Dadansoddi ar gyfer Polisi

YMCHWIL GYMDEITHASOL Y LLYWODRAETH

GSR

GOVERNMENT SOCIAL RESEARCH

Analysis for Policy

Llywodraeth Cymru
Welsh Government

Ymchwil gymdeithasol
Social research

www.cymru.gov.uk

Number: 38/2014(R)

Understanding Wales at the neighbourhood level: small area estimation using the National Survey for Wales

(Report revised March 2015)

Understanding Wales at the neighbourhood level: small area estimation using the National Survey for Wales - Revised

Dr Adam Whitworth, Phil Jones and Dr Bethan Thomas
Department of Geography, University of Sheffield

Views expressed in this report are those of the researcher and not necessarily those of the Welsh Government.

For further information please contact:
Chris McGowan
National Survey team, Knowledge and Analytical Services
Welsh Government
Cathays Park
Cardiff
CF10 3NQ
029 2080 1067
Email: chris.mcgowan@wales.gsi.gov.uk

Welsh Government Social Research, 10 March 2015
ISBN 978-1-4734-1754-0

Note

This report was revised and re-published in March 2015, after some errors were identified in the estimates discussed. The errors do not affect the substantive conclusions.

Table of Contents

Glossary 2
List of Figures 2
Executive summary 4
Introduction to small area estimation 7
Statistical approaches to SAE 9
Spatial microsimulation approaches to SAE 10
Explaining the selection of the IPF method for this project 12
From raw data to small area estimates: A step-by-step guide to the IPF method 13
Results 22
Validation 32
Pushing the boundaries of SAE: Future development and potential next steps 38
Producing small area estimates at the LSOA level: a viable progression? 39
Basing small area estimates on smaller national surveys: can it be done and what are the impacts? 43
Looking at the local authority level: estimating to local authority level using smaller base surveys 48
Discussion 53
References 56
Appendix 59

Glossary

CO	Combinatorial optimisation
IPF	Iterative proportional fitting
GREGWT	Generalised regression weighting
LSOA	Lower Layer Super Output Area
SAE	Small area estimation

List of Tables

Table 1: Small Area figures for number of earners (i.e. individuals in paid employment) derived from Census 2011 for the first MSOA in Wales...................... 18
Table 2: Small Area figures for number of earner (i.e. individuals in paid employment) derived from weighted sum of National Survey dummy variables 18
Table 3: First four survey individuals with adjusted weights after fitting to constraint 1 19
Table 4: Model output for each of the six outcome variables to be estimated down to MSOA level 24
Table 5: Internal validation statistics for the IPF 33
Table 6: Mean absolute error between aggregated small area estimates and direct survey estimates across Welsh local authorities 36
Table 7: Results of the MSOA level multilevel models 59
Table 8: Local authority differences between direct estimates and aggregated small area estimates.. 60
Table 9: Residual level two variance in the LSOA level multilevel models 60
Table 10: Overall local authority distances between central direct National Survey estimate and indirect IPF estimates from smaller base surveys............................... 61

List of Figures

Figure 1: Main methodological approaches to small area estimation 8
Figure 2: Overview of the key steps in the IPF method 14
Figure 3: Percentage of Welsh adults saying that they have… 16
Figure 4: MSOA estimates of percentage of adults satisfied with their local area (systematic 10% sample of MSOAs shown).. 26
Figure 5: MSOA estimates of percentage of adults experiencing financial difficulties (systematic 10% sample of MSOAs shown).. 26
Figure 6: MSOA estimates of percentage of adults who feel unsafe (systematic 10% sample of MSOAs shown) 27

Figure 7: MSOA estimates of percentage of adults satisfied with GP care (systematic 10% sample of MSOAs shown) 27
Figure 8: MSOA estimates of percentage of adults using the internet (systematic 10% sample of MSOAs shown) 28
Figure 9: MSOA estimates of percentage of adults having high satisfaction with performance of Welsh Government (systematic 10% sample of MSOAs shown) 28
Figure 10: MSOA small area estimates for Cardiff 29
Figure 11: MSOA small area estimates for Ceredigion 30
Figure 12: MSOA small area estimates for Wrexham 31
Figure 13: External validation of the aggregated small area estimates against local authority direct estimates 35
Figure 14: Percentage of adults estimated to be internet users across LSOAs in Cardiff 40
Figure 15: Percentage of adults estimated to be internet users across MSOAs in Cardiff 40
Figure 16: Percentage of adults estimated to be internet users across LSOAs in Cardiff (within Cardiff quintiles) 41
Figure 17: Credible intervals around Cardiff LSOA estimates of the percentage of adults using the internet 42
Figure 18: Credible intervals around Cardiff LSOA estimates of the percentage of adults experiencing financial difficulties 42
Figure 19: The impact of sample size on central point estimates of the percentage of adults using the internet in LSOA14 44
Figure 20: The impact of sample size on central point estimates of the percentage of adults feeling unsafe in LSOA75 44
Figure 21: Sensitivity testing the impact of reduced base survey size on the credible intervals around the LSOA estimates of internet use 47
Figure 22: Sensitivity testing the impact of reduced base survey size on the credible intervals around the LSOA estimates of feeling unsafe 47
Figure 23: Direct versus indirect estimates for Blaenau Gwent 49
Figure 24: Direct versus indirect estimates for Wrexham 50
Figure 25: A summary of direct vs indirect estimates across all Welsh local authorities 51

Executive summary

1. Small area estimation (SAE) describes a range of alternative methodological techniques for the estimation of survey data down to small area level where those data do not currently exist at these small spatial scales. Various such variables might need to be estimated at small area level in the UK context including income, healthy lifestyles, digital engagement and well-being, to name but a few. SAE is demanded increasingly by researchers and policy makers who are seeking ever more spatial detail to their knowledge of populations so as to better guide the spatial design and targeting of interventions and resource allocations.

2. The aim of this project was to produce small area estimates for six diverse outcome variables using the National Survey for Wales 2012-13 and, through doing so, to also offer recommendations and considerations to guide future work.

3. To achieve these aims the report provides an overview of the two main methodological frameworks to conduct SAE – statistical approaches and spatial microsimulation approaches – as well as a summary of the main specific methodological techniques within each of these two broad frameworks. The report also provides a detailed step-by-step account of the Iterative Proportional Fitting (IPF) process that is used in this project. This allows a full understanding of the IPF methodology so that the process and the decisions taken within it can be transparently understood and, if desired, applied again in future.

4. The project has produced small area estimates with accompanying 95% credible intervals for the 410 Middle Layer Super Output Areas (MSOAs) across Wales for six diverse outcome variables. These show the estimated percentage of adults aged 16+ across Welsh MSOAs who: use the internet; are experiencing financial difficulties; feel unsafe in the local area after dark; are satisfied with their GP care; are highly satisfied with their local area; and are highly satisfied with the performance of the Welsh Government. The credible intervals around the central MSOA estimates provide a sense of the uncertainty around the point estimates and indicate the range that the ‘true’, but unknown, underlying population value likely falls.

5. The IPF validates well in terms of the internal validation, though this would be expected given the nature of the IPF methodology. The external validation of the aggregated small area estimates relies on a comparison between direct survey estimates for local authorities derived from the National Survey and the indirect MSOA small area estimates aggregated to the local authority level. Scatterplots show that the two sets of estimates are, in general terms, comparable although they show inevitable variation around the line of equality between the two sets of estimates. Statistical analysis of the mean absolute error of the estimates shows that the aggregated small area estimates are on average between roughly two and four percentage points of the local authority estimates taken directly from the National Survey. Most local authorities show smaller differences than this and a minority of local authorities with larger differences pull these mean differences upwards. Although this appears a reasonable degree of accuracy at this scale there are no benchmarks in the literature as to what constitutes acceptable fit;

this is largely a subjective decision based on the use to which the analyst or policy maker is relying on the estimates and the extent to which certainty around precision is important to the policy task at hand.

6. Testing within Cardiff local authority suggests that small area estimation down to the smaller Lower Layer Super Output Area (LSOA) scale is viable. The multilevel regression models required to create the credible intervals around these LSOA estimates would however require a relatively large base survey file – in the order of 7,500 cases or more assuming a similar strategy to that used in the National Survey 2012-13.

7. The local authority analyses highlight the potential for using IPF to create local authority level estimates from smaller base surveys. These results are benchmarked against direct estimates from the National Survey and analyses are presented of the extent of correspondence between these direct survey estimates and indirect IPF estimates. Whilst there is inevitably variation across outcomes, and whilst the decision around how much error is considered to be acceptable is inevitably a subjective and context-specific one, the results are in general supportive of the idea that local authority level estimation from smaller base surveys is a viable proposition. Further work exploring the potential to produce more contextually specific estimates would be beneficial.

8. In terms of potential developments and extensions for future SAE work to consider, several possibilities emerge:
    - Firstly, as with virtually all small estimation projects it is difficult to assess the extent to which the *distributions* of the small area estimates *within* the validated local authority level are accurate. Certainly one would expect greater variability at that smaller scale. In terms of future potential SAE work, one possibility would be to conduct surveys in specific local areas in order to be able to calculate sufficiently precise direct estimates at the small area level against which to externally validate the small area estimates at this small area scale. This would represent the gold standard test in terms of the external validation of these small area estimates and of the IPF methodology more broadly. Alternatively, one might test the IPF by creating small area estimates of a variable that is available directly at small area level from Census or administrative data (e.g. long-term limiting illness) and then comparing the estimates against those known values;
    - Secondly, the external validation highlights the better fit of the aggregated estimates in some local authorities compared to others, but the reasons for these differences are at present unclear. A better understanding of these issues may help to identify the factors that affect the accuracy of the estimation process across differing local contexts at the small area level so as to be able to produce better estimates in (potentially atypical) areas where the models may not be as effective at present;
    - Thirdly, targeted testing around potential future developments suggests that it is viable to produce estimates at the smaller Lower Layer Super Output Area (LSOA) scale and that it is viable to produce acceptable estimates based on smaller base surveys. Sensitivity testing demonstrates the likely impacts of such changes on the expected validity of any resulting

estimates and the acceptability of such estimation would be a subjective decision based on the use for which estimates are needed and the degree of precision that is therefore required;

- Finally, whilst IPF approaches typically rely exclusively on individual level constraint variables future work could explore the combination of individual and area level factors in the IPF process. Depending on the nature of the constraint and outcome variables involved, this would be expected to have benefits both for the accuracy of the central point estimates as well as for reductions in the width of the credible intervals.

Introduction to small area estimation

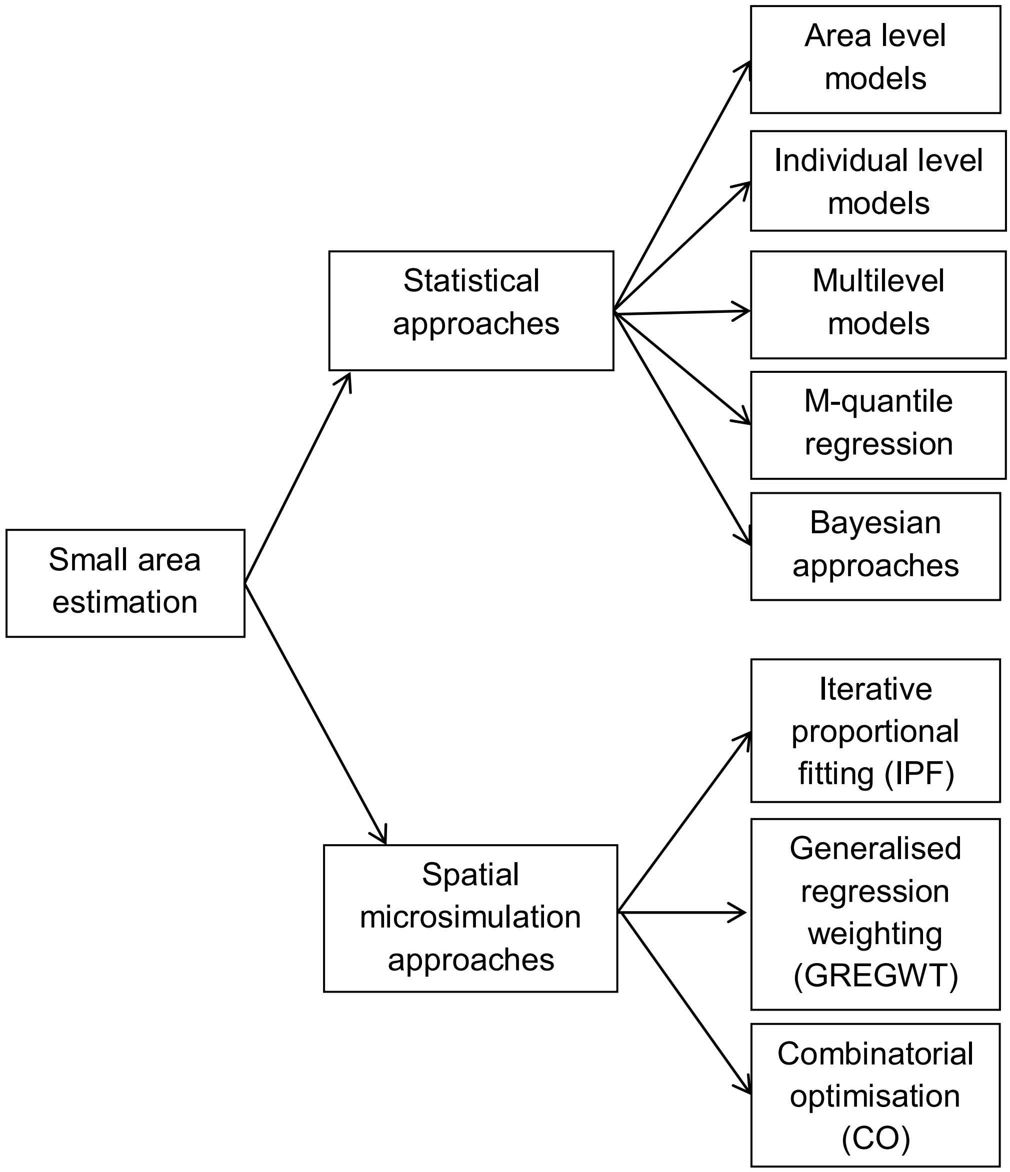
The primary aim of this project is to assess the viability of producing small area estimates of six diverse outcome variables that exist within the National Survey for Wales but which are not available at small area level across Wales. In doing so the project also offers broader guidance and recommendations in order to support the Welsh Government to carry out potential further work with small area estimation in the future.

Small area estimation (SAE) at its most basic level is a methodological approach to estimate data at small area level where those data do not currently exist. Various such variables might need to be estimates down in the UK context including, for example, income, healthy lifestyles, digital engagement and well-being to name but a few. As a next step, SAE might also be used to estimate the spatial impacts of a potential or actual policy change at small area level by simulating not just the impact on *people* of the policy change but, additionally, the *spatial* nature and distribution of those effects. As a result, SAE is demanded increasingly by academics and policy makers searching for finer spatial detail to their knowledge and understanding about the nature of their populations, the targeting of interventions and resources, or the effects of those interventions.

As discussed below, a variety of alternative SAE techniques exist but all share common principles. A first step is to identify (and quantify) relationships in survey datasets between the outcome variable of interest and a set of explanatory variables. As a second step these relationships are then applied to a small area level geography at which the same set of explanatory variables – but not the target outcome variable – exist.

Figure 1 below presents a visual map of the main methodological approaches to SAE and these can be separated at a broad level between statistical approaches and spatial microsimulation approaches. As Figure 1 shows, whilst statistical and spatial microsimulation approaches represent two distinct 'broad churches' in terms of SAE techniques it is also possible to identify different specific SAE methods within each. A brief summary of these main methodological approaches to SAE is provided below in order to provide some context to this project's use of the iterative proportional fitting (IPF) methodology. Interested readers seeking greater detail on these methods are directed to several excellent existing summaries (Rao, 2003; Bajekal et al., 2004; Ballas et al., 2006; Rahman, 2008; Marshall, 2012; Whitworth, 2013).

Figure 1: Main methodological approaches to small area estimation



Statistical approaches to SAE

Statistical approaches to SAE follow from the basic properties of multiple regression modelling. They make use of the ability within such models to use estimates of the relationships between the outcome variable and key explanatory variables to calculate predicted values of the outcome variable for each case. That basic intuition is used widely for a wide range of purposes including the imputation of missing data as well as for the identification of organisations (e.g. firms, schools, hospitals, police forces, etc.) that seem to be either exemplars of best practice (actual outcomes far better than the values predicted given their characteristics) or, conversely, organisations in potential need of attention or intervention (actual outcomes far worse than the values predicted given their characteristics).

In the context of SAE the principle of regression-based prediction remains the same. The difference, however, is that one applies the regression in one dataset (the survey) and applies the estimated coefficients not to the cases within that same survey but instead to values of an identical set of explanatory variables collected for each target small area (usually sourced from Census or administrative data). In doing so, regression-based SAE can predict estimated values for the outcome variable of interest alongside confidence intervals around those estimates. Though all rooted in that basic statistical framework, in practice a range of alternative implementations of the statistical approach to SAE are applied in the literature and these differ in their degree of complexity, likely accuracy and level of data requirements.

Perhaps the most intuitive approach is to run individual-level regression models in the survey so as to identify the most powerful and parsimonious model and then to apply the estimated coefficients to individual-level data for each small area. As with all SAE approaches, this necessitates first of all that all explanatory variables can be coded in the same way in both the survey dataset and the small area covariate data. For individual-level models, however, one also requires individual-level information at the small area level. Unfortunately, this is often difficult to obtain. Typically, SAE uses Census cross-tabulated tables as its source of covariate data for the small areas. For individual-level modelling approaches to SAE, however, it is necessary to have all predictor variables within a single Census cross-tabulation. Due to confidentiality concerns such data are not routinely published and are often difficult to access directly from central statistical agencies. Alternative possibilities for the covariate data do exist, but are imperfect: to seek access to the Census microdata directly (this is difficult to access); to use more readily accessible samples of Census microdata (this lacks spatial comprehensiveness); to commission a bespoke multi-way table from the central statistical agency (potentially possible, but costly); or to estimate oneself a multi-way table (possible, but just an estimate). Finally, even if this is possible there is a risk with individual-level statistical approaches of assuming that the individual-level outcomes can be explained only by individual-level factors. In practice, however, variables at a range of scales besides the individual level (e.g. the area level) may be of relevance in shaping those outcomes.

A second statistical approach is to instead apply area-level models to predict the area-level outcomes by using aggregate values of the explanatory variables in the modelling. Inevitably, a weakness of such area-level models is that they ignore individual-level factors that shape the area outcomes or, indeed, any interactions between individual and area level explanatory factors. Moreover, with ecological models one can be vulnerable to risks of ecological fallacy in which results obtained at one scale (area-level analyses in this case) are taken to be acceptable to model processes that may be occurring at a different scale (often the individual level given the types of outcomes that SAE tends to model). Such cross-scale interpretation can be misleading.

Consequently, a third statistical approach to SAE seeks to combine the advantages, and minimise the weaknesses, of either purely individual-level or area-level models by combining the two within a multilevel regression framework. Multilevel approaches do naturally rely on covariate data being available at individual and area levels, a constraint often not met due to the limited availability of individual level microdata at the small area level. Where possible, however, multilevel approaches are well suited to incorporating both individual and area level factors. Multilevel models are able to recognise the hierarchical nature of much of the survey data from which small area estimates are derived so as to ensure more accurate estimation of coefficients and standard errors. Multilevel models offer the possibility to explore cross-level interactions between variables and can naturally be extended to incorporate three or more levels within the multilevel structure as required. Multilevel regression approaches thus offer an attractive and flexible statistical framework for SAE and have been widely used (Heady et al., 2003; Pickering et al., 2004; Haughton and Haughton, 2011; Whitworth, 2012).

