Estimating Census Health Geographies: Using Synthetic Estimation with Secondary Survey and Census Data

An ESRC Secondary Data Analysis Initiative Project

Working Paper #1

Synthetic estimates of self-assessed health status: The inclusion of individual level data in the estimation process

Graham Moon[1]

Liz Twigg[2]

Joanna Taylor[1]

[1] University of Southampton, [2] University of Portsmouth

Synthetic estimates of self-assessed health status:

The inclusion of individual level data in the estimation process

Section 1 Introduction - the two methodologies for multilevel synthetic estimation

The last decade has seen a growth in the use of synthetic estimation to generate small area estimates in the absence of direct estimates from social surveys. The technique uses a statistical model (such as a multilevel model) applied to survey data on an outcome of interest linked to a set of associated predictor variables. These predictor variables must also be available for all small areas to generate estimates of the outcome for each locality. Two methodologies for synthetic estimation based on multilevel models have been adopted by various studies in recent years. The main difference between the two approaches is that one is based on models which only include area level independent variables, such as rurality or ecological measures of deprivation (Heady *et al.* 2003), whereas the second, more complicated approach is based on models which incorporate individual level explanatory variables such as age and sex as well as area level ones (Twigg *et al.* 2000).

A comprehensive evaluation of synthetic estimation of healthy lifestyles indicators by the National Centre for Social Research compared the performance of area based models with those that incorporate both individual and area level data (Pickering *et al.* 2004). The report's authors concluded that both approaches performed similarly well explaining similar proportions of the area-level variance. The resulting small area estimates were very highly correlated with each other and equally correlated with direct estimates from the Health Survey for England. Furthermore, comparisons with estimates from external surveys did not identify one approach to be consistently more accurate than the other. These findings lead the authors to conclude that they could not advocate one methodology over the other based on statistical criteria. The evaluation recommended area based models due to the ease of implementation, however, they went on to note that while incorporating individual level age-sex data did not manifestly improve the statistical performance of the model[1] it would be "*easier to 'sell' to potential user as, even though the evidence suggests otherwise, it seems more credible that aggregated individual behaviour is better predicted by aggregating*

[1] Their explanation for this was because age and sex distributions do not vary across small areas (in this case wards) to the same extent as other characteristics.

individual level estimates rather than modelling directly at the area-level...there seems little justification for excluding individual-level covariates" (Pickering *et al.* 2004, 56).

From a theoretical standpoint area only models have been heavily criticised for their lack of consideration of reality whereby no account is taken of the connections between individuals and the locality where they live their lives (Duncan *et al.* 1996; Duncan *et al.* 1998). Macintyre, Maciver and Sooman (1993) first introduced the concepts of "*composition*" versus "*context*" around 20 years ago in terms of the explanation of health outcomes. Areal variations in health outcomes were held to reflect both notions. They defined compositional effects as the characteristics of the people living in an area. Context was seen as the characteristics of the area itself; latterly this has been recognises to include both genuine ecological (area) effects as well as aggregate effects reflecting the characteristics of groups of individuals within an area. In a multilevel model, not only is it possible to start separating out compositional from contextual differences, but cross-level interactions can be included to allow for the fact that people may behave or think differently depending on where they live (Moon *et al.* 2005). To answer questions as to whether effects are contextual, compositional or indeed a product of interactions between the two, individual-level data with linked contextual information are needed (Ross and Mirowsky 2008). Synthetic regression models that combine individual and area characteristics within a multilevel framework recognise that individual behaviours or outcomes are dependent on both place and personal characteristics, that is to say factors associated with local context and composition.

The initial aim of this paper was to produce multilevel small area synthetic estimates of the proportion of adults assessing themselves as being in poor health. Because of the advantages of using multilevel models which incorporate both individual and area level data highlighted above, the starting point was therefore to calculate multilevel synthetic estimates based on this methodology. However the resulting estimates were far from satisfactory. This lead us to investigate whether, in certain contexts, multilevel models based solely on area level data actually perform better. Section 2 reviews previous synthetic estimates on general health status and covers the data sources utilised in this paper. Section 3 compares area-plus-individual models with area-only multilevel models before section 4 goes onto to evaluate synthetic estimates based on the two models. The paper concludes with section 5 which offers some possible suggestions for when it may not be advisable to include individual data in the multilevel synthetic estimation process.

Section 2 Previous work and data sources

Numerous studies have highlighted an association between health and deprivation. Riva *et al*. (2007) reviewed 39 multilevel studies on self-reported health status, reporting that significant associations were observed for at least one measure of area socio-economic status in all but two of the studies. Other area factors have also shown to be associated with poor health. Riva *et al*.'s (2009) results showed that rural dwellers were significantly less likely than residents of urban areas to report poor health. Migration has also been shown to be associated with both increases and decreases in place-specific health status (Norman *et al.* 2005). Furthermore, Mitchell *et al*. (2009) has suggested that certain areas can have lower death rates despite persistent economic adversity with the researchers suggesting that attracting (and retaining) residents may explain why these areas behave differently. Informed by the literature and previous synthetic estimates the following area level variables were included in the modelling process (Table 1).

Table 1 Area level variables

Area level variable	Details
Disability benefit claimants	Rate (per 1,000 adults aged 16+) who claimed a disability benefit in February 2011. The disability benefits at the time were Disability Living Allowance and Attendance Allowance.(1)
Working age benefits relating to illness and/or disability	Rates (per 1,000 adults aged 16+) who claimed a working age benefit in February 2011. The benefits at the time were Incapacity Benefit, Severe Disability Allowance and Employment and Support Allowance.(1)
UK wide deprivation measure	See Annex A for more information on this measure based on the Payne and Abel (2012) methodology for combining the English and Welsh individual Indices of Multiple Deprivation.
ONS rurality measure	Rural/urban definition, introduced in 2004 as a joint project between a number of Government Departments and was delivered by the Rural Evidence Research Centre at Birkbeck College. The categories at the MSOA level are "village, hamlet or isolated dwelling", "town or urban fringe" and "urban (>10k population)".
Population turnover	Estimates of previous inflow and outflow between MSOAs (based on 2008/09 figures) from the Office for National Statistics.

