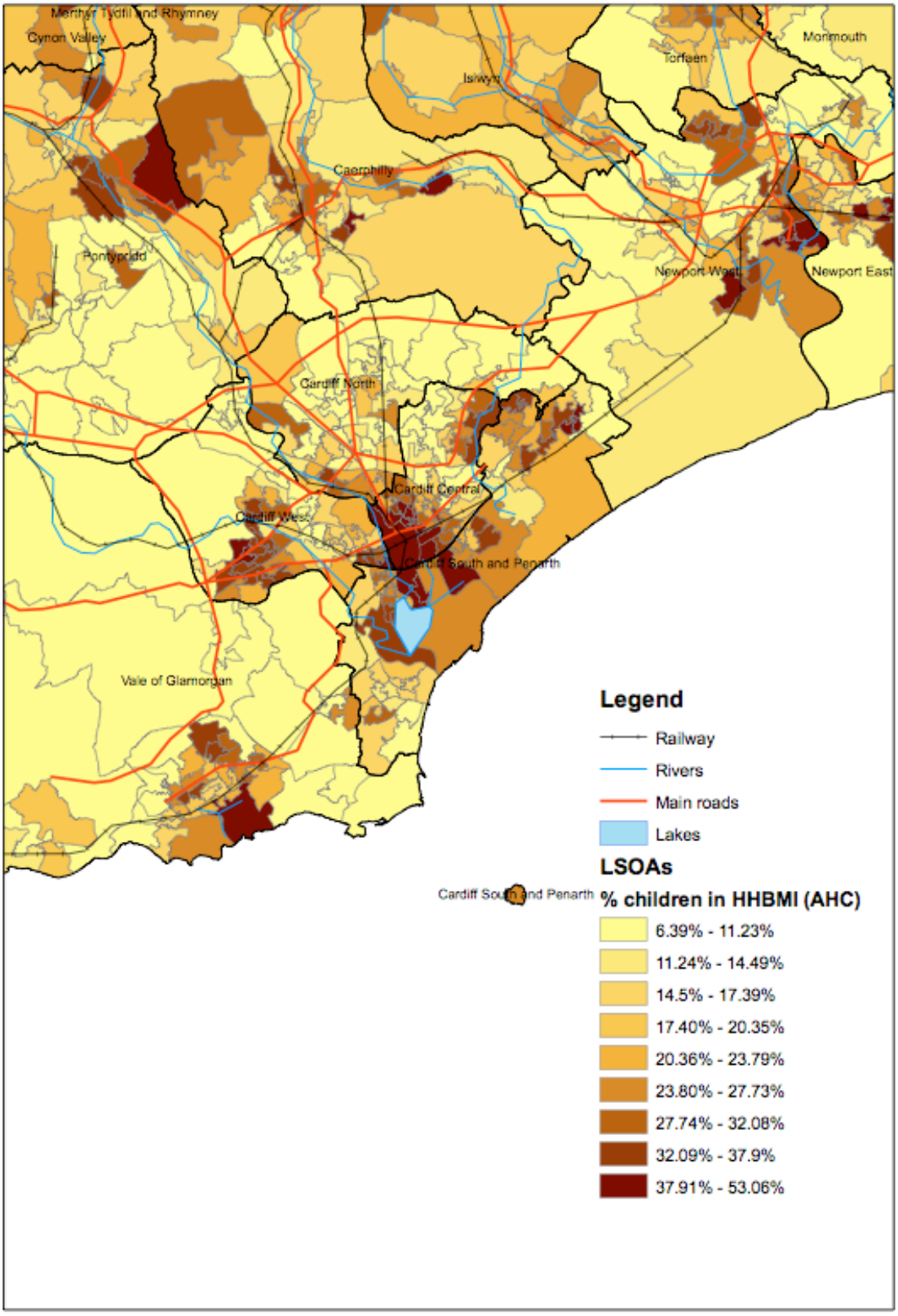

Figure 16: Spatial distribution of % children in HHBMI in the Cardiff area (AHC, LSOA level, Census 2001, FRS 2003-05 pooled)

8 IMPLICATIONS AND RECOMMENDATIONS

Overall the results of this preliminary work on equivalised HHBMI measures are encouraging. The results provide a synthetic household dataset which reproduces the Welsh % HHBMI (BHC/AHC) as measured by the FRS and which also produces a good fit to the Welsh IMD 2005 income domain score at the LSOA level. This is especially true for the AHC measure.