In addition, more advanced, specific or recent statistical approaches to SAE can also be identified. Amongst these, of particular note perhaps has been activity around Bayesian approaches to SAE (Ghosh and Rao 1994; Gomez-Rubio et al. 2010; Molina and Rao 2010) and in M-quantile approaches. Bayesian approaches to SAE follow naturally from the Bayesian framework of statistical analysis more broadly and Gomez-Rubio et al (2010: 3) propose several potential advantages of a Bayesian approach to SAE. M-quantile regression in contrast focuses on the specific desire to model not just point estimates of the outcome variable to small area level (typically mean or median estimates) but instead to produce estimates of the distribution of the target outcome variable. Although demanding in terms of its need for individual-level covariate data for the small areas, M-quantile models offer the potential for greater distributional informational about the outcome variable in each small area in addition the more usual estimation of mean or median point estimates only (Chambers and Tzavidis 2006; Tzavidis et al. 2010; Marchetti et al. 2012).

Spatial microsimulation approaches to SAE

In contrast to the statistical approaches, spatial microsimulation techniques offer a second strand of methodological techniques for SAE. Spatial microsimulation essentially involves ‘fitting’ survey individuals or households to best match the population profile of each target small area (Edwards and Tanton, 2012; Hermes and Poulson, 2012). The details of how this is achieved, however, differ across the three main techniques – iterative proportional fitting (IPF), generalised regression weighting (GREGWT) and combinatorial optimisation (CO).

IPF and GREGWT act to reweight all survey cases such that they come to optimally match the profile of small area totals across a selected set of characteristics (sex, age, tenure, education, employment status, and so on) that are collected for each target small area (e.g. every MSOA in Wales in this project). The small area totals across these characteristics act as constraints that the reweighting of the survey cases seeks to fit to and these variables are therefore typically referred to as the constraint variables in the literature. The constraints used are selected based on the strength of their relationships with the target outcome variable to be estimated and, as a result, may be understood as similar to explanatory variables within a regression model.

Both IPF and GREGWT are deterministic methods in that given the same data they will each produce the same results each time the code is run. However, the way that the two methods operate, and hence the results that they reach, are somewhat different. IPF – the method used in this project – reweights the survey cases sequentially over each constraint variable in turn so that the weights attached to each survey case are gradually refined until such point that the process reaches stability – what is referred to in the literature as 'convergence' (Anderson, 2007; Ballas et al., 2012). At the end of this process the final weighted survey individuals are taken as the best possible match to the small area population profile across the set of selected constraints. In contrast, GREGWT seeks to optimise the weights in one step using matrix algebra (Tanton et al., 2009; Tanton and Vidyattama, 2010; Tanton et al., 2011), an approach originally developed by survey statisticians for the purposes of creating survey weights to gross up to national totals (Deville and Sarndal, 1992).

In contrast, the third main spatial microsimulation approach, combinatorial optimisation (CO), operates not by reweighting survey cases to best match the small area totals across the constraints but, instead, by selecting the optimal set of survey cases to achieve this (Voas and Williamson, 2000; Williamson, 2013). Hence, in CO the 'correct' number of survey cases are drawn from the survey – i.e. 2500 survey cases are drawn if there are that many individuals in the small area according to Census or administrative data. Survey cases are then randomly swapped with cases not yet selected in an attempt to optimise the fit between the characteristics of the cases selected and the characteristics of the small area. During this process of large-scale, automated swapping of the survey cases a decision is taken for each swap as to whether to keep the case already selected or, if it improves the fit, to keep the newly selected case and to return the previously selected case back to the pool of unselected cases. A range of algorithms can be used to guide this automated decision (e.g. hill climbing, simulated annealing, genetic algorithms). Unlike IPF and GREGWT, CO is a probabilistic method in the sense that randomness is involved in the selection and swapping of survey cases. As a consequence, the method will not produce the same results each time even if repeated with the same data.

Whichever of the three spatial microsimulation techniques is adopted, the end result is a micro-level (i.e. individual or household) synthetic population dataset for each small area (i.e. each Welsh MSOA in this project). Within the dataset, the survey cases are either optimally reweighted (IPF and GREGWT) or optimally selected (CO) to best fit the multi-dimensional population profile of that small area. The small area estimate of the target outcome variable in the survey can then be readily picked off from these reweighted/selected survey cases: in IPF, for example, one can use the reweighted survey cases to calculate a weighted mean, median or, indeed, various points from the distribution of the outcome variable as estimates of the outcome variable of interest for each small area.

Explaining the selection of the IPF method for this project

There is no clear agreement in the literature as to which of these various SAE methods is best. Indeed, it may be possible that the alternative methods are differently effective across alternative spatial contexts – dense urban areas, sparse rural areas, unusual small areas, and so on. This is an area of remaining uncertainty and recent academic activity funded by the Economic and Social Research Council's (ESRC's) National Centre for Research Methods (NCRM) has brought together experts from across these various SAE methodologies to seek to progress our comparative methodological understanding (Whitworth, 2013). Ongoing research seeks to further explore the nature and impacts of the methodological linkages, similarities and differences between the alternative SAE methodologies and the implications of this for performance improvements in SAE.

At present, however, the separate SAE approaches are best understood as operating in parallel to one another, with methods generally treated as different but equal. In this project the IPF method has been selected for four main reasons. Firstly, IPF is a widely used, well understood and flexible methodological approach. Secondly, IPF has already been used specifically within policy focussed research in the UK context (Ballas et al., 2005; Anderson, 2007). Thirdly, IPF is perhaps the most inclusive method to use in that it does not demand significant technical or statistical expertise of the reader to understand. Rather, the central idea within IPF of adjusting survey weights in order to make survey cases better 'fit' to small area totals is a relatively intuitive one to grasp. Moreover, the actual method itself is relatively non-technical in terms of its calculations and steps such that it is readily digestible for a broad, non-specialist audience. Finally, IPF estimates can be delivered with accompanying credible intervals through the incorporation of the residual variance on the area level error term within the preparatory regression modelling. For these reasons IPF is an effective, flexible and inclusive SAE methodology to adopt for this project.

From raw data to small area estimates: A step-by-step guide to the IPF method

As noted above, of the various SAE methods outlined in Section 2 this project makes use of the IPF approach to SAE in order to estimate six target outcome variables of interest from within the National Survey for Wales down to small area level across the country. The survey data used are the National Survey 2012-13 results, provided by the Welsh Government. The small area covariate data come from the UK Census 2011 and are sourced from the NOMIS data portal website[1]. After cleaning and recoding the National Survey provides a base survey of 14,362 cases for the IPF.

The project is designed to produce estimates at sub-local authority level. Given that one wants statistical geographies of roughly equal population size for this purpose, this means that three alternative small area scales are in principle possible to estimate to: Middle Layer Super Output Area (MSOA); Lower Layer Super Output Area (LSOA); and Output Area (OA). The average population size of MSOAs in England and Wales is 7,860, for LSOAs it is 1,630 and for OAs it is 309 (ONS, 2012; 2013). As the scale of estimation shrinks the degree of uncertainty and possible imprecision around the estimates tends to rise, indicated by the width of confidence intervals around the point estimates. It is not considered viable to estimate down to the very small OA level and this leaves either MSOA or LSOA scales for consideration. Both are relatively small scale geographies that offer considerable spatial detail to results. Based on previous experience and knowledge of the literature, MSOAs are preferred given the expectation of smaller levels of uncertainty around the central estimates. As part of the project, however, the final section of this report describes the results of case study testing of the viability of producing small area estimates at LSOA scale in future.

Figure 2 below presents a visual summary of the various steps in the IPF methodology used in the project. These steps are discussed in greater detail below in order to provide a clear overview of the small area estimation process followed and to allow for possible replication in future on other target outcome variables.

---

[1] http://www.nomisweb.co.uk/census/2011

Figure 2: Overview of the key steps in the IPF method

Identify potential target outcome variables in National Survey

↓

Identify covariates that exist in both National Survey and at MSOA level

↓

Prepare covariates identically in both National Survey and at MSOA level

↓

Run regression models in National Survey of potential outcome variables with the covariate variables to:

(i) Select a parsimonious set of powerful covariate variables for the IPF

(ii) Select outcome variables that can be estimated with reasonable predictive power

↓

Run IPF for each MSOA in turn in order to calculate new weights for National Survey individuals, linking them to that MSOA based on their characteristics

↓

Pick off small area estimate of target outcome variable(s) from the reweighted National Survey cases

↓

Calculate credible intervals around the MSOA point estimates

Step 1: Identify potential target outcome variables in the survey data

The first step in the process is to draw up a shortlist of potential target outcome variables from the survey that in principle might be estimated down to small area level. This initial shortlisting of candidate outcome variables will usually reflect the policy or informational priorities of the researcher or project sponsor as well as the nature of the data and model testing. In this project various potential outcome variables were assessed and the following six target outcome variables were selected as covering a range of key topics within the survey.

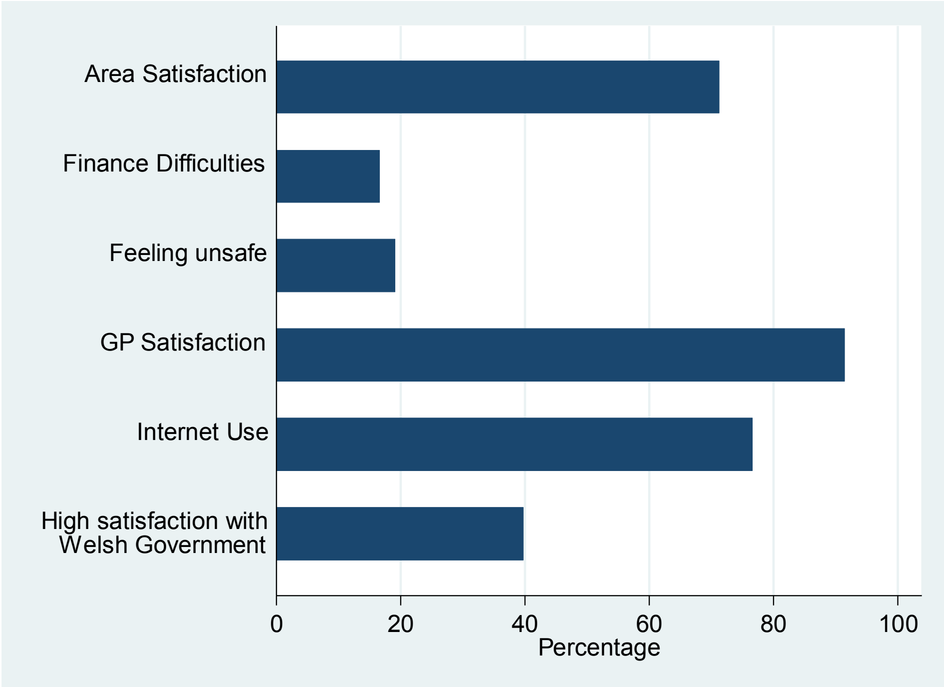
Preliminary analyses of these variables suggested that each of these variables should be transformed to binary outcome variables for the purposes of the small area estimation due to a combination of the ordinal nature of the data collected and because of skewed distributions across the continuous variables in which responses were concentrated into only a few responses in the scale. In the new binary outcome variables, it is the estimated number of Welsh adults in the group that are coded with the value 1 (e.g. using the internet, being satisfied with the local area) and that that will be estimated to small area level. Survey cases not taking this characteristic (e.g. not using the internet, not speaking Welsh, etc) are coded with the value 0 on the binary outcome variable. All survey cases with missing values on the outcome variable must be dropped prior to the IPF.

The six outcome variables selected, and the recoding of these outcome variables, is as follows:

- **area satisfaction** (level of satisfaction with the area lived in) was originally coded on a 0-10 scale with 10 being highest area satisfaction and 0 being lowest area satisfaction. Values 0-7 were recoded as 0 relating to 'low to moderate area satisfaction' and values 8-10 were recoded as 1 relating to 'high area satisfaction';
- **financial difficulties** was originally coded on a 1-5 scale with 1 relating to 'having no difficulties keeping up with bills' and 5 relating to 'having constant difficulties keeping up with bills'. Values 1-2 were recoded as 0 relating to 'no significant financial difficulties' and values 3-5 were recoded as 1 relating to 'significant financial difficulties';
- **feeling unsafe** (feeling safe walking in local area after dark) was originally coded on a 1-4 scale ranging from 1 relating to 'feeling very safe' and 4 relating to 'feeling very unsafe'. Values 1-2 were recoded as 0 relating to 'feeling safe' and values 3-4 were recoded to 1 relating to 'feeling unsafe;
- **satisfaction with GP care** at last appointment was originally coded on a 1-5 scale with 1 relating to 'very satisfied' and 5 relating to 'very dissatisfied'. Values 1-2 were recoded as 1 relating to 'satisfied with GP care' and values 3-5 were recoded as 0 relating to 'dissatisfied or neutral about GP care';
- **internet use** was originally collected as a binary variable with 1 relating to 'I do use the internet' and 0 relating to 'I do not use the internet'. This coding was retained;
- **overall satisfaction with how the Welsh Government is doing its job** was originally coded on a 0-10 scale with 10 being highest satisfaction and 0 being lowest satisfaction with the way in which Welsh Government is doing its job. Values 0-6 were recoded as 0 relating to 'low to moderate satisfaction with how the Welsh Government is doing its job' and values 7-10 were recoded as 1 relating to 'high satisfaction with how the Welsh Government is doing its job'.

Figure 3 below presents weighted estimates of the percentage of all Welsh adults according to the National Survey who agree with each question (i.e. who feel satisfied with the area, who have financial problems, who feel unsafe, who feel satisfied with GP care, who use the internet, and who have high satisfaction with Welsh Government performance).

Figure 3: Percentage of Welsh adults saying that they have…



Step 2: Identify potential covariate variables that exist in both the survey and small area data

The essential idea of any SAE method is to quantify relationships between covariate/explanatory data and a target outcome variable in a survey and then to apply those quantified relationships to the same set of covariates at small area level where data on the outcome of interest do not exist. Whichever method of small area estimation is used, one essential ingredient therefore is to have the same explanatory variables available in both the survey data of interest and at the target small area level desired. For this project, therefore, covariates from the National Survey can only be included in the IPF if aggregated counts of those covariates can also be found at MSOA level.

Step 3: Prepare covariate data identically in both the survey data and MSOA covariate data

In addition to the need for covariates to exist in both the survey data and at small area level it is also necessary that these two sets of covariates are able to be coded in exactly the same way in both datasets. This is necessary because the quantified relationships between the covariates and outcome variable(s) found in the survey dataset occur only on a particular specification of those variables. It is therefore appropriate to apply those modelled relationships only to the same specification of those same covariates at the small area level.

Step 4: Run regression models in the survey data of each potential outcome variable on the set of possible covariate variables

At this stage regression models are fitted to each of the potential outcome variables to be estimated at small area level. These models are gradually developed, adjusted and tested against the range of explanatory (i.e. potential constraint) variables.

Unlike statistical approaches to SAE, although spatial microsimulation approaches to SAE do not use regression models for the actual estimation process itself the regression modelling still fulfils two functions in the preparation of the IPF. Firstly, the models are used to identify a parsimonious set of explanatory variables that offer predictive power in relation to each outcome variable – these covariates then become the constraint variables used within the IPF process. Secondly, the power of these models (assessed typically via the R^2 or, in logit models, pseudo-R^2 values) provides information about the strength of the modelled relationships seen. This gives an initial guide as to likely accuracy of the final small area estimates, even if model power is not the key driver of the IPF's effectiveness and it is the validation process that offers the key insights into the success of the SAE.

Step 5: Run IPF for each MSOA in turn in order to calculate new weights for National Survey respondents linking them to that MSOA based on their characteristics

At this point the analyst is clear about which target outcome variables are most suitable for small area estimation and about which explanatory variables together offer a parsimonious set of constraints for the IPF. In terms of data preparation, the survey dataset is set up with all constraints are coded as binary dummy variables[2] and with all categories coded as dummy variables. This is in slight contrast to the preparation of dummy variables for regression modelling in which the reference category is not coded into dummy variables; in contrast, for the purposes of the IPF *all* categories are coded into binary dummy variables. These exhaustive dummy variables become the constraints within the IPF process.

In terms of the small area data, this dataset is at area-level and shows small area (e.g. MSOA) total counts for each of those dummy constraint variables. For example, the survey dataset might contain a binary dummy variable relating to 'male' and a binary dummy relating to 'female', with men coded one on the former and zero on the later (and vice versa for females). In the small area MSOA dataset the corresponding variables would be one variable called 'male' giving the total number of males in the small area and another variable called 'female' giving the total number of females in the MSOA.

[2] A dummy variable is a binary variable coded 1 if the case takes that characteristic and a value of 0 otherwise (apart from missing values which remain coded as missing). For example, female survey respondents would be coded 1 on a dummy variable called 'female' whilst all men would be coded 0.

IPF relies on the process of adjusting cell totals for small area tables given known marginal (row/column) totals derived from Census or other small area data sources. The IPF takes each small area in turn and 'fits' the survey cases as effectively as possible to the multi-dimensional profile of the MSOA across each of the constraint variables. It does this by sequentially reweighting the survey cases across each constraint in turn based on the extent to which the (re)weighted sum of each constraint in the survey file matches the small area total for that constraint.

Ballas and Anderson (in Whitworth, 2013) provide a worked example of the IPF methodology and an adjusted version of this example is shown below. IPF requires two sets of tables for each constraint and each small area: the Census 2011 (or otherwise sourced) small area tables of totals for the constraints (e.g. Table 1: Small Area figures for number of earners (i.e. individuals in paid employment) derived from Census 2011 for the first MSOA in Wales) and the analogous small area tables constructed by calculating a (re)weighted sum of the dummy variables across the National Survey individuals (e.g. Table 2).

It is necessary that all Census (or otherwise sourced) small area totals sum to the same population totals. In this project the Census 2011 data for Welsh MSOAs sourced from NOMIS contained some minor variation across the different Census tables in the total number of individuals in the MSOA that they summed to. It was necessary therefore to adjust the MSOA constraint variables from the Census tables such that they summed to the same value. To achieve this the MSOA population according to the simple age-sex band Census table was taken as the 'true' MSOA population value and all other Census constraints were adjusted to meet this.

Table 1: Small Area figures for number of earners (i.e. individuals in paid employment) derived from Census 2011 for the first MSOA in Wales

MSOA	Number of individuals	Number of earners = 0	Number of earners = 1	Number of earners = 2	Number of earners = 3+
MSOA1	7840	3970	2210	1420	240

Table 2: Small Area figures for number of earner (i.e. individuals in paid employment) derived from weighted sum of National Survey dummy variables

MSOA	Number of individuals	Number of earners = 0	Number of earners = 1	Number of earners = 2	Number of earners = 3+
MSOA1	12310	5440	3260	3090	520

The IPF process gradually refines the weights attached to the National Survey respondents such that these two sets of totals come to match each other as closely as possible. Intuitively, one can understand in the example tables above that the survey data has 'too many' people in it both in terms of the total number of individuals for MSOA1 as well as in in each of the earner categories: the MSOA that individuals for MSOA1 as well as in in each of the earner categories: the MSOA that we are seeking to fit to has only 7,840 residents but there are 12,310 people in the survey dataset and the MSOA has fewer individuals in each earner category than its equivalent earner category in the survey dataset. In this example, the survey cases therefore need to be down-weighted. More specifically, some earner categories need to be down-weighted more than others because the 'gaps' (i.e. fractions or ratios) between the survey and MSOA totals are not equal across these four earner categories.