Notes:
(1) Denominator based on 2010 population estimates. Could not use 2011 Census information as these based on the 2011 version of MSOAs whereas the CSEW has the 2001 version of the Super Output Area codes attached.

Some of the area level variables detailed above at Table 1 were highly correlated (see the red correlations in Table 2 below). As a consequence the variable summarising the rate of working age benefits was excluded from the modelling process.

Table 2 Correlations between potential area level independent variables (with very high correlations being highlighted in red)

	UK wide deprivation score	Working age benefits	Disability benefits	Inflow
UK wide deprivation score	1.00	0.91	0.65	0.41
Working age benefits	0.91	1.00	0.82	0.33
Disability benefits	0.65	0.82	1.00	0.03
Population inflow	0.41	0.33	0.03	1.00

Notes:
All correlations significant at the 0.01 level (two tailed).

In terms of individual level variables, many other researchers (such as Bentham *et al.* 1995; Shouls *et al.* 1996) have found associations between an individual's socio-demographic characteristics and their health such as sex and age. The decision was made to restrict the individual level variables to just age and sex in order to make our findings directly comparable with those from the National Centre review on synthetic estimation (Pickering *et al.* 2004).[2] Furthermore, Asthana *et al.* (2004) previously showed how the contribution of gender and age to variations in self-reported general health status exceeded that of social class.

The data source for both the dependent variable (poor self-assessed health status) and the individual level independent variables (age and sex) was the 2010/11 sweep of the Crime Survey for England and Wales. At first glance this could seem an odd choice to model health. However, the survey was chosen for a number of pragmatic reasons. The primary reason is that few surveys are currently available with small area geography attached letting the researcher know roughly where the survey's respondents live. This is helpful in terms of the multilevel small area synthetic estimation methodology as it is necessary to be able to attach independent sources of data about the local area to the respondent's answers to the survey questions to give the contextual information about the local area. At the time of writing a geocoded version of the Crime Survey for England and Wales was, available via a special licence from the UK Data Service. Further advantages include the relatively large sample size (46,754[3] with a response rate of 76 per cent) and the fact that the primary sampling units are based on the census geography of Super Output Areas, with the sampling process being a

[2] With the exception of their model for smoking which also included marital status at the individual level.
[3] This number differs from the overall sample size as the latter number only counts those respondents who answered the general health question.

stratified (by Police Force Area) and a partially clustered design (Fitzpatrick and Grant 2011).

The question on self-assessed general health status in the Crime Survey was worded as "*how is your health in general? Would you say it is...READ OUT…very good, good, fair, bad or very bad?*". The answer categories "*fair*", "*bad*" and "*very bad*" were combined to describe someone in poor health. The justification for including fair in the poor health category was twofold. Firstly, the UK's Office for National Statistics (2011) advocates that the general health five point scale can be dichotomised in such a way citing evidence from the 2005 and 2006 General Lifestyles Surveys which found that more than half of those who said described their general health as fair also reported a limiting long standing illness or disability compared with less than ten per cent of those who said their health was either very good or good (Smith and White 2009). Secondly this approach is consistent with previous synthetic estimates on general health status (Heady *et al.* 2003).

Section 3 Comparing the multilevel models

Multilevel models were produced using the software package MLwiN v2.28 (Browne 2009a; Rasbash *et al.* 2009). All the models were initially estimated using iterative generalised least squares based on first order marginal quasi-likelihood approximation. The model coefficients were then checked for stability using Markov Chain Monte Carlo simulation – a Bayesian estimation technique. The default prior distribution applied by the software package for all the parameters was flat. Information on the conditional posterior distributions can be found in Browne (2012). The model was run through 50,000 iterations (with a burn in period of 5,000). The Raftery-Lewis diagnostic (Raftery and Lewis 1992) and the Effective Sample Size (Kass *et al.* 1998) both confirmed that this Markov chain length was sufficiently long..

There were 46,618 individuals at level 1 nested within 3,707 small areas (for the purposes of this paper the census geography of Middle Super Output Areas were used as the level 2 units) which are further nested within 42 Police Force Areas. Although no explanatory variables were included at this higher geography, Police Force Areas were included in the hierarchical structure of the multilevel models for two reasons. To begin with, this three level structure reflects the sampling rationale of the Crime Survey for England and Wales whereby the sample was stratified by Police Force Area. Moreover, the unexplained variation at level 3 can be accounted for in the synthetic estimation process by supplementing the fixed effects with area specific random coefficients (Scholes *et al.* 2008). This reduces the design bias of the synthetic estimates. Twigg, Moon and Walker (2004) used Government Office Region level residuals to improve their estimations of smoking behaviour based on multilevel models which incorporated both individual and area level covariates and Local Authority random effects were incorporated to better capture unexplained sources of variation and area heterogeneity when calculating small area unemployment estimates (Silva and Clarke 2008) based on area only multilevel models.

Table 3 details the multilevel model which incorporated both individual (age and sex) and area level independent variables, here on in referred to Model A.

Table 3 Multilevel model to predict poor self-assessed health status (Model A)

	β	SE(β)
Individual level variables		
Gender (base=male)		
Female	-0.003	0.023
Age (base=16 to 17 years old)		
18 to 19 years old	0.269	0.187
20 to 24 years old	0.309	0.154
25 to 29 years old	0.395	0.151
30 to 34 years old	0.353	0.149
35 to 39 years old	0.717	0.147
40 to 44 years old	0.842	0.145
45 to 49 years old	1.192	0.144
50 to 54 years old	1.518	0.144
55 to 59 years old	1.728	0.143
60 to 64 years old	1.842	0.143
65 to 69 years old	2.078	0.143
70 to 74 years old	2.291	0.143
75 to 79 years old	2.542	0.143
80 to 84 years old	2.713	0.146
85 years old or over	2.772	0.149
Area level variables		
Deprivation	0.032	0.004
Disability benefits	0.002	0.001
Inflow of residents	0.002	0.001
Type of area (base=urban)		
Town or fringe	-0.042	0.042
Village, hamlet or isolated dwellings	-0.142	0.042
Inflow * deprivation	-0.000	0.000
Constant	-3.365	0.163
Unexplained area level variation		
Level 2	0.056	0.012
Level 3	0.011	0.004
DIC	47,334	

Notes:
1. Greyed out explanatory variables were not significant at the 5% level.