The results have also suggested that a focus on %HHBMI, and especially on the AHC indicator would present slightly different spatial distributions of income deprivation than does the WIMD 2005 income domain score. In particular there will be differences where students make up a high proportion of household response persons. This is of course likely to effect specific urban rather than rural areas. In addition the differences between the results for the BHC and AHC indicators mean that consideration needs to be given to which is the 'best' one to use in a future revision of the Welsh IMD. This cannot be answered by this paper as it is dependent on the policy context and the uses to which the Index and its components will be put.

The analysis of errors (SAE) suggests that in some LSOAs the spatial microsimulation method produces a less effective fit than in others for some constraints. This may be because these areas are made up of an unusual combination of household types and future work could investigate extending the spatial microsimulation method to account for such areas and thus to reduce overall error still further.

With respect to children in poverty the results also appear plausible but are perhaps even more vulnerable to the effects of inter-censual change. Provided such change does not affect the overall constitution of an LSOA then the results reported here should hold for 2003-5. However if an LSOA has changed to the extent that more low or high income households with children are present and/or these households have proportionately more children than in 2001 then the results will not give a reliable indicator of the current situation. That said these results provide, based on Census 2001 data, indications of where we would expect to find the highest concentrations of children in poor households and thus where we would expect greatest uptake of relevant benefits. Ground-testing these results against such data would be instructive both in terms of validating the results but also in identifying areas where either uptake is not as expected or the estimates appear incorrect.

9 ACKNOWLEDGEMENTS

Census data was originally created and funded by the Office for National Statistics 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 is copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

The IMD 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 Department of Communities and Local Government.

We would also like to thank Professor Holly Sutherland (ISER, University of Essex) for advice on the treatment of negative income and housing costs.

This work was sponsored by the Welsh Assembly Government.

10 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.

Smith, D., G. Clarke, et al. (2009). "Improving the synthetic data generation process in spatial microsimulation models." Environment and Planning A **41**: 1251 - 1268.

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 Testing Constraints

Table 22: Results of stepwise logistic regression models predicting the BHC equivalised indicators at the household level for Wales in 2003-4, 2004-5 and the pooled 2003-5.

2003-4	b	2004-5	b	2003-5	b
Employment: NS-SEC2 (NS-SEC1)	1.495	Employment: NS-SEC2 (NS-SEC1)	1.918	Employment: NS-SEC2 (NS-SEC1)	1.742
Employment: NS-SEC3	0.506	Employment: NS-SEC3	0.254	Employment: NS-SEC3	0.458
Employment: Inactive	1.728	Employment: Inactive	1.699	Employment: Inactive	1.801
Employment: Retired	0.687	Employment: Retired	0.896	Employment: Retired	0.797
Number of earners	-0.823	Number of earners	-0.833	Number of earners	-0.85
Tenure: Rent from council (Own)	-0.455			Tenure: Rent from council (Own)	-0.544
Tenure: Social rent	-1.063			Tenure: Social rent	-1.206
Tenure: Private rent	0.126			Tenure: Private rent	-0.134
				Gender of HRP (female)	0.307
Constant	-2.256	Constant	-2.396	Constant	-2.392
Pseudo r sq	0.122	Pseudo r sq	0.126	Pseudo r sq	0.129
N	1261	N	1231	N	2501

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

Table 23: Results of stepwise logistic regression models predicting the AHC equivalised indicators at the household level for Wales in 2003-4, 2004-5 and the pooled 2003-5.

2003-4	b	2004-5	b	2003-5	b
Employment: NS-SEC2 (NS-SEC1)	1.592	Employment: NS-SEC2 (NS-SEC1)	1.638	Employment: NS-SEC2 (NS-SEC1)	1.728
Employment: NS-SEC3	0.966	Employment: NS-SEC3	0.447	Employment: NS-SEC3	0.72
Employment: Inactive	2.177	Employment: Inactive	1.643	Employment: Inactive	1.999
Employment: Retired	0.352	Employment: Retired	-0.422	Employment: Retired	0.856
Tenure: Rent from council (Own)	0.851	Number of earners	-0.98	Tenure: Rent from council (Own)	0.637
Tenure: Social rent	0.941	Tenure: Rent from council (Own)	0.697	Tenure: Social rent	0.561
Tenure: Private rent	1.186	Tenure: Social rent	0.373	Tenure: Private rent	0.892
Number of earners	-0.823	Tenure: Private rent	0.887	Number of earners	-0.951
				age 25 to 34 (16-24)	-0.382
				age 35 to 44	-0.267
				age 45 to 54	-0.274
				age 55 to 64	-0.921
				age 65 to 74	-1.311
				age 75 to 84	-1.47
				age 85 or over	-1.41
Constant	-2.506	Constant	-1.786	Constant	-1.641
Pseudo r sq	0.234	Pseudo r sq	0.219	Pseudo r sq	0.232
N	1270	N	1231	N	2501

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