To begin the IPF process all individuals are given their adult weight as provided within the National Survey. These weights enable the survey respondents to reflect, and gross up to, the total adult population of Wales. For each constraint in turn the weights for each individual in the survey are then adjusted using the formula below:

*New weight = Previous Weight ** (*MSOA constraint total / Weighted survey constraint total*)

Table 3 shows a worked example of the calculations performed by this formula for this first constraint – the number of earners. For each survey case the original survey weight is adjusted so that the survey sample fits the Census data on this one dimension and a new weight is calculated. Given that there are 'too many' individuals in each of these earner categories in the survey then the effect of the reweighting is to reduce the size of the weights and, more specifically, to reduce them to differing degrees dependent on their distance from the Census small area total for each earner category.

This new weight then becomes the starting weight for the fitting on the next constraint. The weighted survey constraint total for this second constraining step is based on the new weights produced in the prior constraining step, not on the original survey weights.

This process continues until the IPF has passed over all of the constraints once. Each constraining step refines the weights further so that the reweighted survey cases gradually become a more optimal fit to the MSOA constraint totals across the multi-dimensional set of constraints.

Table 3: First four survey individuals with adjusted weights after fitting to constraint 1

Survey Case	Number of earners	Initial survey weight	New weight after fitting to constraint 1 (number of earners)
1	1	51.2	= 51.2 * (2210/3260) = 34.7
2	0	76.3	= 76.3 * (3970/5440) = 55.7
3	2	33.7	= 33.7 * (1420/3090) = 15.5
4	1	125.3	= 125.3 * (2210/3260) = 84.9
..	..	…	..

Once the IPF has passed over all the constraints once the process loops back to constraint one and moves sequentially over all of the constraints for a second time. There is no agreement within the literature about how many times the IPF should be set to iterate around the full set of constraints, with Ballas et al (2005) recommending 5-10 times and Anderson (2007) suggesting 20 times. In this project we found that 10 iterations were enough to produce stable weights.

For each MSOA the end result is a reweighted version of the National Survey data file with adjusted weights such that the reweighted survey individuals represent the 'fractional existence' of that kind of individual in that MSOA. The reweighted survey file as a whole can be understood as a synthetic population micro-dataset for the MSOA given that it is fitted on, or close to, the MSOA totals across the whole set of constraint variables. Once this had been achieved the results are saved and the IPF process is repeated for the next MSOA.

Step 6: Pick off small area estimate of target outcome variable(s) from the reweighted National Survey cases

Having completed the re-weighting process, calculating the small area estimate from each MSOA is achieved by picking off weighted values of the target outcome variable from the reweighted National Survey cases for that MSOA. The most commonly calculated small area estimates are weighted sums (e.g. the total number of individuals in the MSOA estimated to feel unsafe), weighted mean values (e.g. the average income level of the MSOA) or weighted median values (e.g. the median income level of the MSOA).

As described above in Step 1, in this project all target outcome variables were binary variables and weighted sums were therefore calculated so as to given an estimate of the total number of individuals in the MSOA affected. Given that the populations are known for each MSOA, these sums can easily be expressed as the percentage of each MSOA's residents that are estimated to be affected.

Step 7: Calculate credible intervals around the MSOA point estimates

Typically, the IPF process calculates a point estimate of each target outcome variable for each MSOA in Wales – typically (though not necessarily) a weighted mean, median or sum. Unlike statistical approaches to SAE, however, spatial microsimulation approaches to SAE (whether IPF, GREGWT or combinatorial optimisation) do not tend to also calculate confidence intervals around these point estimates, despite the recognition that the small area estimates are just that – estimates with uncertainty around them.

For this project it was considered important to provide a sense of uncertainty around the point estimates. Although statistical and spatial microsimulation approaches to small area estimation work typically exist separately from one another, this project links the two through a hybrid statistical-spatial microsimulation approach to the calculation of credible intervals around the IPF central point estimates, making use of the preparatory regression modelling used to identify the constraints for the IPF. Developing an approach used to calculate credible intervals within the statistical literature (Heady et al., 2003; Bajekal et al., 2004; Pickering et al., 2004), the estimates of the area level uncertainty as estimated within a multilevel regression model can be understood to reflect the level of unexplained variance in the model at the area level. Step 4 above discusses the use of a multiple regression model to identify the appropriate constraint variables for the IPF and the levels of R^2 (or, in logit models such as these, pseudo-R^2) in these models provide an indication of the overall explanatory power of these models. Given that the data (and the IPF process itself) show a possible multilevel structure of individuals (level one) nested inside target small areas (level two – MSOAs in this case) it is possible to understand these as hierarchical data structures that are suitable for multilevel regression models. Technically it is more appropriate to fit multilevel models to hierarchical data in order to ensure accurate estimates of coefficients and standard errors; these are typically very similar to those obtained in single level models though not always and single level models can occasionally be misleading in these circumstances.

In a multilevel regression structure the error variance is not presented as a single R^2 value but is instead partitioned across the different levels in the model. In the empty multilevel model (i.e. the model with no explanatory variables) this represents the intraclass correlation coefficient (ICC) and shows the share of the error variance that is accounted for at the area level as compared with the individual level. This ICC value is often taken as an indicator of the importance or relevance of the multilevel structure (and hence the multilevel model) to the data. In the full multilevel model (i.e. the model with explanatory variables) the residual error variance at the different levels is reported after having accounted for the variables in the model. A comparison of those error variances between the empty and the full multilevel models indicates the extent to which the explanatory factors are able to account for the variance in the outcome at each level of the model.

Given that the emphasis in the small area estimation process is on the estimates at the area level it is the residual variance on the level two (i.e. MSOA) error term that is of central interest in terms of the credible intervals as this is the key indicator of remaining uncertainty at the area level. In order to compute the credible intervals around the central point estimates we therefore take the central IPF point estimate and, across 10,000 separate repetitions, add an additional term relating to a random draw from the distribution of this estimated residual variance on the level two error term – with mean of zero and residual variance[3] as estimated by the multilevel model. The values of the 250th and 9,750th largest cases (i.e. the 2.5th and 97.5th percentiles) of this resulting distribution of 10,000 values are taken as the lower and upper limits of the credible intervals around the central IPF point estimate.

[3] All models were estimated in Stata. The rnormal() function is used to create the credible intervals and this is in fact based on standard deviation rather than variance.

Results

As discussed above, a necessary step in preparing the IPF is to run initial binary logistic regression models for each of the target outcome variables against a series of explanatory variables. The primary aim in doing so is to identify the most powerful set of explanatory factors that shape each outcome variable so that an effective and parsimonious set of constraint variables can be identified for the IPF.

Table 4 below shows the final regression models following the process of gradual model building and testing[4]. Log odds are reported in Table 4 along with an indication of their statistical significance. These results give an indication of the direction, strength and statistical significance of the relationships between the explanatory variables and the outcome variables and help to identify those variables to can be expected to act as predictively useful constraints within the IPF. The log odds centre around a value of 0 with negative log odds showing a lower likelihood of being coded 1 on the outcome variable (e.g. feeling unsafe) compared to the reference category for that factor and with positive log odds conversely showing an increased likelihood of being coded 1 on the outcome variable (i.e. feeling unsafe) compared to the reference group. Log odds that are statistically significant (at the 5% level) are denoted by an asterisk.

The final row of Table 4 shows the pseudo-R^2 values for each of the separate binary logistic regression models and these give an indication of the explanatory power of each model as a whole. The values are expressed as percentages and so can take a value from 0 up to a possible maximum of 100. In terms of what these values mean, a value of 100% would mean that all of the variation in the outcome variable across the survey cases can be explained systematically by the set of explanatory variables (i.e. the model is a perfect predictor at the individual). At the other extreme, a value of 0% would mean that none of the variation in the outcome variable can be explained systematically (i.e. the model is a poor predictor at the individual level).

The final row shows that the pseudo-R^2 values are in general relatively low, with the exception of the model relating to internet usage where the pseudo-R^2 equals 38%. These generally low pseudo-R^2 values across these outcome variables are not entirely unexpected and, as discussed below, not necessarily critical in terms of the viability of the final small area estimates themselves. Firstly, given that SAE is limited only to those explanatory variables that can be found in both the survey data and at small area level then it is typically the case that the regression models underpinning the preparation of SAE are somewhat less powerful than general models which do not face that data constraint. Secondly, many of these particular outcome variables are attitudinal variables whose total variation can be expected to be in significant entirely random rather than as variation that relates systematically to, and that therefore can be accounted for, by a set of observed explanatory factors within models. As a result, for a large portion of the total variance of several of these

[4] Other variables tested in the process but not retained in the final models were occupational status and household composition.

outcome variables there may well be no possible explanatory variables that could be identified or collected to systematically explain much of their variation given that it is random and not systematic variation. Additional explanatory variables that might be imagined to be of potential benefit to these sorts of outcome variables (e.g. attitudinal factors, psychometric profiles) are not available in the National Survey or at small area level and, indeed, are rarely ever collected.

In terms of implications for the IPF and small area estimates themselves, one would ideally desire greater predictive power in the underlying regression models in order to seek to better explain the variance in the outcomes and to better reflect that variance across the small areas. Nevertheless, the relatively low power of these underlying regression models is not necessarily problematic for the effectiveness of the IPF itself given that these regression models are quite different in terms of their focus and scale to the IPF process itself. The regression models focus on identifying the key explanatory variables in terms of their predictive power at the individual level which will be used as constraint variables in the IPF. The focus of the IPF, however, is on creating estimates at the small area level. Whilst the two are related they are not equivalent: although the model power of the underlying regressions may be an indicator of likely ‘success’ in producing valid small area estimates (given that they indicate an ability to more effectively account for the outcome variable in a systematic fashion) this is not necessarily the case. Rather, it is the internal and external validation discussed later that offers the best insights into the acceptability of the small area estimates themselves.

Following this modelling phase the project has created small area estimates for all six outcome variables using the set of constraint variables identified in Table 4. For each of these outcome variables the IPF produces an estimate of the average number of adults aged 16 or above who, respectively, use the internet, feel unsafe, experience financial difficulties and are satisfied with the local area, with GP care or with the performance of the Welsh Government. These estimated totals are then expressed as a percentage of the adult population in each MSOA so as to give estimated rates of prevalence of the six outcomes across all 410 MSOAs in Wales.

For each MSOA, 95% credible intervals are provided around the central IPF point estimates using the residual level two error variance estimated within equivalent multilevel logit models (i.e. with the same set of explanatory variables as the single level regression models) as described in Step 7 of the methods section above. The multilevel models underpinning the creation of these credible intervals via the estimation of the residual variance on the level two error terms are presented in Table 7 in the Appendix. These credible intervals provide a lower bound and upper bound around each estimate that indicate the likely range within which one can be statistically confident that the ‘true’, but unknown, percentage for the MSOA population can be expected to fall. In this project only individual level constraint variables were employed and future would could explore the possibility of incorporating additional area level constraints in order both to seek to improve central point estimates as well as to seek to reduce the width of the credible intervals given that these rely on the residual variance in the error at the area level.

Table 4: Model output for each of the six outcome variables to be estimated down to MSOA level

Explanatory Variables		Satisfaction with Area	Feeling unsafe	Satisfaction with Welsh Gov	Satisfaction with GP care	Finance Problems	Internet Use
Tenure (ref=owned)	Social Rent	-0.22*	0.20*	0.16*	-0.14	0.77*	-0.58*
	Private Rent	-0.03	-0.04	0.23*	0.33*	0.62*	-0.21*
Employment Status (ref= unemployed)	Working	0.47*	-0.41*	0.24	0.09	-0.92*	-0.10
	Retired	1.53*	1.62*	0.32*	1.22*	-1.76*	-3.89*
	Inactive	0.38*	-0.08	0.28*	0.16	-0.50*	-0.65*
	Full-time Student	0.37*	-0.30	0.52*	0.99*	-0.69*	0.63
(ref= no health problems)	Health Problems	-0.41*	0.79*	-0.10	-0.66*	0.64*	-0.15*
(ref = no car)	Access to Car	0.09	-0.25*	-0.21*	0.07	-0.16*	1.19*
Qualifications (ref= no/low qualifications)	Medium Qualifications	-0.06	-0.16*	-0.11*	-0.13	-0.20*	1.01*
	High Qualifications	-0.17*	-0.41*	-0.33*	-0.29*	-0.53*	1.86*
(ref = no child in household)	Dependent Child	-0.04	0.09	0.13*	-0.03	0.29*	0.58*
House Type (ref=detached)	Semi-Detached	-0.57*	0.33*	0.01	0.02	0.47*	-0.13*
	Terraced	-0.98*	0.50*	-0.06	-0.17	0.57*	-0.25*
	Flat	-0.90*	0.67*	0.07	-0.12	0.21	-0.06
	Other Dwelling	-0.16	-1.08	-0.57	-0.40	1.09*	0.08
Age-Sex Group (ref= Male 16-29)	Male 30-49	0.24*	0.16	-0.50*	0.20	0.24*	-0.92*
	Male 50-64	0.59*	0.13	-0.13	0.62*	0.03	-1.96*
	Male 65+	-0.08	-1.20*	0.07	0.24	0.08	0.31*
	Female 16-29	0.08	1.50*	-0.13	-0.18	0.03	0.05
	Female 30-49	0.29*	1.34*	-0.44*	0.31	0.32*	-0.67*
	Female 50-64	0.57*	1.36*	-0.25*	0.84*	-0.00	-1.86*
	Constant	0.73*	-2.35*	-1.30*	1.98*	-1.24*	1.88*
	Observations	14327	14108	13250	11428	13985	14359
	Pseudo-R^2	7%	11%	1%	4%	12%	38%

Note: * denotes where p<=0.05

Figures 4 to 9 give a visual sense of the variation in the small area estimates across Welsh MSOAs in the form of caterpillar plots. To simplify their presentation, these figures select a systematic 10% sample of MSOAs across the full range of small area estimates on that outcome variable. For each MSOA the charts show the small area estimates in the form of a cross with the 95% credible intervals drawn around that value. Some of the outcome variables show greater variability than others across the MSOAs: high satisfaction with the performance of the Welsh Government and satisfaction with GP care are particularly flat across the MSOAs whilst the estimates for satisfaction with the local area show much greater variability from around 60% of adults in some MSOAs up to just over 80% of adults in others.

The credible intervals provide a sense of the uncertainty and plausible error around the point estimates. These credible intervals are relatively wide and, as is often the case in small area estimation work, are often overlapping. In terms of their meaning, the credible intervals show the estimated range within which we can be statistically confident that the ‘true’, but unknown, underlying small area population values for these outcomes falls. Where the credible intervals of different small areas overlap, therefore, this suggests that we cannot be statistically confident in stating that the areas necessarily have different underlying population values for the outcome, even if the central small area point estimates do suggest a difference. Given the methodology used here to create the credible intervals future work could explore the potential to incorporate additional area level factors into the IPF to seek to try and reduce the width of the intervals as well as to seek to better explain the variance in the outcomes and widen the spread of the central point estimates themselves.

As noted above, given that the width of the intervals depends upon the level of residual variance in the level two error term within the multilevel models future work could explore the possibility of incorporating additional area level factors into the process in order to seek to reduce the width of the credible intervals. This may also seek to better account for the variation in the outcome variables and in so doing to widen the spread of the central point estimates across the small areas. There is however a limit to what can be expected to be achievable in this regard however, and relatively wide and frequently overlapping intervals are a common feature of any small area estimation work.

Figure 4: MSOA estimates of percentage of adults satisfied with their local area (systematic 10% sample of MSOAs shown)

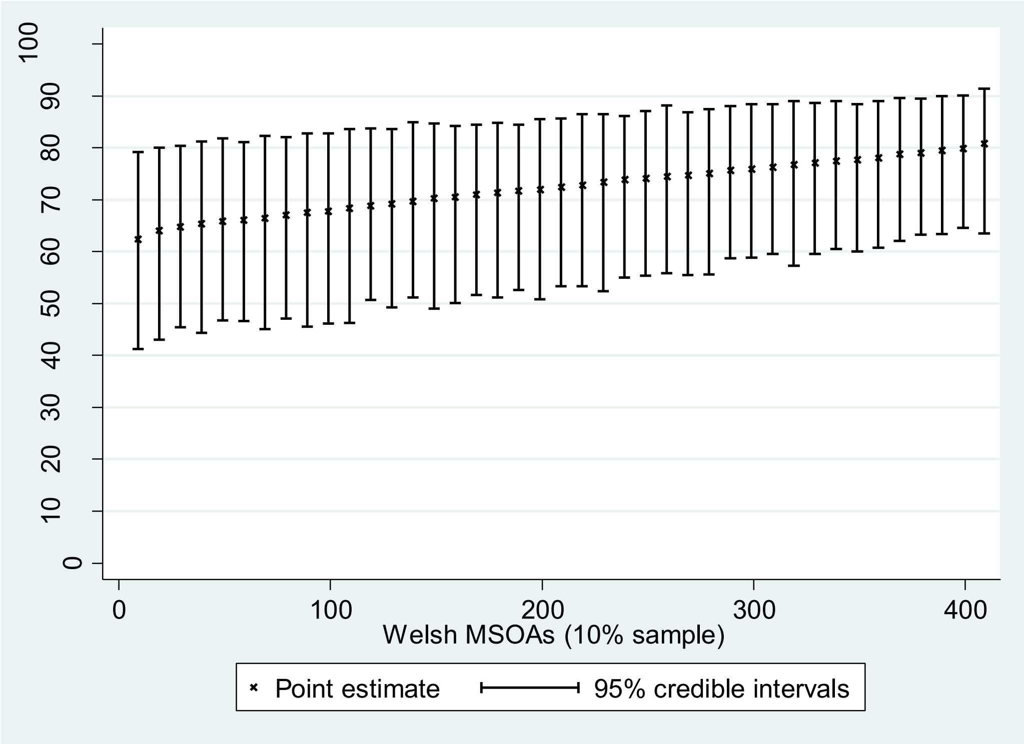


Figure 5: MSOA estimates of percentage of adults experiencing financial difficulties (systematic 10% sample of MSOAs shown)

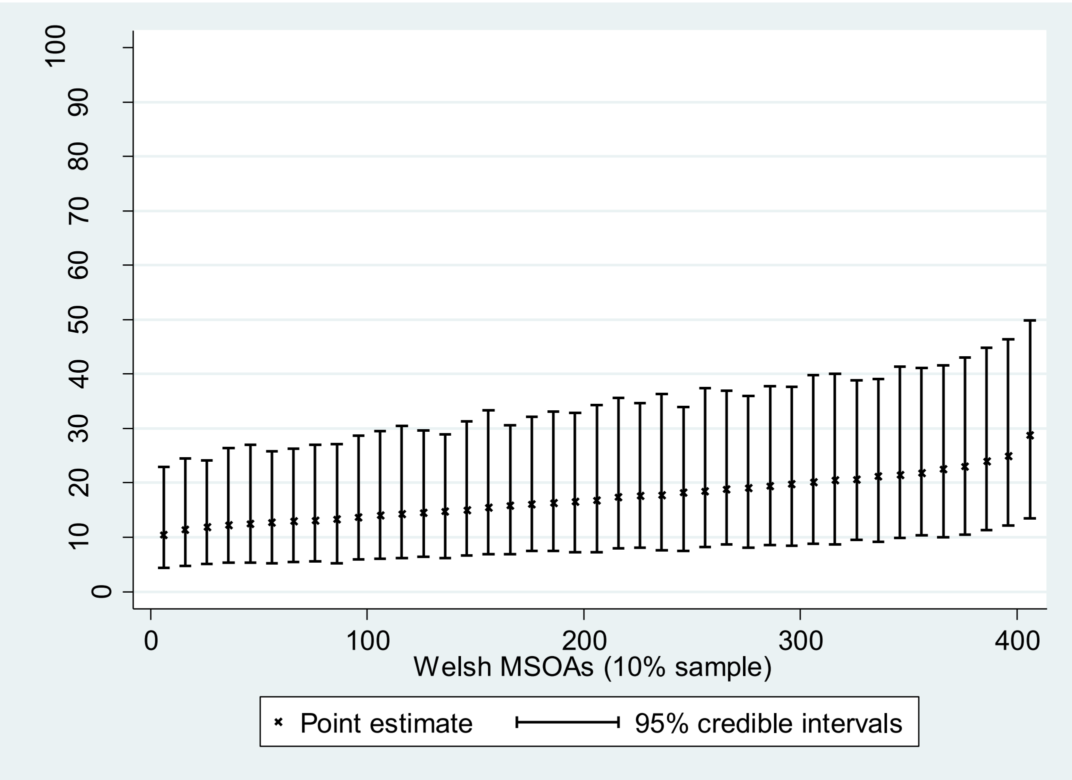


Figure 6: MSOA estimates of percentage of adults who feel unsafe (systematic 10% sample of MSOAs shown)