A further model was then run, with the same area level explanatory variables as Model A but without the individual level variables (age and sex). Removing the individual level variables rendered many of the area level variables non-significant. Therefore the model was re-run until only significant area level variables remained. From here on in this model is referred to as Model B (Table 4).

Table 4 **Multilevel model (area level explanatory variables only) to predict poor self-assessed health status (Model B)**

	β	SE(β)
Area level variables		
Deprivation	0.006	0.001
Disability benefits	0.006	0.000
Constant	-1.753	0.042
Unexplained area level variation		
Level 2	0.042	0.004
Level 3	0.011	0.010
DIC	51,986	

An assumption behind all multilevel models is that the residuals at each level follow a Normal distribution. Figures 1 and 2 are Normal probability plots of the standardised and ranked residuals with the relatively straight lines indicating that the Normality assumption is valid for both Model A and Model B (Hox 2002).

Figure 1 **Residual plots for Model A**

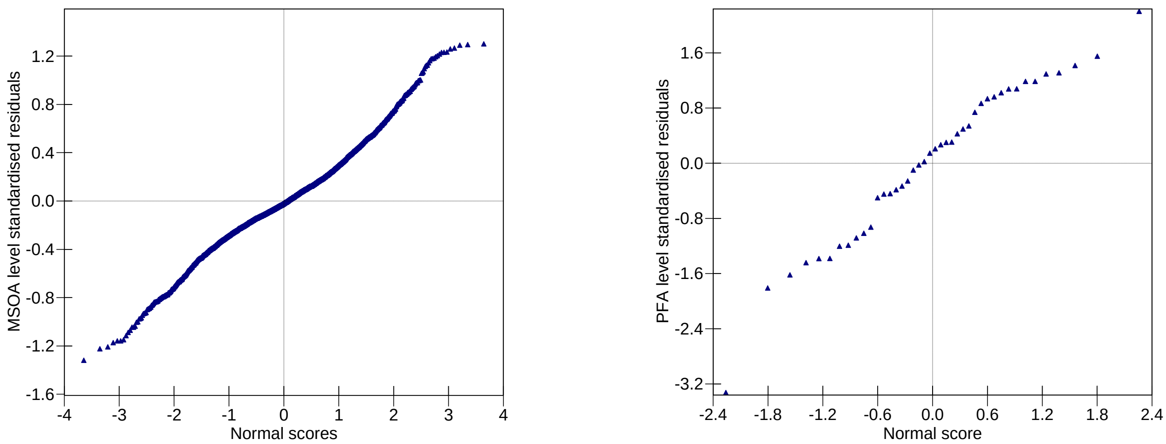
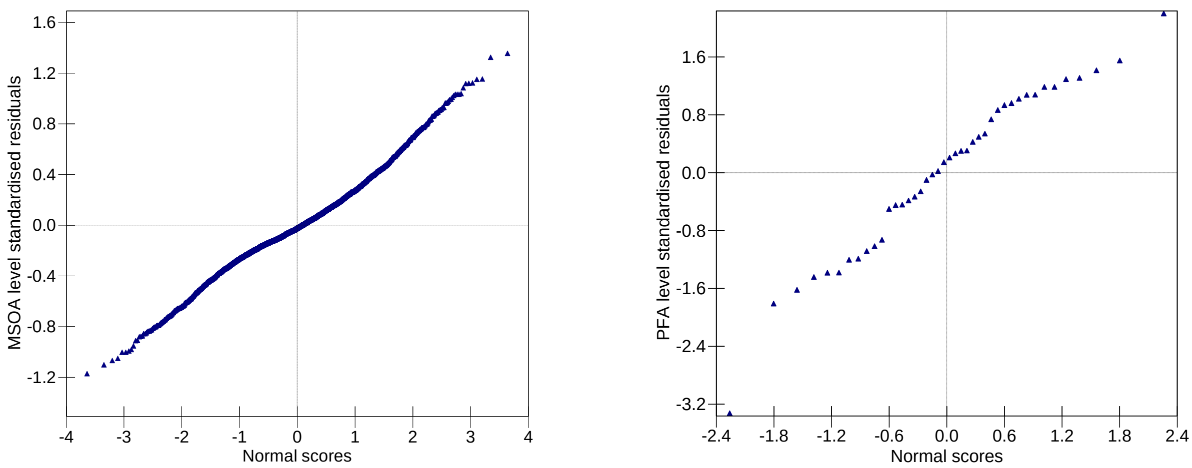


Figure 2 **Residual plots for Model B**



A useful tool to compare models is the Deviance Information Criterion or DIC (Spiegelhalter *et al.* 2002). This can be thought of as a measure of how well the model fits the data. The DIC diagnostic accounts for the number of parameters in each model. Consequently any two DIC values are directly comparable and so any decrease in the DIC suggests a better model (Browne 2009b). Following experience with the more frequently used Akaike's Information Criterion (Akaike 1974) a rule of thumb has been developed that differences of four or more suggest that the model with the higher DIC statistic has considerably less support (Burnham and Anderson 2002).[4] The DIC statistic for Model A at 47,334 is substantially lower than the corresponding statistic for Model B (51,986) indicating that the simpler model (Model B) has less support and, based on the DIC statistic alone, we should focus on Model A.

We can further compare the models in terms of the percentage of respondents which each model correctly predicts as reporting being in poor health (Field 2005). Two cut off points based on the 2010/11 sweep of the CSEW were tested – one based on the unweighted data and one based on weighted data. The percentages reporting poor health were 25.2 per cent and 21.3 per cent respectively. Based on the unweighted cut off point the multilevel model predictions based on Model A match 66 per cent of respondents' actual answers from the Crime Survey (Table 5) compared with 59 per cent for Model B (Table 6). The difference using the weighted cut off point is starker – 61 per cent (Model A) versus just 40 per cent correct (Model B).

Table 5 Cut off analysis for synthetic estimates based on Model A

	Unweighted cut off point		Weighted cut off point	
	Model prediction good health	**Model prediction poor health**	**Model prediction good health**	**Model prediction poor health**
CSEW good health	48.5	26.3	41.8	33.0
CSEW poor health	8.0	17.2	6.1	19.1

Notes:
1. Shaded percentages indicate a 'match' between the synthetic and census estimates.

[4] The DIC estimates the number of degrees of freedom as part of the model fitting process whereas the AIC does not.

Table 6 Cut off analysis for synthetic estimates based on Model B

	Unweighted cut off point		Weighted cut off point	
	Model prediction Good health	Model prediction Poor health	Model prediction Good health	Model prediction Poor health
CSEW good health	47.4	27.4	19.8	55.0
CSEW poor health	13.4	11.8	4.8	20.4

Notes:
1. Shaded percentages indicate a 'match' between the synthetic and census estimates.

Taking all the diagnostic tests together indicate that, as expected, the more expansive model (Model A) which incorporates both individual and area level variables is the preferred model. The next stage of our analysis was to use both multilevel models to generate synthetic estimates of poor self-assessed health status for every MSOA in England and Wales. Full details of the methodology for calculating synthetic estimates based on models which incorporate both individual and area level variables and those restricted to just area level variables can be found at Twigg *et al*. (2000) and Heady *et al*. (2003) respectively.

Section 4 Comparing the synthetic estimates

A common methodology to validate a set of synthetic estimates is to compare them to an alternative set of estimates for the same small areas (Scarborough *et al.* 2009). Consequently Figures 3 and 4 illustrate the scatter plots of the synthetic estimates (x axis) against the findings from the 2011 Census (y axis) at the MSOA level. Although it is expected that there will be a wide scatter (due to the large confidence intervals around the synthetic estimates), for a good model estimate the scatter should be around the line $x = y$, in other words a regression line should have a gradient close to one and an intercept around zero. The synthetic estimates based on Model A do not fulfil this criterion. Furthermore the Spearman's rank correlation is weak at 0.22 (Table 7).

The synthetic estimates based on Model B perform better with a Spearman's rank correlation of 0.92. Although the confidence interval for the intercept does not include zero, the gradient of the regression line is exactly one (Table 7). Of the two coefficients which make up a regression line it is arguable that the gradient is more important than the intercept, as it represents the relative relationship or rank between the two sources, whereas the intercept reflects any absolute differences between the sets of estimates.

Table 7 Synthetic estimates versus the 2011 Census

	Intercept	Lower CI	Upper CI	Contains zero?	Gradient	Lower CI	Upper CI	Contains one?	ρ(3)
Model A	16.90	16.20	17.53	×	0.40	0.35	0.45	×	0.22**
Model B	-2.45	-2.70	-2.19	×	1.00	0.99	1.01	✓	0.92**

Notes:
1. ** indicates correlation significant at the 0.01 level (two-tailed).
2. CIs represent 95 per cent confidence intervals.
3. ρ indicates Spearman's rank correlation.

Figure 3 Synthetic estimates from Model A

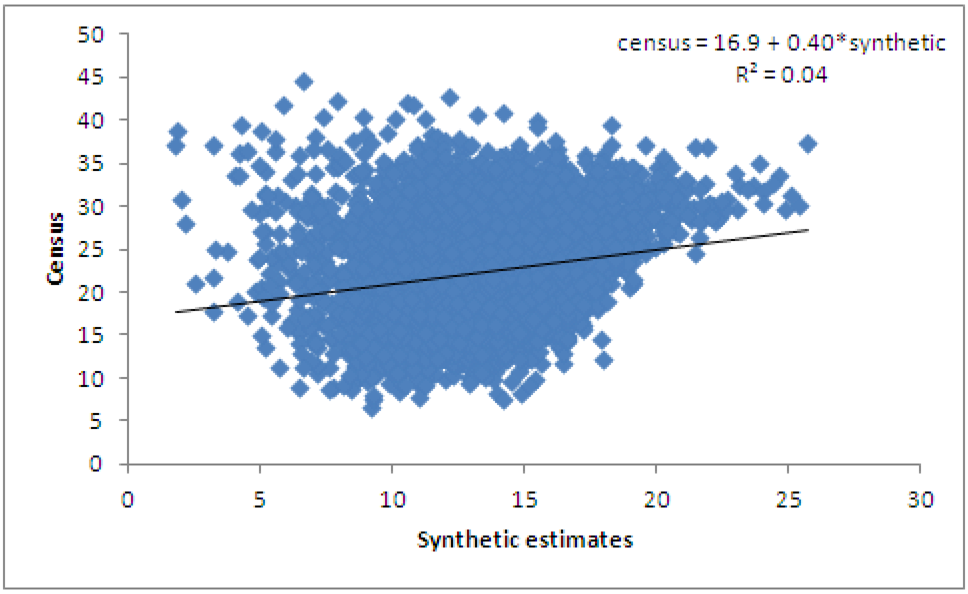
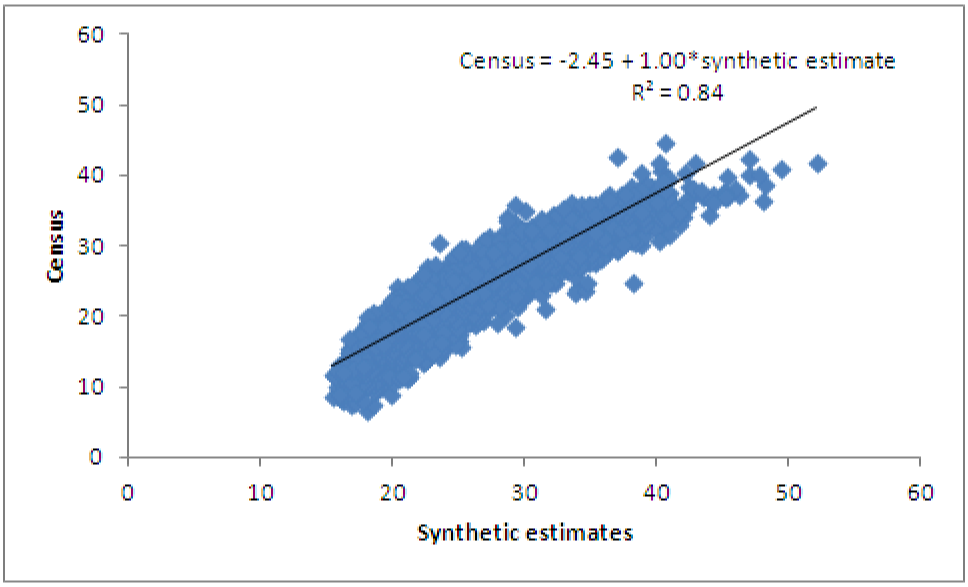