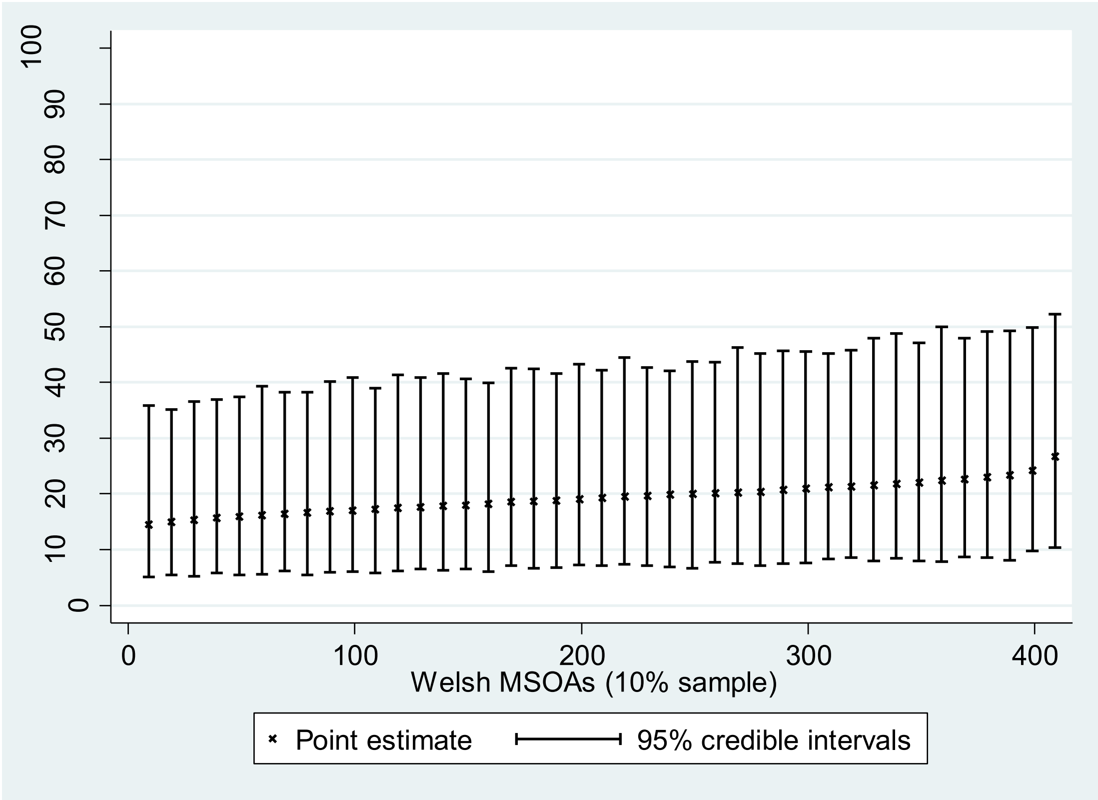


Figure 7: MSOA estimates of percentage of adults satisfied with GP care (systematic 10% sample of MSOAs shown)

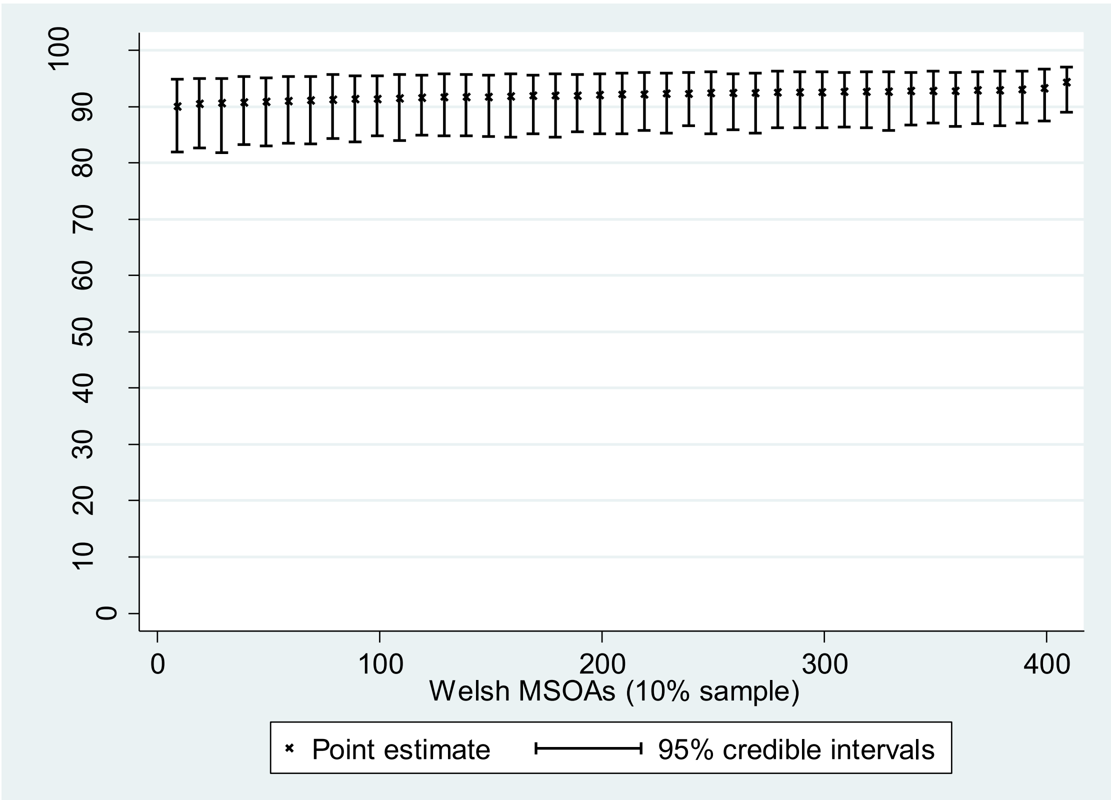


Figure 8: MSOA estimates of percentage of adults using the internet (systematic 10% sample of MSOAs shown)

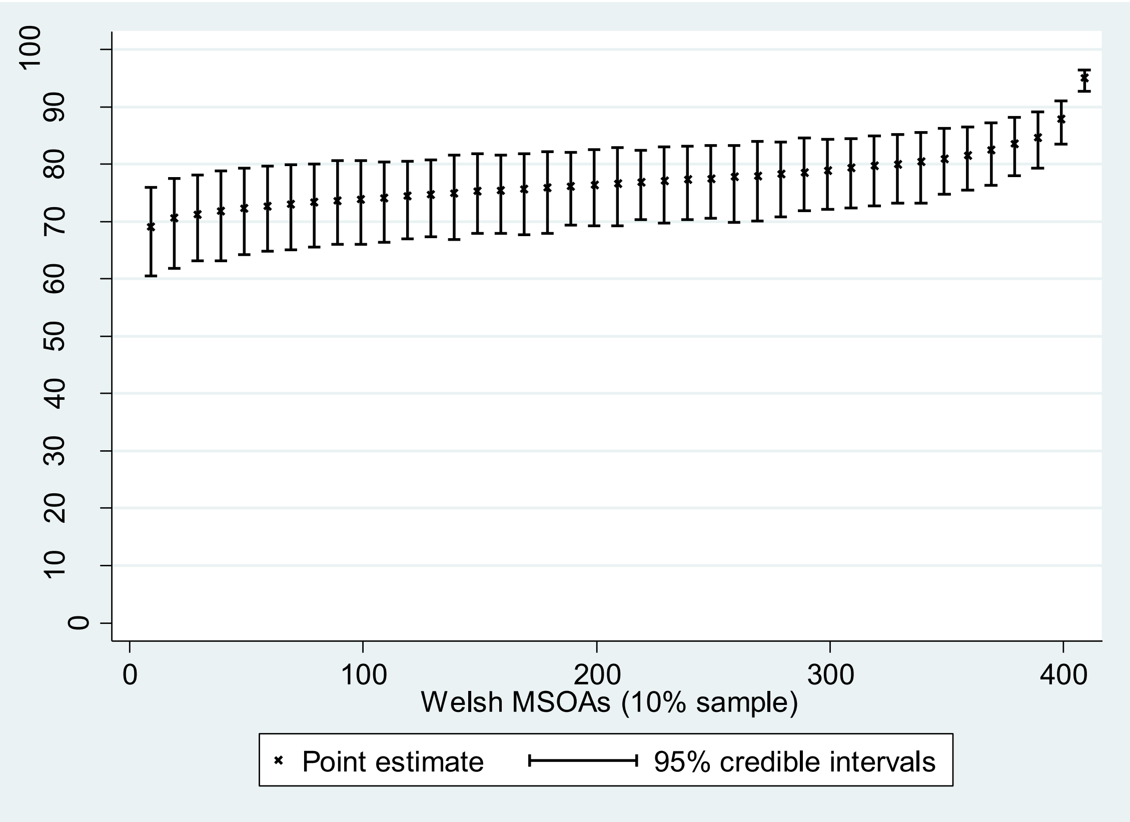
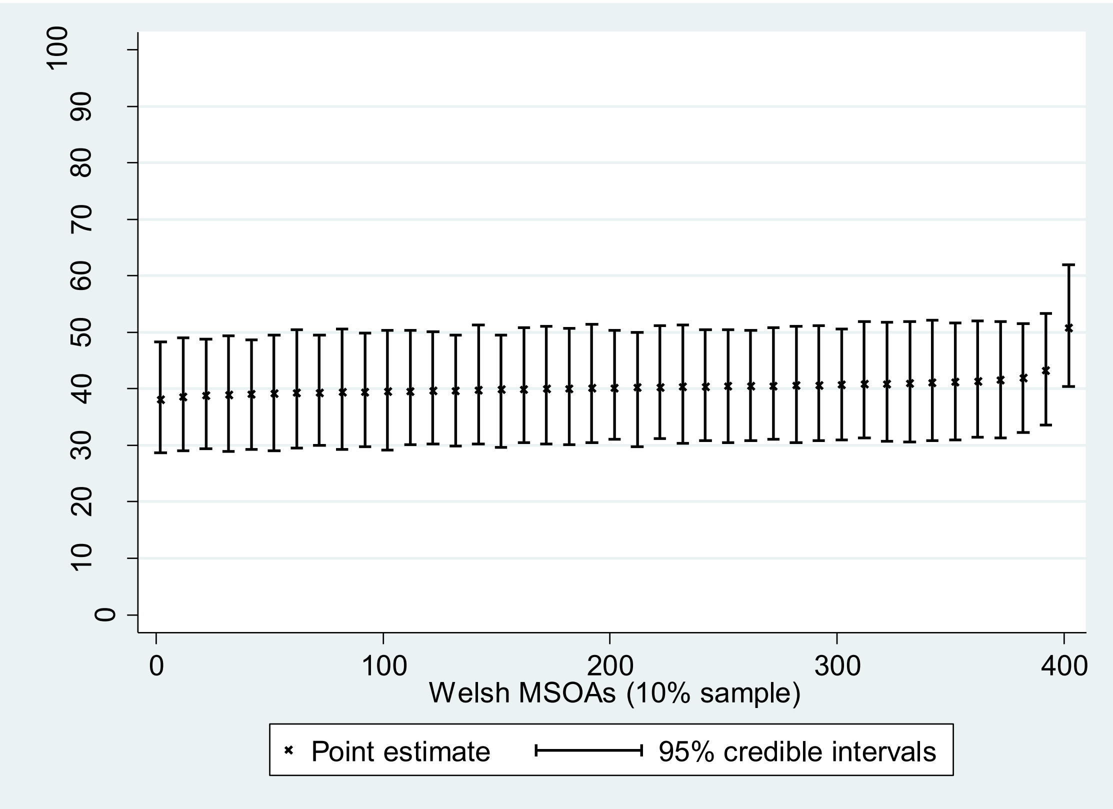


Figure 9: MSOA estimates of percentage of adults having high satisfaction with performance of Welsh Government (systematic 10% sample of MSOAs shown)



Figures 10 to 12 map the six sets of MSOA small area estimates across three varied local authority areas – Cardiff, Ceredigion and Wrexham. Each map is drawn across five national quintiles each containing 20% of Welsh MSOAs. Different patterns emerge both inside and between each of these local authority areas across the six sets of MSOA estimates.