Figure 4 Synthetic estimates from Model B



Section 5 Discussion and conclusions

Why are the synthetic estimates from Model A performing so much worse than those based on the simpler model which does not include individual level explanatory variables (Model B)? Table 8 below gives some insight by detailing the percentage of the variance explained by each of the models. Pickering *et al.* (2004, 56) in their testing and evaluation study of synthetic estimation of healthy lifestyle indicators for the Department of Health recommended that 40 per cent of area level variance should be explained by the multilevel model as an absolute minimum. However, others such as Bauer (2009) and Snijders and Bosker (2012) have warned against such calculations for binary multilevel models. They argue that direct comparisons with the null model are not possible because the addition of new independent variables implicitly rescales both the fixed coefficients and the variances of the random effects due to the fact that level 1 variance is fixed (at ${\pi^2}/{3}$ in the case of the logit link). Snijders and Bosker (2012, 306) advocate that researchers should look at the total explained proportion of the variance calculated as:

$$Explained\ proportion\ of\ variance = \frac{\sigma_F^2}{\sigma_F^2 + \sigma_v^2 + \sigma_u^2 + {\pi^2}/{3}}$$

Table 8 Variance explained multilevel models on general health based on the 2010/11 Crime Survey for England and Wales

	Variance of the linear predictors σ_F^2	Intercept variance level 3 σ_v^2	Intercept variance level 2 σ_u^2	Explained % of the total variance
Null model	0.00	0.03	0.09	0.0%
Null model + age and sex	0.62	0.04	0.14	15.1%
Model A	0.70	0.01	0.06	17.2%
Model B	0.07	0.01	0.04	1.9%

Table 7 indicates that when the level 1 variables (sex and age) are added to the null model the area level variance (in this case levels 2 and 3 combined) actually increases from 0.12 to 0.18. This implies that area variations are stronger once we take account of the age and gender of the individuals who make up the local populations. This phenomenon of the inclusion of level 1 variables increasing the higher level unexplained variance has been observed in other multilevel studies (see for example Jones and Bullen 1993), especially in instances where the newly added variable explains mainly within group variation.

There are two possible explanations for this occurrence. The first is substantive, with the example often cited of house prices (Jones 1992, 246); "*it is possible for the inclusion of level 1 variables to result in an increase in the variance of the level 2 random term...it is being suggested that 'contextual' differences between areas may increase, when the 'composition' of properties in an area is taken into account".* The other reason is statistical whereby direct comparisons between models are often not possible when the outcome is binary or ordered-categorical. The problem is that the addition of a new predictor or random effect to the model implicitly rescales the coefficients of the prior predictors as well as the variances and covariances of any random effects. This rescaling makes it difficult to compare the results of sequentially fitted models for binary or ordered-categorical outcomes. See Bauer (2009) for criticism for the way that it has been done to date.

It is certainly plausible in the case of self-reported health status that the substantive explanation holds – in other words, when the age and sex mix of the local area is taken into account the associations with area level factors such as deprivation actually increases. This reflects debates fully discussed in Macintyre *et al*. (2002) on the role of the individual's socio-economic status versus the level of deprivation in the area as a whole. This paper, with its relatively basic models, is not designed to add to this debate as the multilevel models were built for the specific purpose of generating synthetic estimates. However, in terms of synthetic estimation, the phenomenon of individual level variables increasing the level of area level variation is a potential cause of the problem of the relative performance of the synthetic estimates based on Model A compared to Model B.

It is difficult and furthermore unwise to draw conclusions as to whether it is more appropriate to base synthetic estimates on multilevel models purely based on area measures as opposed to those which include both individual and area level independent variables based on this one example. What we can say, given the findings presented here, is that researchers should proceed with caution when the inclusion of individual level fixed effects actually increases the level of unexplained area level variation. In these instances, until rules of thumb have been established through further research, it would be prudent to generate two sets of synthetic estimates based on multilevel models with and without the individual level variables.

Annex A Combined UK index of multiple deprivation

The process followed the methodology first advanced by Payne and Abel (2012) with two important differences. Firstly, England was used as the baseline to generate adjusted UK wide scores (the Payne and Abel methodology used Scotland as the baseline). England was chosen due to the number of small areas covered. Secondly the health domain was excluded from the overall measure due to the fact that the dependent variables of interest were health related. The latest available versions of the four indices were included in the UK measure (Table A1).

Table A1 Countries indices of deprivation

Country	Year	Source
England	2010	https://www.gov.uk/government/publications/english-indices-of-deprivation-2010
Northern Ireland	2010	http://www.nisra.gov.uk/deprivation/nimdm_2010.htm
Scotland	2012	http://simd.scotland.gov.uk/publication-2012/
Wales	2011	http://wales.gov.uk/topics/statistics/theme/wimd/wimd2011/?lang=en

Regression models were calculated for each of the four constituent countries with the overall index (excluding health) as the dependent variable and the income and employment domains as the independent variables (Table A2).

Table A2 Relationships between the overall index and income and employment domains for all four countries in the UK

	β_0 (constant)	β_1 (income)	β_2 (employment)	R^2	σ
England (2010)	0.576	0.838	0.578	0.940	3.230
Northern Ireland (2010)	-4.687	0.601	0.610	0.962	2.574
Scotland (2012)	-1.082	0.783	0.672	0.962	2.716
Wales (2011)	-2.711	0.866	0.399	0.923	3.492

The results from these regression analyses were then used to calculate a UK wide adjusted deprivation score using the results from England as the baseline.

For example to calculate the Northern Ireland UK adjusted deprivation scores as:

$$IMDnohealthUK = 0.576 + (0.838 * INCOMErate) + (0.578 * EMPLOYrate) + {}^{residuals\,*\,3.230}/_{2.574}$$

Figures A1 to A3 compare Northern Ireland, Scotland and Wales' adjusted deprivation scores with their original scores (always excluding the health domains). Both Northern Ireland and Wales' the adjusted scores are greater than the original scores with the difference increasing for areas experiencing high levels of deprivation. The adjustments make little difference to the Scottish scores.