Figure 10: MSOA small area estimates for Cardiff

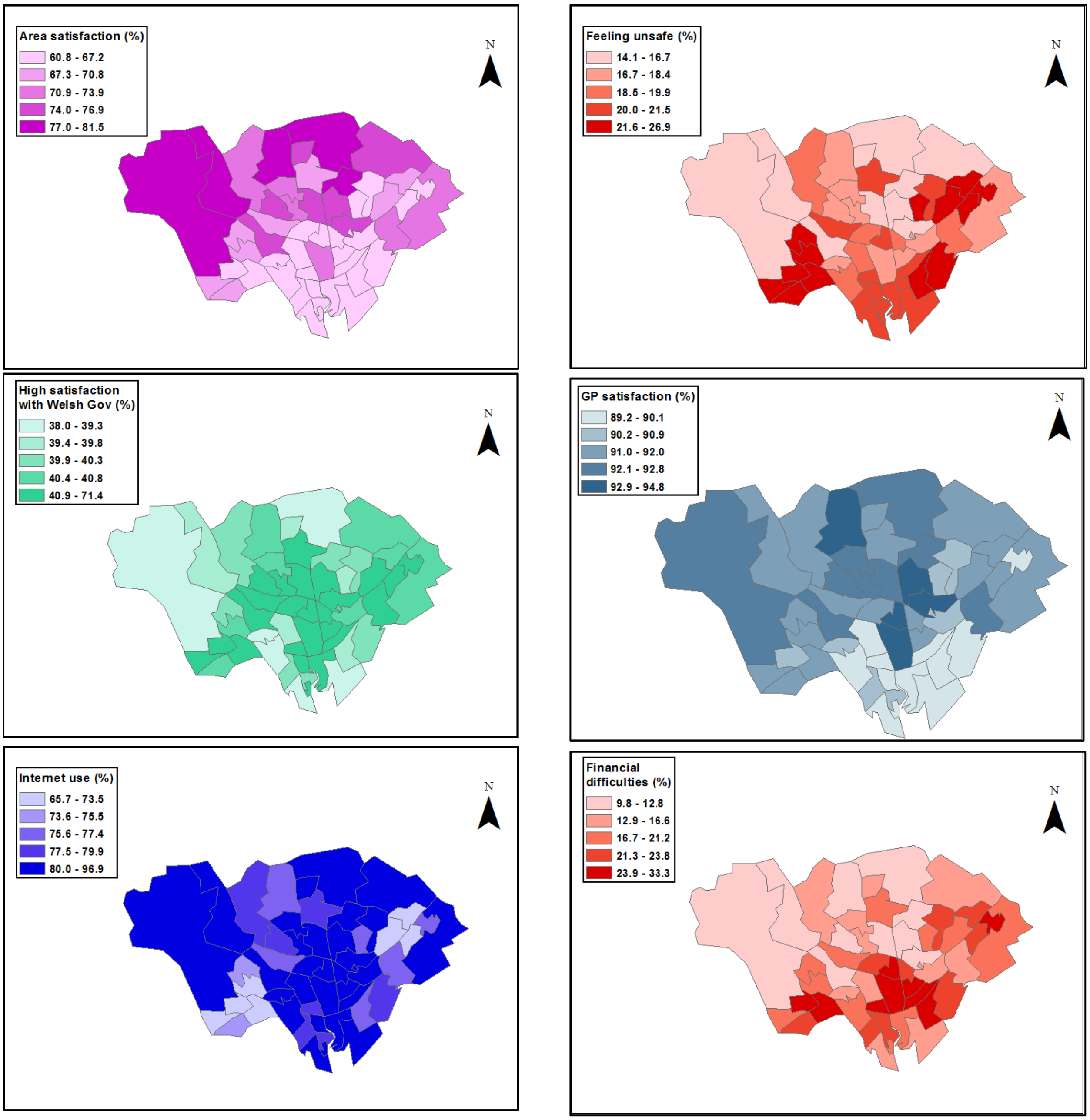


Figure 11: MSOA small area estimates for Ceredigion

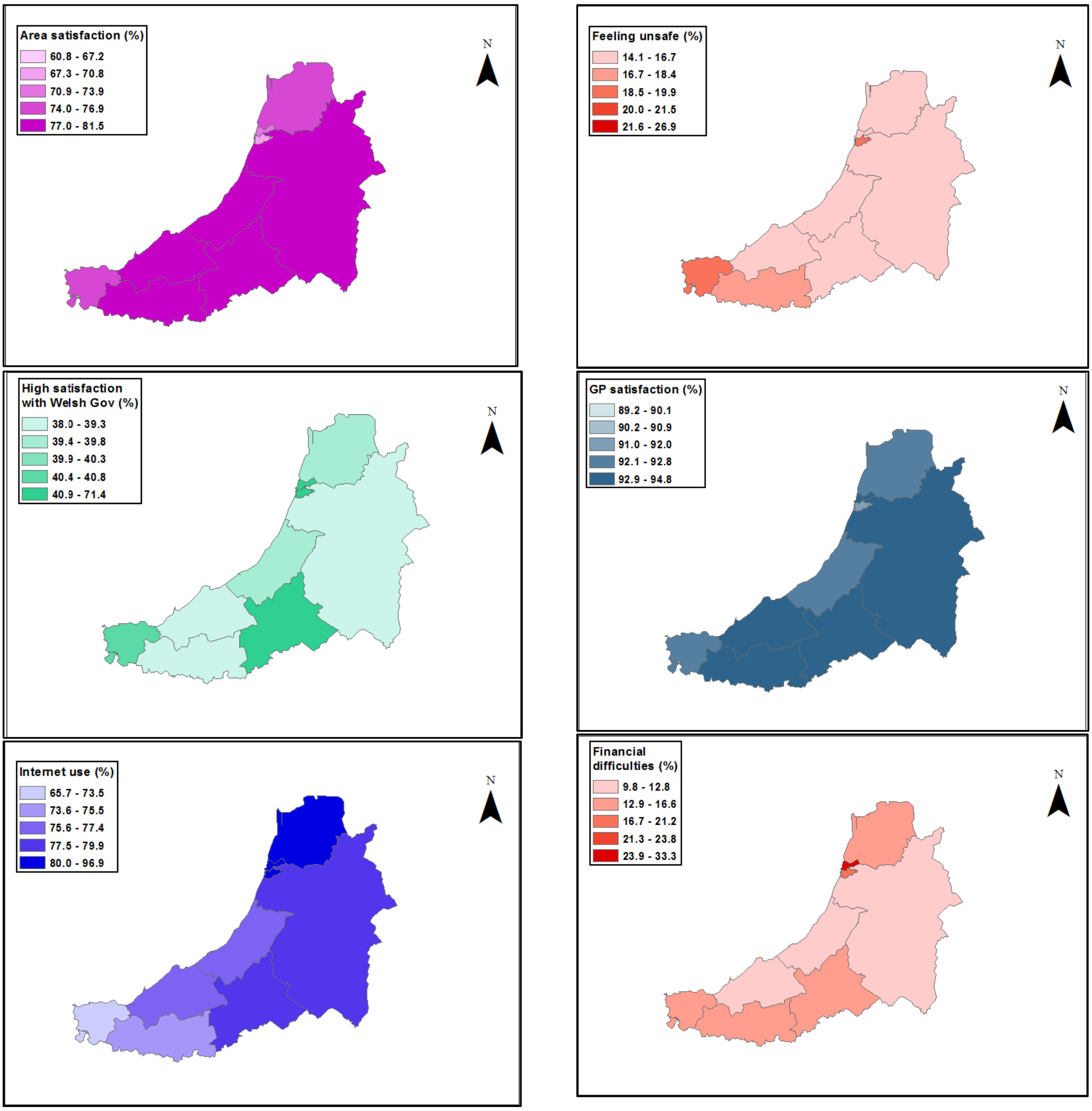
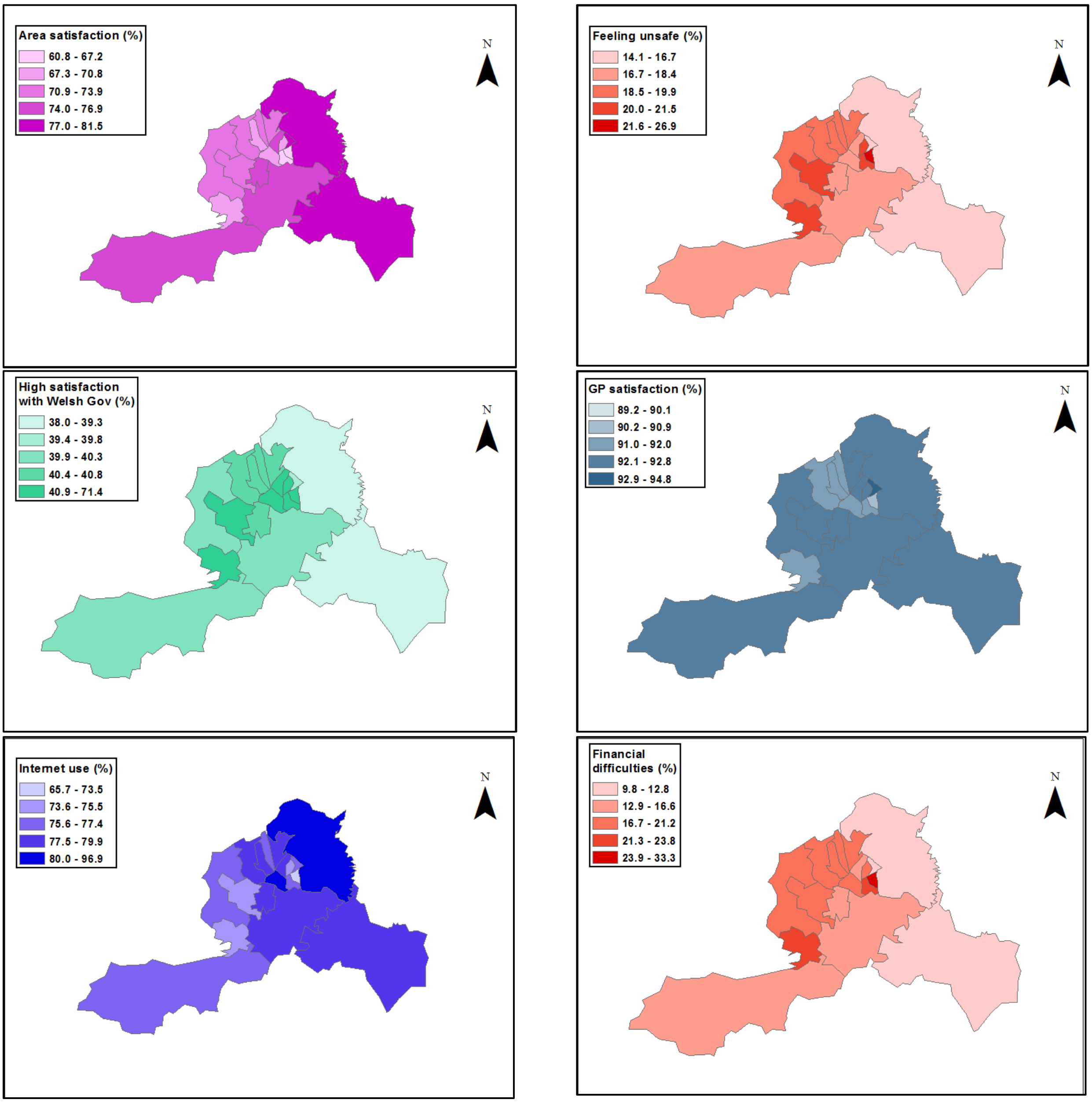


Figure 12: MSOA small area estimates for Wrexham



Validation

An important aspect of any small area estimation process is to validate the estimates produced in order to check their likely validity in terms of being an accurate reflection of the 'true', but unknown, small area population value. Two separate types of validation can be conducted. Firstly, internal validation refers to the process of assessing the model itself in terms of its power and fit. Secondly, external validation refers to the process of assessing the extent to which the estimates themselves can be corroborated by other data, whether at the same scale (e.g. MSOA scale here) or whether at higher spatial scales.

For IPF approaches to SAE, internal validation focuses on the extent to which the final weights give weighted sums of the constraint variables in the National Survey that are well fitted to the small area totals (in this case the MSOA totals as derived from the Census 2011). Table 5 below shows two commonly used fit statistics for the internal validation (Anderson, 1997; Voas and Williamson, 2001; Smith et al., 2009).

For each constraint, the first column of data (% Fit) shows the percentage of Wales' 410 MSOAs that have weighted survey estimates within 20% of the actual total as derived from the Census 2011. Table 5 shows excellent fit across the constraints. Secondly, the mean percentage of standardised absolute error (% Error) is calculated for each constraint as an alternative measure of internal validation. This figure is calculated by taking the difference between the IPF's weighted constraint total for the MSOA and the actual MSOA total from the Census and then expressing that difference as a percentage of the actual Census total for that constraint. With the exception of the economic status constraints all variables show excellent levels of standardised absolute error. Although not unexpected given the nature of the IPF process, these internal validation statistics are certainly acceptable when taken together.

Table 5: Internal validation statistics for the IPF

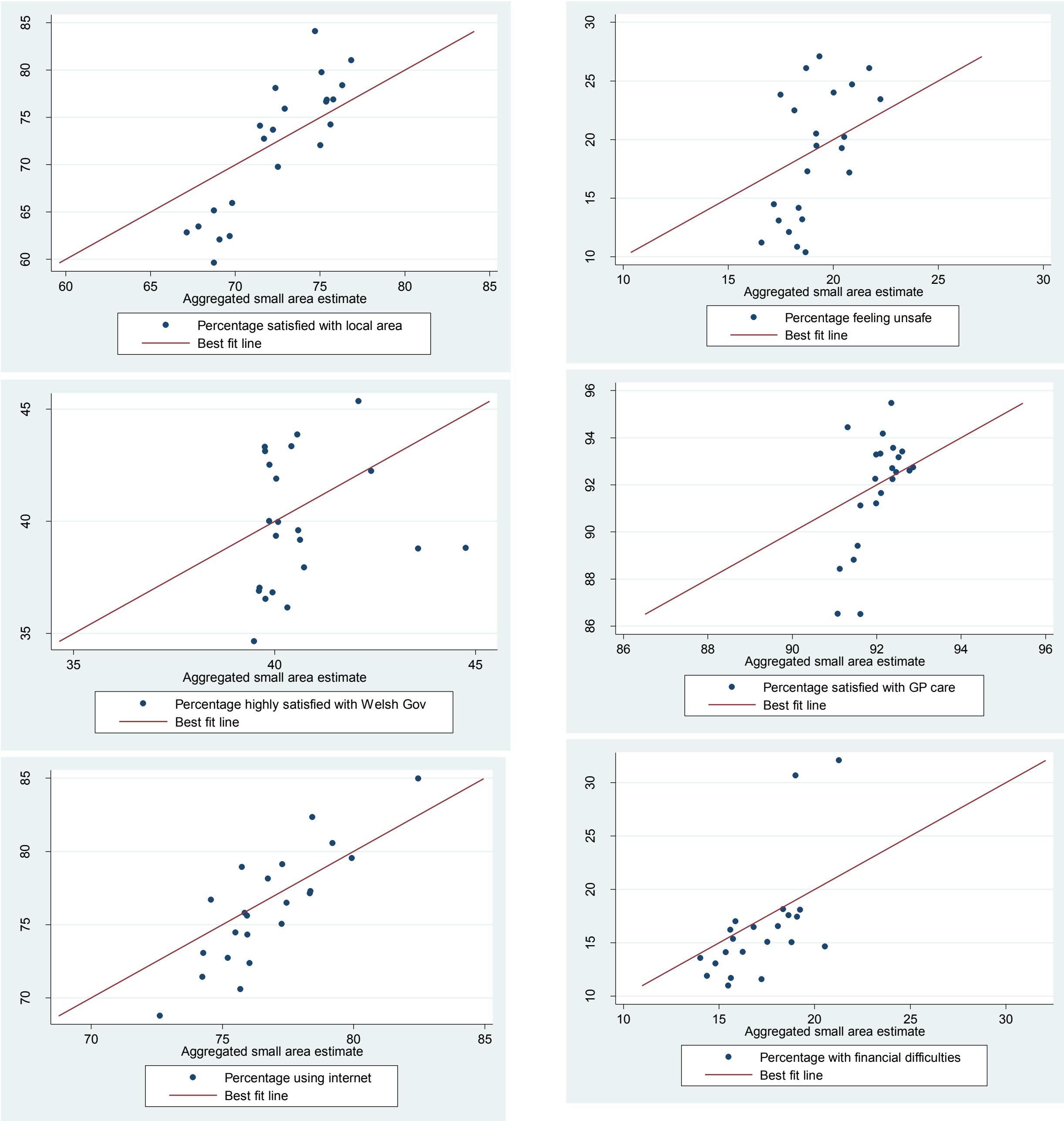
	Constraint	% Fit	% Error		Constraint	% Fit	% Error
Age-Sex Groups	Female 16-29	100	0.0	Health Status	No Health Problem	100	0.0
	Female 30-49	100	0.0		Health Problem	100	0.0
	Female 50-64	100	0.0	Vehicle Access	No Car	100	0.0
	Female 65+	100	0.0		Car	100	0.0
	Male 16-29	100	0.0	Education	High Quals	100	0.0
	Male 30-49	100	0.0		Medium Quals	100	0.0
	Male 50-64	100	0.0		Low/No Quals	100	0.0
	Male 65+	100	0.0	Children	Dependent Child	100	0.0
Tenure	Social Renter	100	0.0		No dependent Child	100	0.0
	Private Renter	100	0.0	House Type	Detached	100	0.0
	Owner Occupier	100	0.0		Semi-detached	100	0.0
Economic Status	Working	100	3.3		Terraced	100	0.0
	Unemployed	100	3.3		Flat	100	0.0
	Inactive	100	3.3		Other	100	0.0
	Full-Time Student	100	3.3				
	Retired	100	9.9				

External validation assesses the extent to which the estimates themselves can be corroborated by other data and this in many ways gives the best indicator of the robustness of the final small area estimates. External validation is often problematic given that typically the reason for conducting small area estimation in the first instance is that no such data exist. However, it is sometimes possible to externally validate estimates at the same spatial scale if suitable proxy variables can be found: Anderson (2007) for example externally validates LSOA income estimates across England by comparing them with LSOA Index of Multiple Derivation ranks. Often, however, proxies at the same spatial scale cannot be found and external validation is instead carried out by aggregating the small area estimates up to a higher geographical scale where they can then be compared to direct estimates taken from a survey to that higher scale.

In the absence of suitable comparable data at MSOA level across Wales for these six outcome variables, the approach taken here is to externally validate the MSOA estimates by aggregating them to local authority level and by comparing them against the mean survey estimates for local authorities taken directly from survey weighted analyses of the National Survey. By comparing the two sets of local authority estimates (one direct and one aggregated SAE estimates) one can gain a sense of the external plausibility of the small area estimates.

Figure 13 below shows scatterplots comparing the percentage of adults affected on each outcome variable according to the direct local authority estimates from the National Survey (vertical axis) and according to the aggregated total of the MSOA small area estimates (horizontal axis). The red line presents a line of equality between the two sets of estimates where points along this line would reflect identical results on the two sets of estimates. Some of these sets of local authority estimates appear visually to match more closely than others along this line of equality: the percentage of adults using the internet is in general well matched for example around this slope whilst other outcomes show more variability around the slope. Three outcomes – feeling unsafe, high satisfaction with Welsh Government and satisfaction with GP care – are flatter in the small area estimates than in the direct survey estimates, suggesting that the IPF is not always able given these constraints to replicate the level of variance seen at local authority level. Further exploration with additional (especially area level) variables may help to better reflect this variation in future work. Local authority outliers are also apparent on some charts and whilst this is usually a gradual phenomenon it is for example focused on just two local authorities on the chart relating to the percentage of adults experiencing financial difficulties. Further work to assess whether specific characteristics of those local authorities might be used to improve the estimates in those areas could be explored.

Figure 13: External validation of the aggregated small area estimates against local authority direct estimates



More formal external validation can be provided in the form of the mean absolute error of these aggregated estimates as compared to the direct estimates from the National Survey. In order to calculate mean absolute error, for each local authority the difference in the percentage of adults estimated to be affected by each particular outcome variable is calculated between the local authority survey estimates and the small area estimates aggregated to local authority level. This is an absolute percentage point difference in which all differences are expressed positively. The mean of this absolute percentage point difference is then calculated across all 22 Welsh local authorities. These results are provided in Table 6 below.

Table 6: Mean absolute error between aggregated small area estimates and direct survey estimates across Welsh local authorities

Outcome Variable	% Mean Absolute Error
Internet Use	2.0
Financial Difficulties	3.0
High Satisfaction with Welsh Government	2.7
Satisfaction with GP Care	1.5
Feeling Unsafe	4.1
Area Satisfaction	3.8

On average, the mean of these differences lies roughly between two and four percentage points across the local authorities. There are no benchmarks as to what should be considered acceptable fit and this is largely a subjective decision based on the degree of accuracy that the analyst or policy maker demands of the estimates. In general, the level of these mean absolute error statistics appears reasonable at this scale.

Two points should be noted. Firstly, for each outcome variable these figures reflect mean differences in the two sets of estimates across all 22 local authorities. Typically, however, these mean values are somewhat skewed by a small number of local authorities that show notably larger differences than the remaining majority of local authorities. This is positive in the sense that the majority of local authorities show notably smaller differences than the means in Table 6 suggest. At the same time, those local authorities that might be considered to be outliers due to their larger differences are perhaps candidates for further attention in terms of consideration of the possibility for more locally specific estimation. The results for each local authority are presented in more detail in Table 8 in the Appendix.

Secondly, although one may be comfortable with the extent of error at the local authority level it is not possible to be certain about the distribution of this area at the small area scale within the local authority, even if the aggregated MSOA estimates appear reasonable at the local authority level. Certainly one would expect greater variability at that smaller scale. This is almost always true for any small area estimation project given the frequent lack of suitable data against which to perform external validation at the small area level. In terms of potential future validation work, in order to test for the validity of the small area estimates at the small area level itself then the Welsh Government may wish to consider conducting surveys in selected case study geographies in order to generate sufficiently large local samples to produce robust direct survey estimates in these selected test case study areas. This would present the gold standard test in terms of externally validating the small area estimation. Alternatively, one might test the IPF by creating small area estimates of a variable that is available directly at small area level from Census or administrative data (e.g. long-term limiting illness) and then comparing the estimates against those known values. Given that this is one step removed from the validation of the actual target variable of interest it does reflect a proxy validation exercise but this would give a general indication of the likely success of the SAE.

Pushing the boundaries of SAE: Future development and potential next steps

The focus of this project has been on the viability of producing small area estimates for six diverse outcome variables at the Middle Layer Super Output Area (MSOA) level across Wales using the National Survey dataset. However, a final aspect of the project is to provide some insight into the viability of three inter-related potential future developments:

- the production of small area estimates at the even smaller scale of the Lower Layer Super Output Area (LSOA) geography;
- the potential to produce small area estimates from smaller national surveys;
- the potential to produce local authority estimates from smaller national surveys.

The remainder of this report explores each of these three potential future developments in detail.

Producing small area estimates at the LSOA level: a viable progression?

In terms of the scale of any final small area estimates, MSOA estimates do offer significant spatial detail and represent a considerable improvement in geographical understanding compared to local authority information. If it were possible to estimate down to LSOA level, however, then this would offer even greater spatial detail and sensitivity still. Across England and Wales LSOAs have an average population size of around 1,600 residents compared to an average of around 7,800 residents for MSOAs. Hence, whilst both are small scale geographies there are on average around five LSOAs nested inside each MSOA. In terms of the survey data, greater knowledge about the potential viability of using survey datasets with fewer cases than the National Survey may open up opportunities to make additional use of existing smaller, possibly more thematically focussed datasets or to be able to resource the special commissioning of such new datasets.

The testing of whether it is viable to produce estimates at the LSOA level focuses on the case study area of Cardiff and the single outcome variable of internet use. Figure 14 shows these small area estimates of the percentage of adults in each LSOA in Cardiff that use the internet. For ease of comparison, the MSOA estimates of internet usage across Cardiff shown in the bottom left pane of Figure 10 above are reproduced in Figure 15. With the exception of some slightly wider values at the top and bottom ends of the LSOA range, both maps use identical ranges to allow for direct comparison across the two maps. In broad terms the two maps inevitably show similar spatial patterns but the finer levels of detail and subtlety in the LSOA map *within the MSOAs* is also evident.

Although these equivalent ranges allow for direct comparison between the LSOA and MSOA small area estimates, policy makers in a particular local authority may instead potentially be more interested in exploring LSOA variation in internet usage *within* their local authority in greater detail. In order to highlight more of the variation across Cardiff's LSOAs, therefore, these internet usage estimate are redrawn in Figure 16 based on Cardiff's internet usage quintiles – the lightest blue shading relates to the 20% of Cardiff LSOAs with the lowest internet usage estimates whilst the darkest blue shading relates to the 20% of Cardiff LSOAs with the highest internet usage estimates. By exploring LSOA quintiles within Cardiff, Figure 16 is better able to draw attention to the relative differences inside this one local authority area.