Kendall's τ for the association between adjusted and original IMD (excluding health in both cases) is 0.984, 0.967 and 0.964 for Northern Ireland, Scotland and Wales respectively.

Figure A1 Adjusted versus original results for Northern Ireland

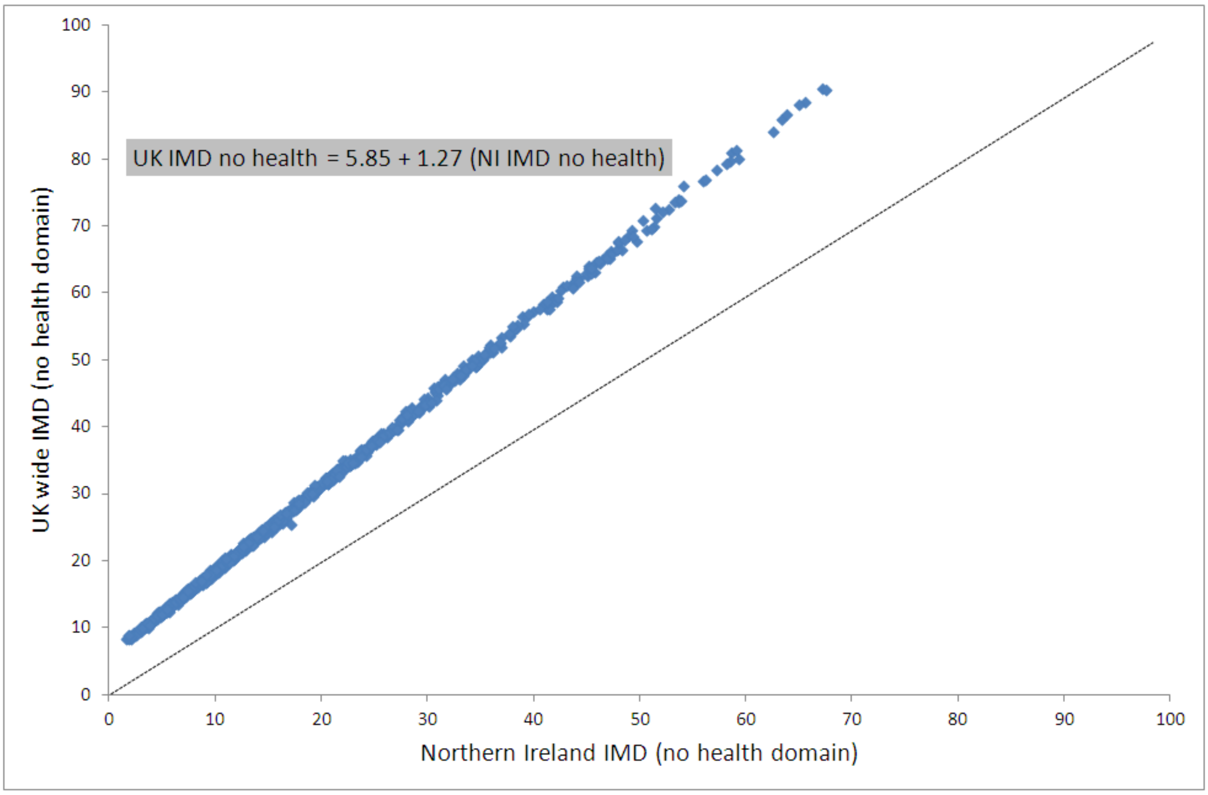


Figure A2 Adjusted versus original results for Scotland

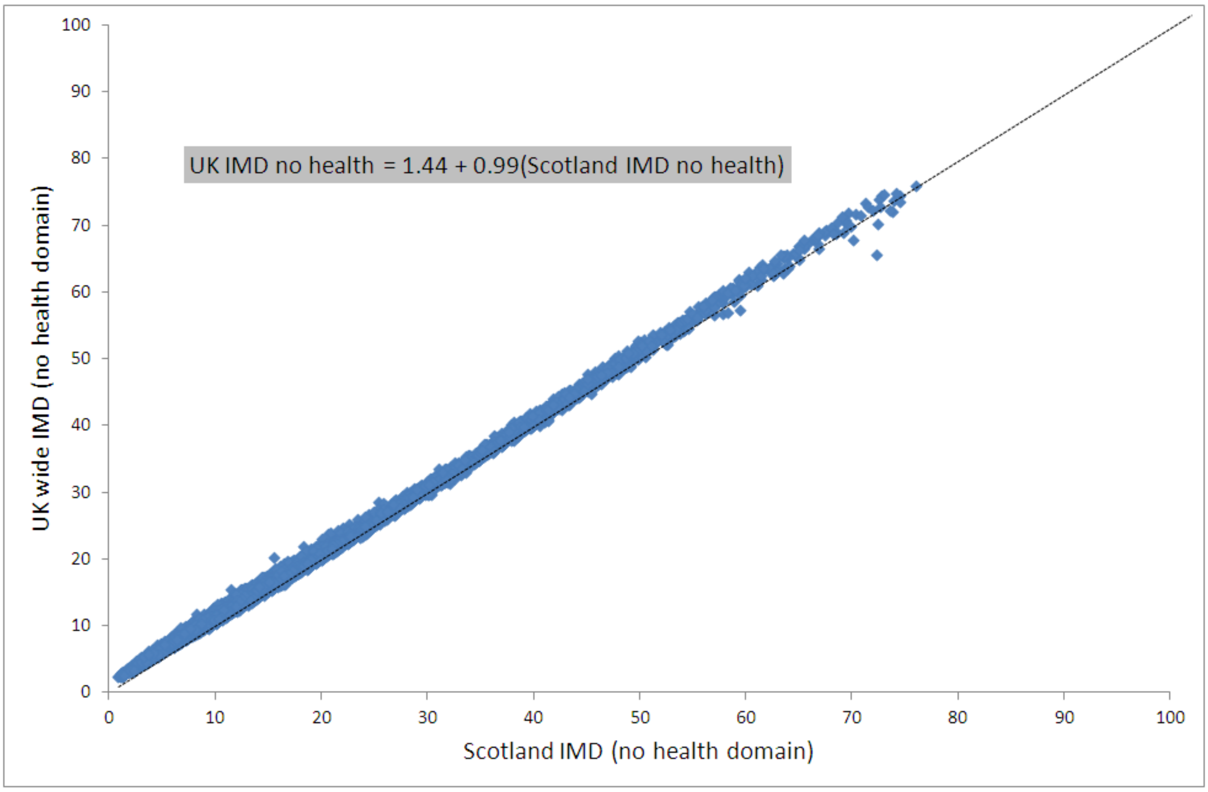
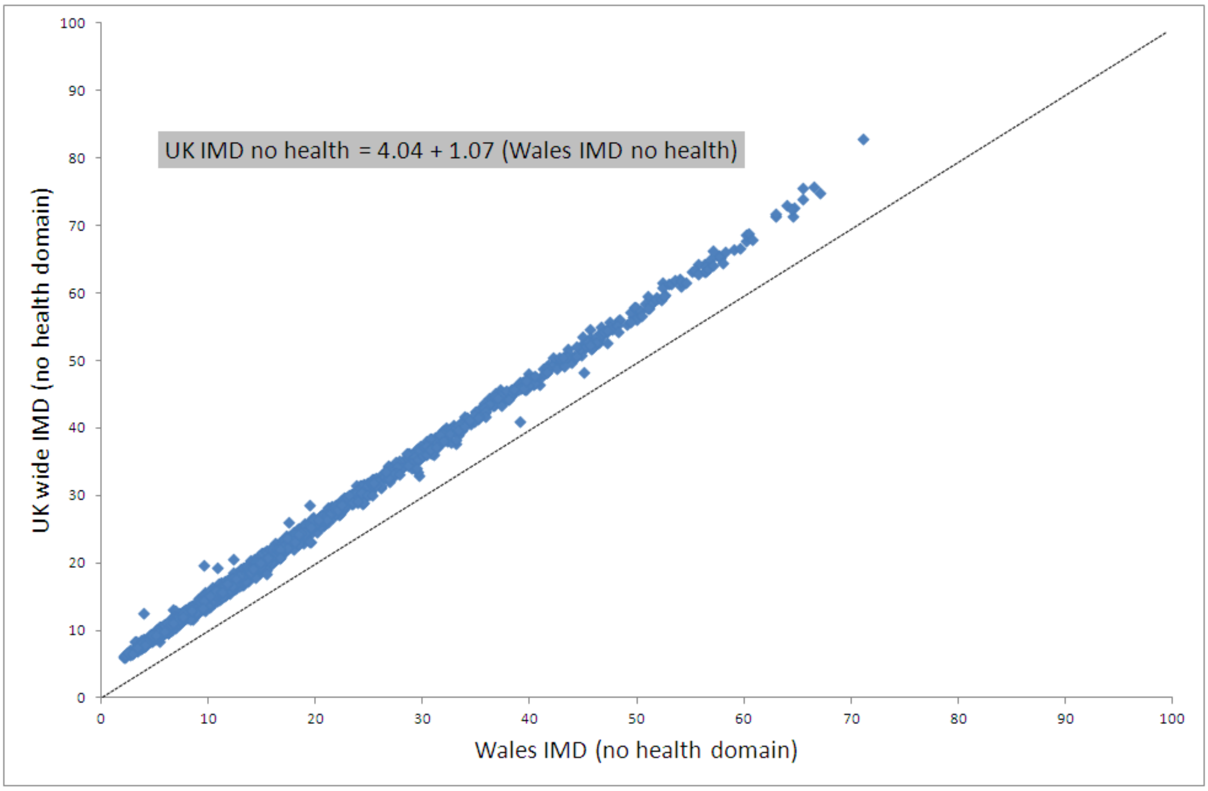


Figure A3 Adjusted versus original results for Wales



Box 1 Additional note on the transformation undertaken

To combine the remaining domains into an overall index the indices needed to be standardised and transformed as they are based on very different units of measurement. The process outlined in McLennan *et al.* (2011) and adapted for the reduced composite indicator described above involved ranking all small areas with one being the least deprived.

Then for each area, the scaled rank R was calculated where the range is [0,1] whereby $R = {}^{1}/_{N}$ for the least deprived and $R = {}^{N}/_{N}$ (in other words R=1) for the most deprived, with N=the number of LSOAs in England.

Each domain (x) was then transformed using equation [1] below:

$$x = -23 \times \ln\left\{1 - R\left[1 - exp^{\left(-100/_{23}\right)}\right]\right\}$$

Equation [1]

The exponential transformation procedure reduces the extent to which lack of deprivation in one domain can cancel the effect of deprivation in another, when combining all the domains into the overall multiple deprivation measure. The