Figure 14: Percentage of adults estimated to be internet users across LSOAs in Cardiff

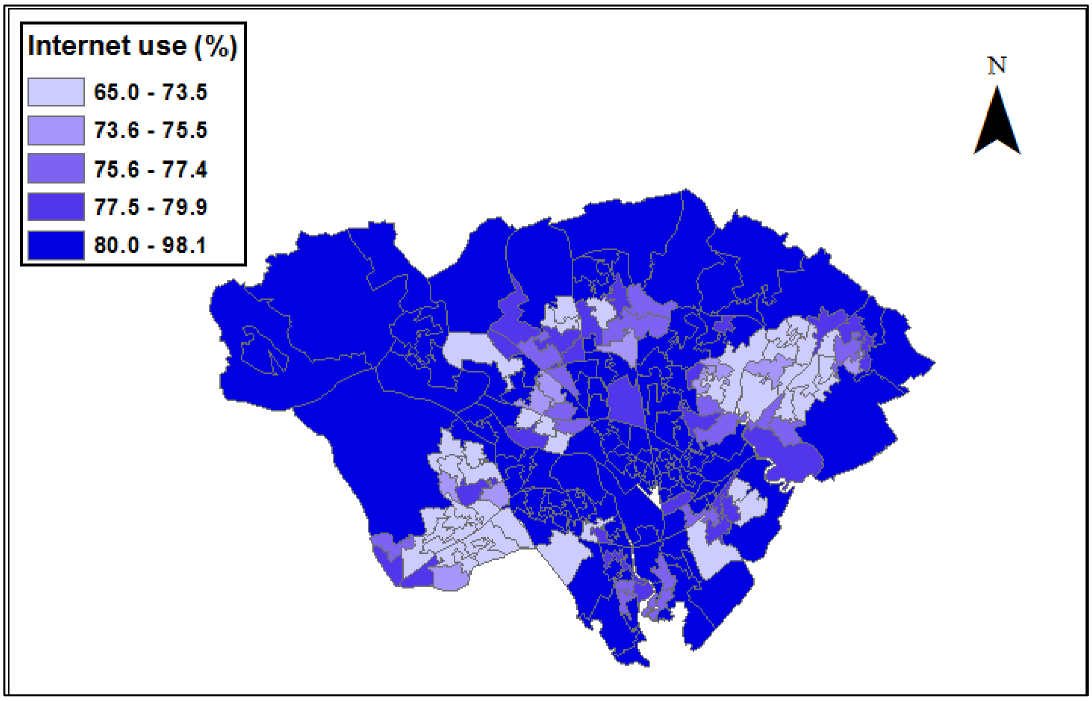


Figure 15: Percentage of adults estimated to be internet users across MSOAs in Cardiff

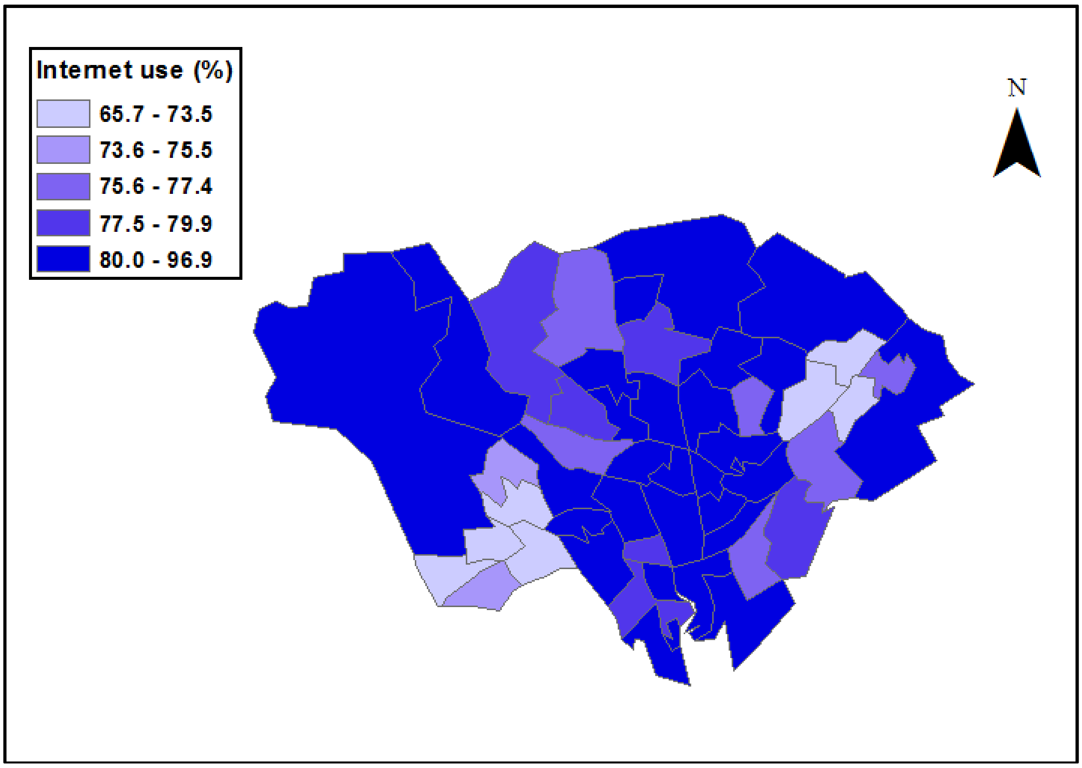
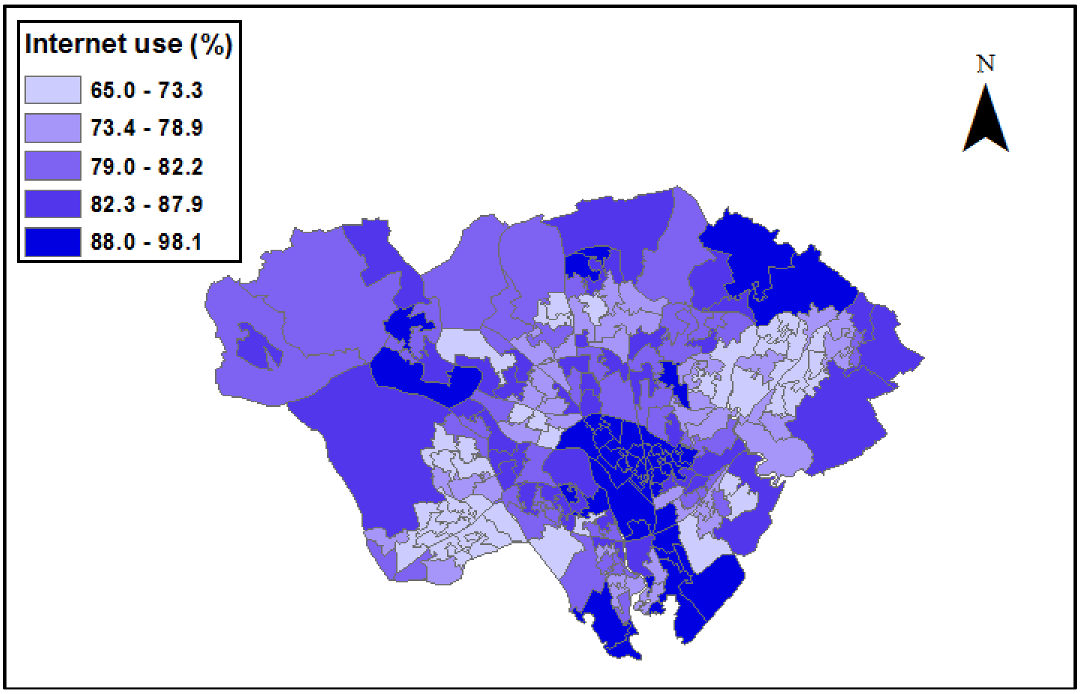


Figure 16: Percentage of adults estimated to be internet users across LSOAs in Cardiff (within Cardiff quintiles)



It is therefore viable to produce central point estimates at the smaller LSOA scale and this does, at least in this example, offer additional spatial detail to results. However, in terms of the overall potential of producing small area estimates at the LSOA scale two key additional considerations are, firstly, the viability of producing credible intervals at this smaller LSOA scale and, secondly, the width of any such credible intervals.

Figures 17 and 18 below show resulting credible intervals around Cardiff's LSOA estimates for internet use and experiencing finance difficulties based on equivalent multilevel models to those used for these outcomes at MSOA level. For ease of presentation the focus is again on a systematic 10% sample of Cardiff's LSOAs across the full range of estimates. Table 9 in the Appendix shows the residual variance in the level two error terms estimated within LSOA multilevel models that were used to derive these intervals. These LSOA models use the same set of explanatory factors for each outcome as in the equivalent MSOA levels reported earlier. As expected, the intervals are wide at this scale but it is considered viable to produce estimates and credible intervals at LSOA level given the current sample size and smapling approach taken by the National Survey. As noted earlier, this project's implementation of the IPF approach relied upon the use of individual level constraint variables only. Further work might additionally seek to incorporate additional area level explanatory factors to seek to reduce the width of these intervals.

Figure 17: Credible intervals around Cardiff LSOA estimates of the percentage of adults using the internet

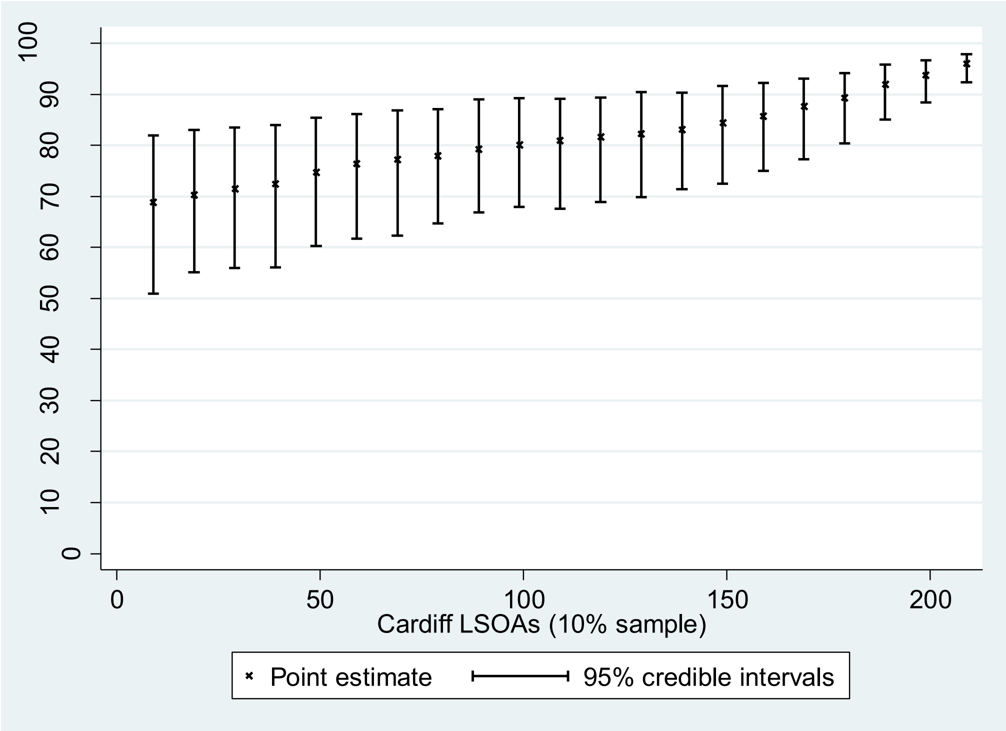
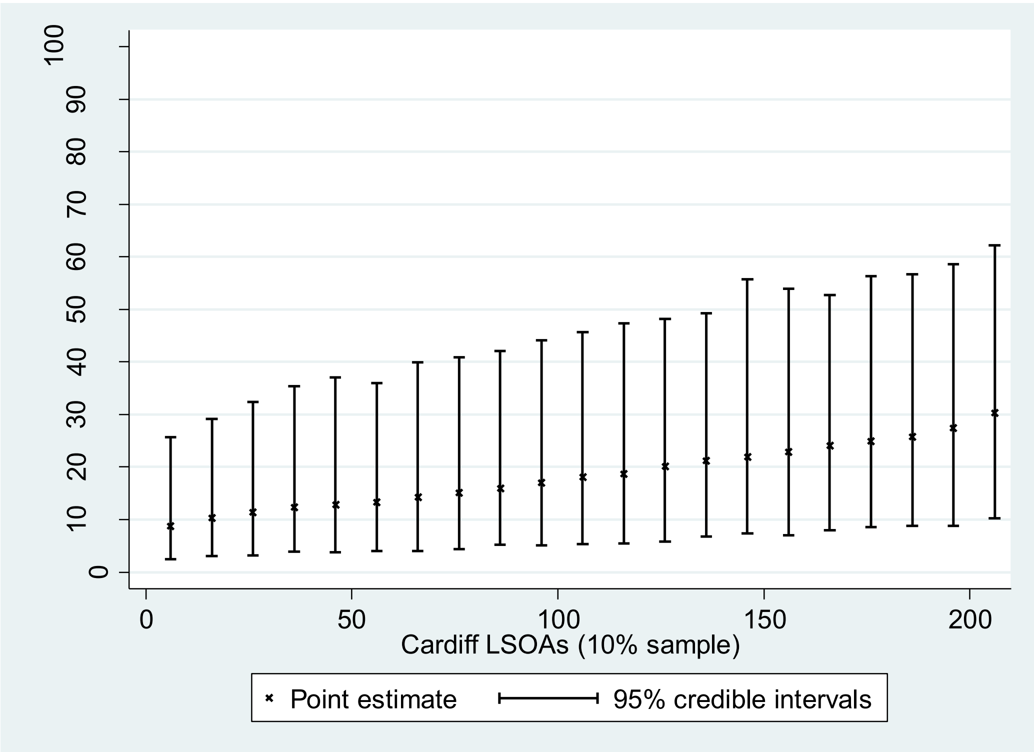


Figure 18: Credible intervals around Cardiff LSOA estimates of the percentage of adults experiencing financial difficulties



Basing small area estimates on smaller national surveys: can it be done and what are the impacts?

A second issue of interest is the extent to which any future small area estimation work could viably be based on smaller national surveys and, if so, what the impact of doing so would be both on the central point estimates and the intervals around them. To begin to test this issue, Figures 19 and 20 below show distributions of LSOA estimates based on 100 repeated sub-samples of differing percentage sizes of the full National Survey (14,362 cases), each drawn randomly within the local authority strata:

- Sample 1: 75% stratified random sample of National Survey (10,292 cases);
- Sample 2: 50% stratified random sample of National Survey (6,869 cases);

Figures 19 and 20 below focus on two LSOA test cases and focus on the estimation of internet use and feeling unsafe in two LSOA areas using 100 sub-samples of 75% and 50% of the full National Survey. These test cases naturally cannot be generalised to all small area in terms of their specifics but their general messages and principles would be expected to hold elsewhere. Each curve in these figures plots the resulting IPF small area estimates from the 100 repeated sub-samples of that size taken from the full National Survey. In terms of the meaning of these curves, the idea is that any such future surveys might plausibly be of these smaller sizes. Naturally, only *one* survey of this size would be collected and any resulting small area estimates would inevitably be affected by both the specific size and nature (e.g. diversity) of that one sample. The issue, however, is that one would not know how that one sample collected differs to all other possible samples that *could* have been collected. In effect, therefore, each curve is similar to a sampling distribution showing the likely range across which the central point estimates would be expected to fall in these test examples if *one* survey of these respective sample sizes were drawn. A dashed vertical line is placed onto each chart and this displays the central point estimate estimated from the full National Survey file.

The modal (i.e. most common) points in the curves on each chart are relatively similar and are close to the central estimates calculated from the full National Survey. As might be expected, there is a gradual tendency for the likely range of the central point estimates to widen as the sample size reduces, though the range of likely variation remains relatively small in these test cases even with notably smaller base surveys.

Figure 19: The impact of sample size on central point estimates of the percentage of adults using the internet in LSOA14

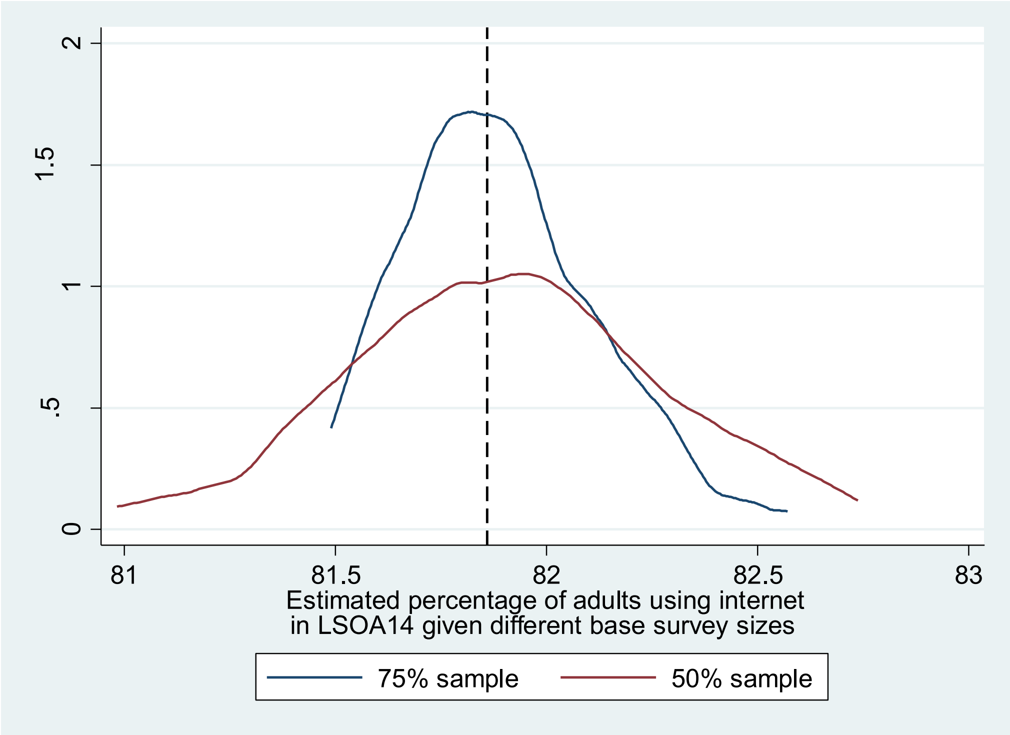
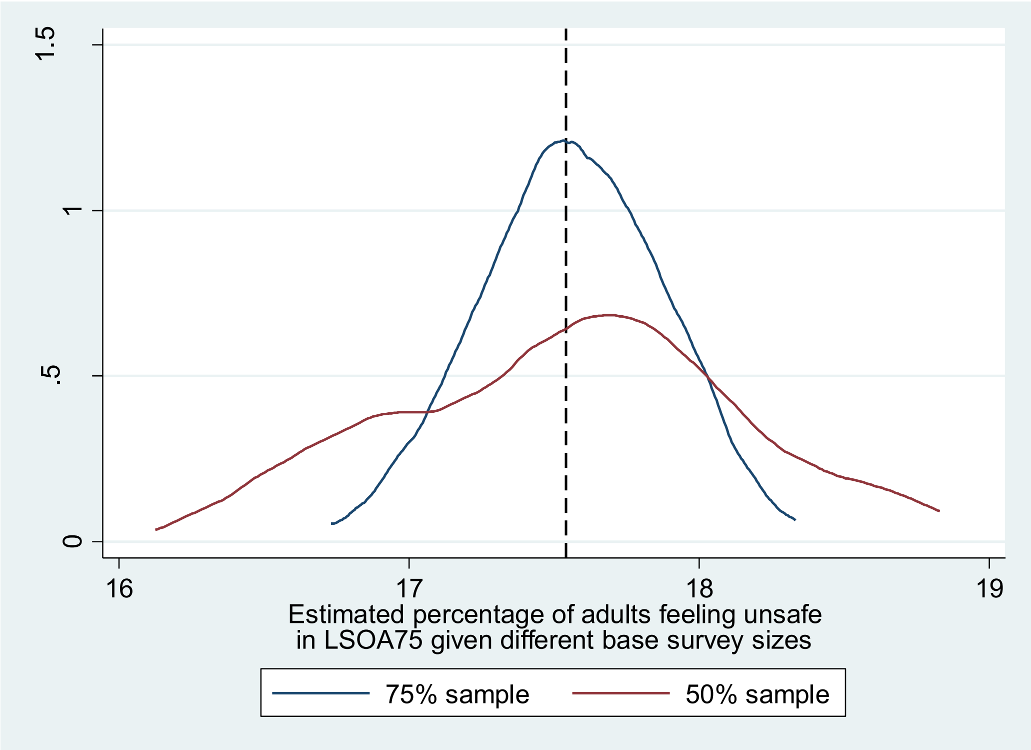


Figure 20: The impact of sample size on central point estimates of the percentage of adults feeling unsafe in LSOA75



The previous section described how the creation of the credible intervals around the central IPF point estimates at LSOA level is viable given the current sample size and sampling strategy of the National Survey. If the survey remained (stratified) random but reduced in size, however, would it remain viable to construct the credible intervals?