domains were then combined using the weighting factors detailed above[5] to create the new composite indicator used throughout this thesis. IMD data are only available at the LSOA level, therefore weighted population averages (based on 2010 mid-year population estimates) of the index were calculated to aggregate the data up to Middle Layer Super Output Areas where applicable.

[5] Inside living environment weight is 0.062 and outside living environment weight 0.031.

Akaike, H. (1974). 'A new look at the statistical model identification'. *IEEE Transactions on Automatic Control*, 19(6), 716-723.

Asthana, S., A. Gibson, G. Moon, P. Brigham and J. Dicker (2004). 'The demographic and social class basis of inequality in self reported morbidity: an exploration using the Health Survey for England'. *Journal of Epidemiology and Community Health*, 58(4), 303-307.

Bauer, D. (2009). 'A note on comparing the estimates of models for cluster-correlated or longitudinal data with binary or ordinal outcomes'. *Psychometrika*, 74(1), 97-105.

Bentham, G., J. Eimermann, R. Haynes, A. Lovett and J. Brainard (1995). 'Limiting long-term illness and its associations with mortality and indicators of social deprivation'. *Journal of Epidemiology and Community Health*, 49(Suppl 2), S57-S64.

Browne, W. (2009a). *MCMC estimation in MLwiN version 2.1*. Bristol, Centre for Multilevel Modelling, University of Bristol.

Browne, W. (2009b). *MCMC estimation in MLwiN version 2.13*. Bristol, Centre for Multilevel Modelling.

Browne, W. (2012). *MCMC estimation in MLwiN version 2.26*. Bristol, Centre for Multilevel Modelling, University of Bristol.

Burnham, K. and D. Anderson (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. New York, Springer.

Duncan, C., K. Jones and G. Moon (1996). 'Health related behaviour in context: a multilevel modelling approach'. *Social Science and Medicine*, 42(6), 817-830.

Duncan, C., K. Jones and G. Moon (1998). 'Context, composition and heterogeneity: Using multilevel models in health research'. *Social Science and Medicine*, 46(1), 97-117.

Field, A. (2005). *Discovering Statistics Using SPSS*. London, Sage.

Fitzpatrick, A. and C. Grant (2011). *2010-11 British Crime Survey (England and Wales) Technical Report Volume I*. London, TNS-BMRB.

Heady, P., P. Clarke, G. Brown, K. Ellis, D. Heasman, S. Hennell, J. Longhurst and B. Mitchell (2003). *Model-based small area estimation series number 2: Small area estimation project report*. London, ONS.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. New Jersey, Lawrence Erlbaum Associates.

Jones, K. (1992). 'Amendment to 'Specifying and estimating multilevel models for geographical research' by Kelvyn Jones (1991) Trans. Inst. Br. Geogr. 16: 148-160'. *Transactions of the Institute of British Geographers*, 17(2), 246.

Jones, K. and N. Bullen (1993). 'A multilevel analysis of the variations in domestic property prices: Southern England, 1980-87'. *Urban Studies*, 30(8), 1409-1426.

Kass, R., B. Carlin, A. Gelman and R. Neal (1998). 'Markov Chain Monte Carlo in practice: A roundtable discussion'. *The American Statistician*, 52(2), 93-100.

Macintyre, S., A. Ellaway and S. Cummins (2002). 'Place effects on health: how can we conceptualise, operationalise and measure them?'. *Social Science and Medicine*, 55(1), 125-139.

Macintyre, S., S. Maciver and A. Sooman (1993). 'Area, class and health - should we be focusing on places or people'. *Journal of Social Policy*, 22, 213-234.

McLennan, D., H. Barnes, M. Noble, J. Davies, E. Garratt and C. Dibben (2011). *The English indices of deprivation 2010*. London, Department for Communities and Local Government.

Mitchell, R., J. Gibbs, H. Tunstall, S. Platt and D. Dorling (2009). 'Factors which nurture geographical resilience in Britain: A mixed methods study'. *Journal of Epidemiology and Community Health*, 63(1), 18-23.

Moon, G., S. Subramanian, K. Jones, C. Duncan and L. Twigg (2005). 'Area-based studies and the evaluation of multilevel influences on health outcomes'. In *Handbook of health research methods: Investigation, measurement and analysis*. A. Bowling (eds). Berkshire, McGraw-Hill Education, 266-292.

Norman, P., P. Boyle and P. Rees (2005). 'Selective migration, health and deprivation: A longitudinal analysis'. *Social Science and Medicine*, 60(12), 2755-2771.

ONS (2011). *Harmonised concepts and questions for social data sources primary standards general health and carers v3.2*. London, ONS.

Payne, R. and G. Abel (2012). *UK indices of multiple deprivation - a way to make comparisons across constituent countries easier*. London, Office for National Statistics.

Pickering, K., S. Scholes and M. Bajekal (2004). *Synthetic estimation of healthy lifestyles indicators: Stage 2 report*. London, National Centre for Social Research.

Raftery, A. and S. Lewis (1992). 'How many iterations in the Gibbs sampler?'. In *Bayesian statistics four*. J. Bernardo, J. Berger, A. Dawid and A. Smith (eds). Oxford, Oxford University Press, 763-773.

Rasbash, J., C. Charlton, W. Browne, M. Healy and B. Cameron (2009). *MLwiN version 2.1*. Bristol, Centre for Multilevel Modelling, University of Bristol.

Riva, M., S. Curtis, L. Gauvin and J. Fagg (2009). 'Unravelling the extent of inequalities in health across urban and rural areas: Evidence from a national sample in England'. *Social Science and Medicine*, 68(4), 654-663.

Riva, M., L. Gauvin and T. A. Barnett (2007). 'Toward the next generation of research into small area effects on health: A synthesis of multilevel investigations published since July 1998'. *Journal of Epidemiology and Community Health*, 61(10), 853-861.

Ross, C. and J. Mirowsky (2008). 'Neighborhood socioeconomic status and health: Context or composition?'. *City and Community*, 7(2), 163-179.

Scarborough, P., S. Allender, M. Rayner and M. Goldacre (2009). 'Validation of model-based estimates (synthetic estimates) of the prevalence of risk factors for coronary heart disease for wards in England'. *Health and Place*, 15(2), 596-605.

Scholes, S., K. Pickering and C. Deverill (2008). *Healthy lifestyle behaviours: Model based estimates for middle layer super output areas and Local Authorities in England, 2003-2005: Stage 1 report*. Leeds, The NHS Information Centre for health and social care.

Shouls, S., P. Congdon and S. Curtis (1996). 'Modelling inequality in reported long term illness in the UK: Combining individual and area characteristics'. *Journal of Epidemiology and Community Health*, 50(3), 366-376.

Silva, D. and P. Clarke (2008). *Some initiatives on combining data to support small area statistics and analytical requirements at ONS-UK*. The International Association for Official Statistics Conference on Reshaping Official Statistics, 14 to 16 October 2008. Shanghai.

Smith, M. and C. White (2009). 'An investigation into the impact of question change on estimates of general health status and healthy life expectancy'. *Health Statistics Quarterly*, 41, 28-41.

Snijders, T. and R. Bosker (2012). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*, SAGE Publications.

Spiegelhalter, D., N. Best, B. Carlin and A. van der Linde (2002). 'Bayesian measures of model complexity and fit'. *Journal of the Royal Statistical Society Series B - Statistical Methodology*, 64(4), 583-639.

Twigg, L., G. Moon and K. Jones (2000). 'Predicting small-area health-related behaviour: A comparison of smoking and drinking indicators'. *Social Science and Medicine*, 50, 1109-1120.

Twigg, L., G. Moon and S. Walker (2004). *The smoking epidemic in England*. London, Health Development Agency.