It is worth reflecting on the linkages between the National Survey and the approach used to estimate these credible intervals in order to clarify the demands that LSOA level estimation places upon the data and the resultant requirements in terms of base survey sample size. It will be remembered that the method used to create the credible intervals relies upon the estimation in the survey data of the residual variance on the area level error term within a multilevel regression model in which survey individuals (level one) are nested inside the target small area (i.e. LSOAs in this case). One constraint here is that reliable estimation of variance within such multilevel models relies upon having an adequate number of level two units (LSOA or MSOA areas in this case) as well as an adequate number of level one units (i.e survey individuals) inside those level two units . Whether placing MSOAs or LSOAs at level two there remains an adequate number of groups at this level. At level one, strict criteria to define 'adequate' sample size of survey individuals within those level two groups does not exist and depends in part upon the nature of the estimation at hand. As a general rule of thumb a level one sample size of ten inside the level two groups is becoming small and five is pushing the lower limits (Snijders and Bosker, 1999). Hence, if one is seeking to estimate at LSOA level rather than MSOA level then one is placing greater pressure on the sample size in terms of this need for an adequate sample of survey individuals inside the higher level target spatial units. Our sensitivity testing here therfore focuses on the more demanding LSOA target scale.

In these National Survey 2012-13 data there are 14,362 cases used and on average there are 10.3 individuals nested inside each LSOA area, with a median value of 9 individuals and 6 individuals at the 25th percentile (i.e 25% of LSOAs have fewer than this, 75% have more than this). These numbers are small but are viable for the estimation of the variance parameters within the multilevel model. One area of interest is to consider not just whether the estimation of the central point estimates is viable in smaller base surveys but, and more demanding in terms of the data requirements, whether the present approach to the estimation of the credible intervals around those point estimates remains viable in smaller base surveys.

If the same sampling approach were used (i.e. random sampling stratified by local authorities) but the survey were only 75% of the size then analyses suggest that there would be around 10,775 survey cases and an expectation of around 8 individuals in each LSOA on average. The median number of individuals in an LSOA would be around 7 and with 5 individuals at the 25th percentile. These numbers would be expected to remain viable for the estimation of the variance parameters within the multilevel model that are required to calculate the credible intervals. If the same sampling approach were used but the survey were only half the size, however, then one would quite possibly be unable to reliably estimate the required variance parameters in the multilevel model at this demanding LSOA scale. On average one would expect around 5.5 individuals in each LSOA in such a survey, with a median value of 5 individuals and just 3 individuals in each LSOA at the 25th percentile. This is very much at the lower end of what would be considered viable for the reliable estimation of the variance parameters.

The project included an empirical investigation of the impact of these alternatively-sized base surveys on the ability to derive LSOA level credible intervals. The analyses below take 100 separate stratified random samples of 75% and 50% respectively of the full National Survey and each time run the full multilevel model with individuals at level one nested inside LSOAs at level two. The estimate of the residual variance on the level two error term – the key term in defining the width of the credible intervals – is saved each time and plotted below in Figures 21 and 22 for the outcomes relating to internet use and feeling unsafe. On each chart a dashed vertical line is included to show the level of residual variance in the level two error term estimated within the same multilevel model when using the full National Survey 2012-13.

The mean estimates of these curves lie close to the dashed vertical line on each chart and this suggests that *on average* there are not dramatic impacts on the size of the residual variance in the level two error two – and hence in the width of the intervals – across base surveys of different sizes. However, in practice one would not be in a position to take the average of the level two residual error variance terms across 100 sub-samples given that one would not have access to this sampling distribution but would instead have collected just one of these (smaller) survey samples. Figures 21 and 22 also highlight threfore that although the means may be similar across the repeated sub-samples that there is greater *variability* in the residual variance in the level two error terms estimated across the 100 sub-samples when the survey sample size is smaller. The implication is that, other things equal, smaller base surveys induce a higher chance of delivering credible intervals that are wider *or* narrower than is the case with larger base survey files where the residual variance on the level two error term clusters more tightly around the central value.

Taken together these analyses suggest firstly that the calculation of credible intervals to LSOA level based on the approach used here is viable but that it requires relatively large base survey sizes – in the order of 7,500 randomly sampled cases at a minimum in the base survey. The analyses suggest secondly that there is reduced potential volatility in the potential size of the resulting residual variance on the level two error term (and hence resulting credible intervals around the central IPF point estimates) in larger base survey samples compared with smaller base survey files.

Figure 21: Sensitivity testing the impact of reduced base survey size on the credible intervals around the LSOA estimates of internet use

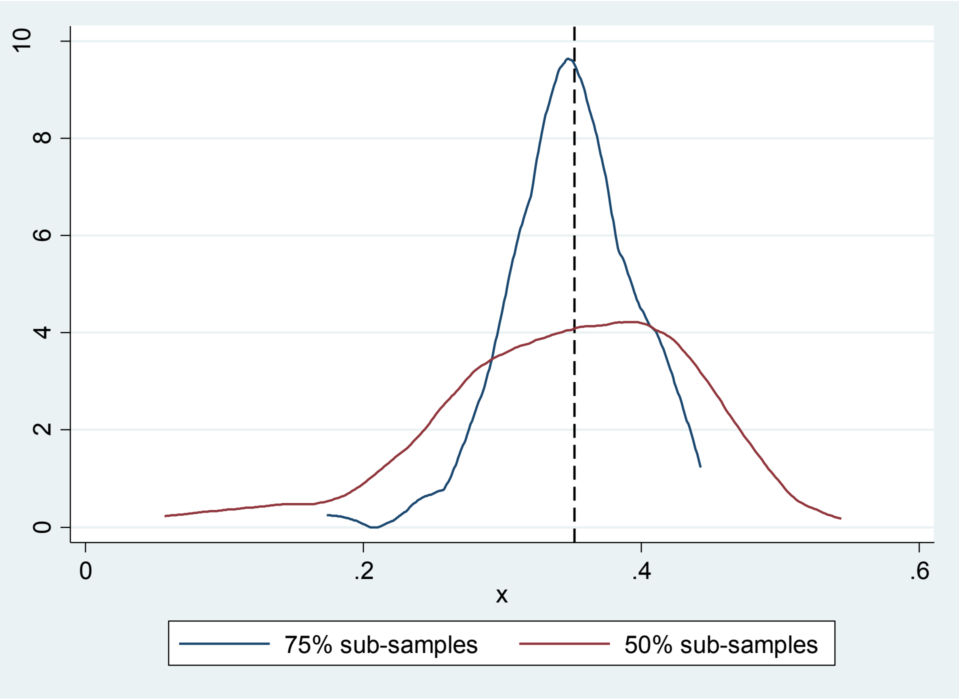
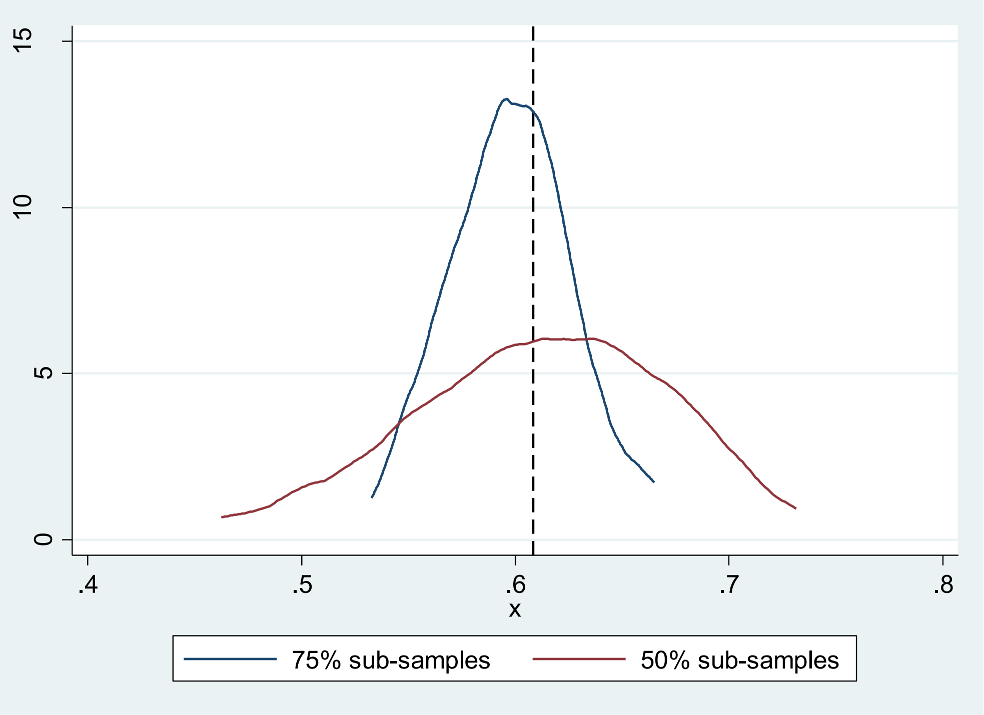


Figure 22: Sensitivity testing the impact of reduced base survey size on the credible intervals around the LSOA estimates of feeling unsafe



Looking at the local authority level: estimating to local authority level using smaller base surveys

A final and related area of interest is to explore the possibility of using small area estimation techniques to create indirect estimates at the larger local authority level. If possible, such work might help to reduce costs by enabling smaller surveys than the current National Survey sample size or, alternatively, may make smaller scale but more specifically focussed surveys more viable in future.

In order to explore this issue, the natural starting point is to take the current best estimates that can at present be derived by weighted analyses of the National Survey. As described above, the National Survey 2012-13 contains over 14,000 cases and is a stratified sample across local authorities, with roughly equal numbers of cases in each authority. The key advantage of this large and stratified sampling approach is that relatively precise survey estimates can be gained at local authority level. The main disadvantage is that the data collection required is extensive given that each local authority requires a relatively large sample size.

This local authority evaluation exercise begins by creating National Survey estimates of the percentage of adults in each local authority with each outcome (e.g feeling unsafe, using the internet, etc). This represents the current best local authority estimate of these six variables.

In order to create indirect local authority estimates on base surveys of different sizes, the IPF method is carried out for each local authority using the same set of constraint variables as above. Two differently sized smaller base surveys are tested here based on (i) half the size of the full National Survey and (ii) a quarter the size of the full National Survey with each drawn from the local authority survey strata such that all local authorities retain an even (albeit reduced) sample size.

In order to reflect the potential variability of the actual sample drawn, 50 different samples of each sub-sample size (i.e. 50% and 25% of the full National Survey) are drawn stratified randomly from the full National Survey and separate IPF estimates for the six outcome variables are calculated for each sample. All 50 samples are plausible and, consequently, this distribution of estimates is presented as an indication of the likely range that any *one* such set of indirect estimates would have been expected to fall into *if* the National Survey 2012-13 had instead been far smaller and *if* IPF had been used to create the local authority estimates.

Figures 23 and 24 below shows all results for two single local authority areas: Blaenau Gwent in Figure 23 and Wrexham in Figure 24. The distribution of estimates is presented as a boxplot for each outcome variable: the central shaded area shows the interquartile range[5] of the distribution of estimates with the median marked as a

[5] The interquartile range covers the central 50% of cases beginning at the value one quarter of the way through the distribution (i.e. leaving one quarter of cases with values lower than this point) and ending at the value three quarters of the way through the distribution (i.e. leaving one quarter of cases with values greater than this point).

horizontal line within that box; the vertical lines – or 'whiskers' – coming out of the box cover the remaining values up to 1.5 times the interquartile range; and the dots show outliers beyond this range. The central direct survey estimate from the National Survey is shown as a single horizontal line for each outcome. These direct survey estimates are estimated with confidence intervals around them (typically in the order of 3 to 5 percentage points either side of these central survey estimates) but these are not displayed below and the focus remains on the fit of the IPF estimates to the central survey estimate.

For Blaenau Gwent, the IPF estimates relating to feeling unsafe are most similar to the National Survey direct estimates whilst the estimates for area satisfaction and for financial difficulties are less successful. For Wrexham, in contrast, it is feeling unsafe which stands out as the least successfully estimated but with most other outcomes showing a relatively good degree of fit.

Figure 23: Direct versus indirect estimates for Blaenau Gwent

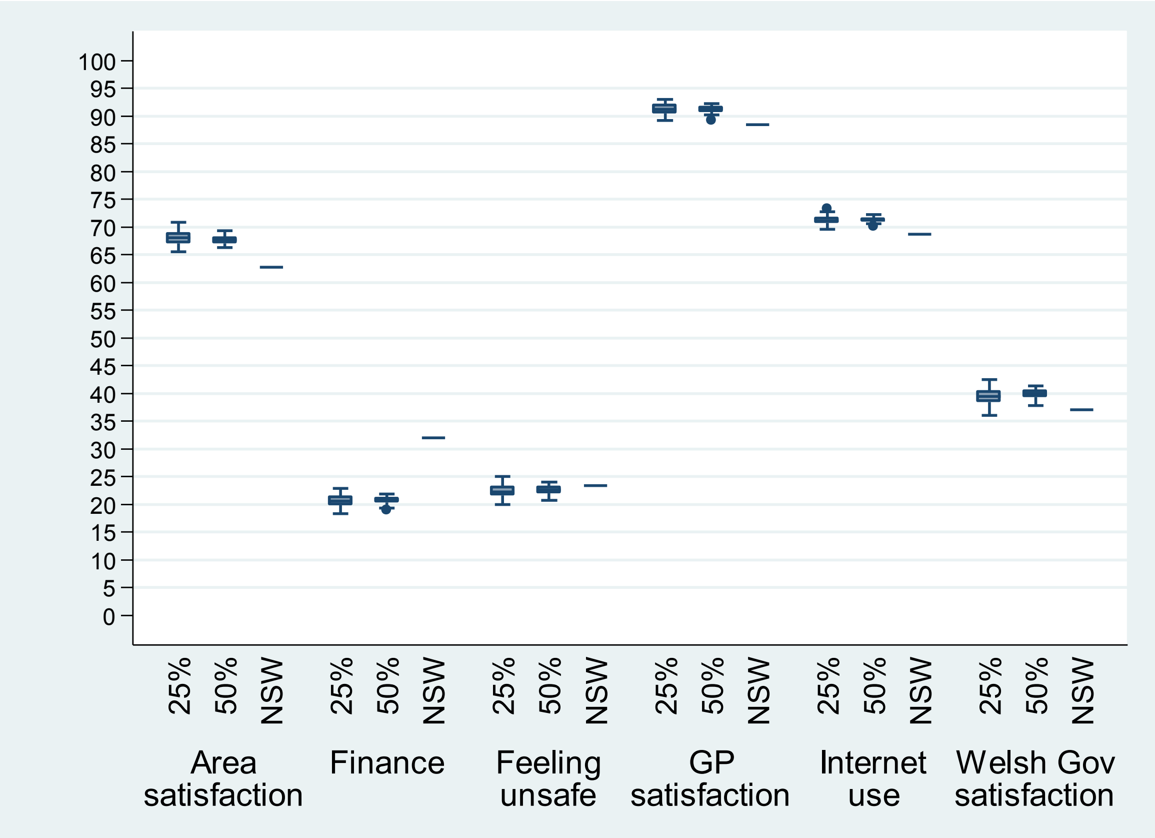


Figure 24: Direct versus indirect estimates for Wrexham

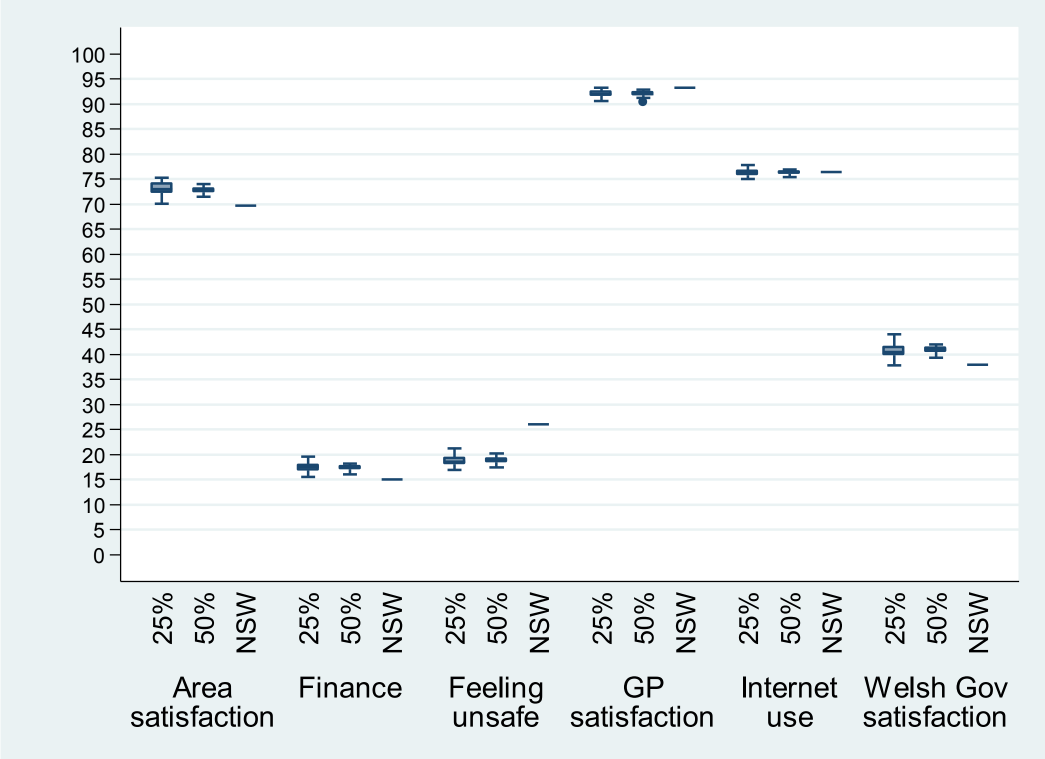
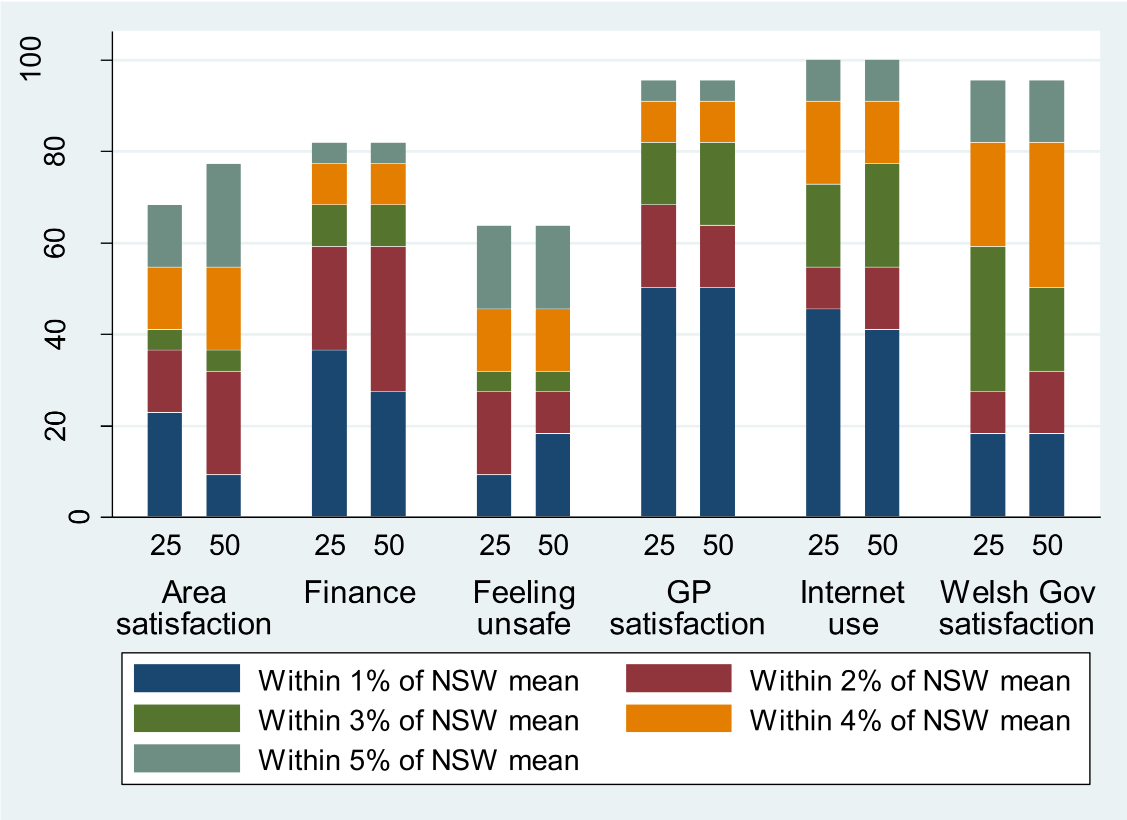


Figure 25 below presents an overview of the performance of the IPF estimation of all outcome variables across all of Wales' 22 local authorities. The raw data underpinning Figure 25 are provided in Table 10 in the Appendix. For each outcome variable every local authority has 50 sets of IPF estimates based on samples either one half or one quarter the size of the National Survey respectively. For each local authority the mean estimate is calculated from these 50 estimates and this mean value is then compared to the direct survey estimate from the National Survey. The stacked bars in Figure 25 show the percentage of Wales' local authorities that are within between 1 and 5 percentage points of the National Survey direct estimates based on these mean IPF values. Figure 25 therefore offers a concise visual summary of the degree of correspondence that one could expect from such local authority level IPF estimation.

As an example, if one focuses on the IPF estimation of internet use from a base survey 25% of the full National Survey then just over 40% of Wales' local authorities have a mean value for those IPF estimates within 1 percentage point of the mean direct estimate from the National Survey. On this same measure a little over 70% of Wales' local authorities have a mean value for those IPF estimates within 3 percentage points of the mean of the internet use direct survey estimates, and almost all local authorities are within 5 percentage points of the central direct survey estimate.

Looking across the six sets of estimates as a whole, for three outcomes (GP satisfaction, internet se, high satisfaction with Welsh Government) all or almost all local authorities' IPF estimates are within 5 percentage points of the direct National Survey estimates. This is the case for over 80% of Welsh local authorities in relation to the estimates of adults experiencing financial difficulties and over 60% of local authorities for the IPF estimates relating to area satisfaction and feeling unsafe. If one is very stringent with the degree of correspondence demanded then one can see that around 60% of local authorities can be estimated to within 2 percentage points of the means of the National Survey direct estimates for three of the outcomes (financial difficulties, GP satisfaction, internet use) but this applies to only around 30% of local authorities in relation to the remaining three outcomes (area satisfaction, feeling unsafe, high satisfaction with Welsh Government).

Figure 25: A summary of direct vs indirect estimates across all Welsh local authorities



One would always expect a degree of error in any estimation process. However, it may be that estimation could be improved somewhat by pre-estimation case tailoring and/or post-estimation calibration/constraining. Pre-estimation tailoring could involve either more locally specific SAE models either through tailoring the selection of cases (eg via selection cases only in that region, geodemographic type or deprivation quintile for example) and/or the selection of constraint variables. In terms of post-estimation calibration/constraining, one might seek to use information about either known trends in error (e.g. or information from external sources (potentially administrative data, Census, or other surveys) to systematically adjust the small area estimates such that they are constrained to sum to ‘true’ local authority level (or other higher scale) values. One might do so based on an assumption of uniform spatial error at small area level.

Alternatively, it may be possible to learn something about the small area spatial distribution of the error through the parallel small area estimation of an outcome variable for which ‘true’ small area values are known (e.g. general health status). Although the variable used for calibration is unlikely to be the target variable being estimated (else there would be no need for the small area estimation), this approach may give some insight into the reasons why results in some local authorities are systematically less accurate. This would require further consideration of whether trends on this proxy outcome variable can be assumed to also hold for the target variable of interest.

Although post-estimation calibration/constraining is not common it has been used previously within the small area estimation literature (ONS, 2003; Scholes et al., 2007). Further exploration and testing would be required to assess the possible magnitude of any potential improvements from such pre-estimation selection and/or post-estimation calibration strategies.

Discussion

The main aim of this project was to produce small area estimates for six diverse outcome variables using the 2012-13 National Survey for Wales. Through doing so the projects was also tasked with offering back to the Welsh Government recommendations and considerations to guide future small area estimation work.

To achieve these aims the report has provided an overview of the two main methodological frameworks to conduct SAE – statistical approaches and spatial microsimulation approaches – as well as a summary of the main specific methodological techniques within each of these two broad overarching frameworks. This offers broad understanding around the range of potential SAE methodologies that are available to be used potentially in any future work as well as the principles, strengths and weaknesses of each of these alternative approaches. This overview also serves to situate the particular methodological approach used in this project – iterative proportional fitting (IPF). The report provides a detailed step-by-step account of the IPF process so that its steps and decisions taken within it can be transparently understood and, if desired, followed or amended in future.

The project has produced small area estimates with accompanying 95% credible intervals for the 410 Middle Layer Super Output Areas (MSOAs) across Wales for six diverse outcome variables. These show the estimated percentage of adults aged 16+ across Welsh MSOAs who: use the internet; are experiencing financial difficulties; feeling unsafe in the local area after dark; are satisfied with their GP care; who are highly satisfied with their local area; and who are highly satisfied with the performance of the Welsh Government. Credible intervals have been provided around the central MSOA estimates and these provide a sense of the uncertainty and plausible error around the point estimates. In terms of their meaning, the credible intervals show the estimated range within which we can be statistically confident that the ‘true’, but unknown, underlying small area population values for these outcomes falls. These credible intervals are relatively wide and, as is often the case in small area estimation work, are often overlapping. Where the credible intervals of different small areas overlap, therefore, this suggests that we cannot be statistically confident in stating that the areas necessarily have different underlying population values for the outcome, even if the central small area point estimates do suggest a difference.

The IPF validates well in terms of the internal validation, though this would be expected given the nature of the IPF methodology. The external validation of the aggregated small area estimates relies on a comparison between direct survey estimates for local authorities derived from the National Survey and the indirect MSOA small area estimates aggregated to the local authority level. Although based only on 22 local authority observations, scatterplots show that the two sets of estimates are, in general terms, comparable, though they show inevitable variation around the line of equality between the two sets of estimates. Some of the estimated outcomes show flatter distributions compared with the greater variation between local authorities seen in the equivalent direct survey estimates and this suggest that for those outcomes this IPF specification may not be fully able to capture the variation in these outcomes. Further work might explore the potential to incorporate further (particularly area level) variables into the IPF in order to seek to better explain the variation in these outcomes. Statistical analysis of the mean absolute error of the estimates shows that the aggregated small area estimates are on average between roughly two and four percentage points of the local authority estimates taken directly from the National Survey. Most local authorities show smaller differences and a minority of local authorities with sometimes notably larger differences pull these mean differences upwards. There are no benchmarks in the literature as to what constitutes acceptable fit and this is largely a subjective decision based on the use to which the analyst or policy maker is relying on the estimates and the extent to which certainty around precision is critical.

The local authority analyses highlight the viability of using IPF to create local authority level estimates from smaller base surveys. These results are benchmarked against direct estimates from the National Survey and analyses are presented of the extent of correspondence between these direct survey estimates and indirect IPF estimates. Whilst there is inevitably variation across outcomes, and whilst the decision around how much error is considered to be acceptable is inevitably a subjective and context-specific one, the results are in general supportive of the idea that local authority level estimation from smaller base surveys is a viable proposition.

In terms of future potential SAE work, four possibilities emerge. Firstly, future work could explore the incorporation of additional (especially area level) factors in the IPF process. Depending on the nature of the constraint and outcome variables involved, this may have benefits firstly for the accuracy of the central point estimates by increasing the amount of the variation in the outcome that is able to be accounted for. This may, in turn, increase the spread of the small area estimates across the small areas. Secondly, this would be expected to have benefits in terms of reductions in the width of the credible intervals by reducing the amount of residual variance on the area level error term within the multilevel regression models from which the credible intervals are calculated. This would be expected to be especially powerful where area level factors play an important role in explaining variation in the outcome variable.

Secondly, as with virtually all small estimation projects it is inevitably difficult to assess the extent to which the *distributions* of the small area estimates *within* the validated local authority level are accurate. Certainly one would expect greater variability at that smaller scale. One possibility would be to conduct surveys in specific local areas in order to be able to calculate sufficiently precise direct estimates at the small area level against which to externally validate the small area estimates at this small area scale. This would represent the gold standard test in terms of the external validation of these small area estimates and of the IPF methodology more broadly. Another possibility may be to combine multiple years of survey data in order to create sufficiently precise small area direct estimates against which to compare the indicate small area estimates at those small scales, assuming that datasets across different years contain comparable indicators that can meaningfully be combined.

Thirdly, the external validation exercise highlights that the aggregated small area estimates are a better fit to the direct survey estimates in some local authorities compared to others. It is not clear, however, why certain local authorities appear to produce more robust estimates than others or why some local authorities might appear as outliers in this validation exercise. A better understanding of these issues may help to identify the factors that affect the accuracy of the estimation process across differing local contexts so as to be able to produce better estimates in (potentially atypical) small area areas where the SAE may not be as effective at present.

Fourthly, targeted testing of the viability of producing small area estimates of internet use to the smaller LSOA scale across Cardiff suggests that this may well be possible. Although MSOAs do offer significant advances to policy understanding in terms of the level of spatial detail that they offer, this level of spatial detail and subtlety inevitably improves somewhat if estimates could in future be produced at LSOA level. Moreover, it does also appear viable to produce acceptable local authority level estimates based on smaller base surveys and further work around the potential to enhance the local specificity of these local authority estimates could also be explored. Sensitivity testing demonstrates the likely impacts of such changes on the expected validity of any resulting estimates and the acceptability of such estimation would be a subjective decision based on the use for which estimates are needed and the degree of precision that is therefore required.

References

Anderson, B (2007) Creating *small area income estimates for England: spatial microsimulation modelling*, a report to the Department of Communities and Local Government. London: Department of Communities and Local Government.

Ballas, D., Clarke, G., Dorling, D., Eyre, H., Thomas, B and Rossiter, D (2005) 'SimBritain: A Spatial Microsimulation Approach to Population Dynamics', *Population, Space and Place*, 11, 13-34.

Ballas, D., Campbell, M., Clarke, G., Hanaoka, K., Nakaya, T and Waley, P (2012) 'A spatial microsimulation approach to small area income estimation in Britain and Japan', *Studies in Regional Science*, 42, 163-187.

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

Ballas, D., Dorling, D., Anderson, B. and Stoneman, P (2006) *Assessing the feasibility of producing small area income estimates*. London: Office for the Deputy Prime Minister.

Chambers, R and Tzavidis, N (2006) 'M-quantile models for small area estimation', *Biometrika,* 93, 255-268.

Deville, J-C and Sarndal, C-E (1992) 'Calibration estimators in survey sampling', *Journal of the American Statistical Association*, 87, 418, 376-382.

Edwards, K and Tanton, R (2013) *Spatial microsimulation: A reference guide for users.* London: Springer.

Ghosh, M. and Rao, J. (1994) 'Small area estimation: an appraisal', *Statistical Science*, 9, 1, 55-76.

Gomez- Rubio, V., Best, N., Richardson, S., Li, G and Clarke, P (2010) *Bayesian statistics small area estimation*, an NCRM Methodological Review Paper http://eprints.ncrm.ac.uk/1686/

Haughton, D and Haughton, J (2011) 'Multilevel models and small-area estimation' in Haughton, D and Haughton, J (eds.) *Living standards analytics: Development through the lens of household survey data.* London: Springer

Heady, P., Clarke, P., Brown, G., Ellis, K., Heasman, D., Hennell, S., Longhurst, J. and Mitchell, B (2003) *Model based area estimation series No. 2: Small area estimation project report*. London: Office for National Statistics.

Hermes, K and Poulson, M (2012) ‘Current methods to generate synthetic spatiaol microdata using reweighting and future directions’, *Computers, Environment and Urban Systems,* 36, 281-290.

Marchetti, S., Tzavidis, N and Pratesi, M (2012) ‘Non-parametric bootstrap mean squared error estimation for M-quantile estimators of small area average, quantiles and poverty indicators’, *Computational Statistics and Data Analysis,* 56, 10, 2889-2902.

Marshall, A (2010) *Small area estimation using ESDS government surveys – An introductory guide*. Economic and Social Data Service.

Molina, I and Rao, J. (2010) ‘Small area estimation of poverty indicators’, *The Canadian Journal of Statistics*, 38, 3, 369-385.

ONS (2003) *Model-based ward estimates of crime, 2001/02: Feasibility Report.* Unpublished draft report.

Pickering, K., Scholes, S and Bajekal, M (2004) *Synthetic estimation of healthy lifestyles indicators: Stage 3 report.* London: NatCen.

Rao, J (2003) *Small Area Estimation.* New York: Wiley.

Rahman, A (2008) *A review of small area estimation problems and methodological developments.* University of Canberra: NATSEM Discussion Paper Issue 66.

Scholes, S., Pickering, K and Rayat, P (2007) *Healthy lifestyle behaviours: Model-based estimates for Middle Layer Super Output Areas and Local Authorities in England, 2003-2005: User Guide*. London: NatCen.

Snidjers, T and Bosker, R (1999) *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. London: Sage.

Tanton, R., Vidyattama, Y., McNamara, J., Vu, Q and Harding, A (2009) ‘Old, Single and Poor: Using Microsimulation and Microdata to Analyse Poverty and the Impact of Policy Change among Older Australians’, *Economic Papers: A journal of applied economics and policy*, 28, 2, 102–120.

Tanton, R and Vidyattama, Y (2010) ‘Pushing it to the edge: Extending generalised regression as a spatial microsimulation method’, *International Journal of Microsimulation*, 3, 2, 23–33.

Tanton, R., Vidyattama, Y., Nepal, B and Mcnamara, J (2011) ‘Small area estimation using a reweighting algorithm’, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174, 4, 931-951.

Tzavidis, N., Marchetti, S and Chambers, R (2010) ‘Robust estimation of small area means and quantiles’, *Australian and New Zealand Journal of Statistics*, 52, 2, 167-186.

Voas, D and Williamson, P (2000) ‘An evaluation of the Combinatorial Optimisation approach to the creation of synthetic microdata’, *International Journal of Population Geography*, 6, 349-366.

Whitworth, A. (2012) ‘Sustaining evidence-based policing in an era of cuts: Estimating fear of crime at small area level in England’, *Crime Prevention and Community Safety*, 14,1, 48-68

Whitworth, A (2013) (ed.) *Evaluations and improvements in small area estimation methodologies.* Economic and Social Research Council: National Centre for Research Methods methodological review paper.

Williamson, P (2013) ‘An evaluation of two synthetic small-area microdata simulation methodologies: synthetic reconstruction and Combinatorial Optimisation’ in Tanton, R and Edwards, K (eds) *Spatial microsimulation: a reference guide for users*. London: Springer.

Appendix

Table 7: Results of the MSOA level multilevel models

Explanatory Variables		Satisfaction with Area	Feeling unsafe	Satisfaction with Welsh Gov	Satisfaction with GP care	Finance Problems	Internet Use
Tenure (ref=owned)	Social Rent	-0.23*	0.19*	0.05	-0.15	0.83*	-0.58*
	Private Rent	-0.07	-0.00	0.13*	0.33*	0.66*	-0.21*
Employment Status (ref= unemployed)	Working	0.46*	-0.35*	0.22*	0.09	-0.98*	-0.11
	Retired	1.50*	1.83*	0.12	1.23*	-1.87*	-3.92*
	Inactive	0.38*	-0.01	0.16	0.18	-0.57*	-0.66*
	Full-time Student	0.39*	-0.24	0.70*	1.00*	-0.77*	0.62
(ref= no health problems)	Health Problems	-0.42*	0.81*	-0.19	-0.66*	0.68*	-0.15*
(ref = no car)	Access to Car	0.06	-0.22*	-0.16	0.07	-0.16*	1.19*
Qualifications (ref= no/low qualifications)	Medium Qualifications	-0.08	-0.14*	-0.04	-0.14	-0.18*	1.02*
	High Qualifications	-0.22*	-0.36*	-0.01	-0.31*	-0.52*	1.87*
(ref = no child in household)	Dependent Child	-0.03	0.09	0.14*	-0.03	0.32*	0.58*
House Type (ref=detached)	Semi-Detached	-0.47*	0.18*	0.16*	0.05	0.41*	-0.13*
	Terraced	-0.83*	0.30*	0.05	-0.13	0.51*	-0.25*
	Flat	-0.78*	0.47*	0.13	-0.10	0.19	-0.06
	Other Dwelling	-0.28	-1.07	0.08	-0.37	1.06*	0.08
Age-Sex Group (ref= Male 16-29)	Male 30-49	0.23*	0.16	-0.36*	0.19	0.20*	-0.92*
	Male 50-64	0.59*	0.16	-0.22*	0.61*	0.00	-1.96*
	Male 65+	-0.05	-1.29*	0.08	0.25	0.07	0.31*
	Female 16-29	0.09	1.53*	-0.19	-0.19	0.01	0.05
	Female 30-49	0.30*	1.38*	-0.38*	0.31	0.30*	-0.67*
	Female 50-64	0.56*	1.44*	-0.27*	0.84*	-0.02	-1.87*
	Constant	0.74*	-2.48*	-0.46*	2.01*	-1.23*	1.90*
	Observations	14327	14108	13250	11428	13985	14359
	Residual variance on level two error	0.182	0.345	0.047	0.127	0.223	0.036

Table 8: Local authority differences between direct estimates and aggregated small area estimates

	Internet	Financial difficulties	Feeling unsafe	Satisfaction with WG	Satisfaction with GP care	Satisfaction with local area
Blaenau Gwent	3.9	10.8	1.1	2.6	2.7	4.3
Bridgend	1.4	5.7	0.2	0.1	0.3	1.4
Caerphilly	5.1	11.7	0.3	1.8	5.1	3.9
Cardiff	2.5	1.2	7.7	4.8	0.5	3.6
Carmarthenshire	3.7	1.1	5.4	0.7	0.1	9.4
Ceridigion	0.4	2.1	5.4	6.0	0.1	4.2
Conwy	1.2	1.8	1.5	1.0	0.2	1.3
Denbigshire	0.1	4.0	4.3	3.3	0.8	1.4
Flintshire	1.1	0.6	6.3	2.9	0.1	3.0
Gwynedd	2.2	0.4	8.3	3.3	2.0	3.0
Isle of Anglesey	1.7	1.3	5.8	3.5	3.1	1.4
Merthyr Tydfil	2.8	5.9	4.4	3.2	4.6	4.4
Monmouthshire	3.9	0.5	2.7	2.7	0.3	1.1
Neath Port Talbot	2.1	0.2	1.2	4.2	0.8	2.6
Newport	1.8	1.1	4.0	1.5	2.1	7.3
Pembrokeshire	3.2	4.5	7.5	2.6	1.2	4.6
Powys	0.4	2.5	4.3	4.9	0.6	2.0
Rhonda, Cynon, Taf	1.0	1.7	3.6	0.1	2.6	9.1
Swansea	1.2	1.5	1.3	0.2	0.5	1.0
Vale of Glamorgan	1.3	0.4	4.2	3.1	1.3	5.7
Torfaen	2.5	3.8	3.8	3.3	3.1	7.0
Wrexham	1.0	2.5	7.4	2.8	1.2	2.8

Table 9: Residual level two variance in the LSOA level multilevel models

Outcome Variable	Residual variance in level two error
Internet Use	0.123
Financial Difficulties	0.461
High Satisfaction with Welsh Government	0.091
Satisfaction with GP Care	0.181
Feeling Unsafe	0.370
Area Satisfaction	0.271

Table 10: Overall local authority distances between central direct National Survey estimate and indirect IPF estimates from smaller base surveys

	% of local authorities within various percentage point differences from the central direct National Survey estimate					Outcome
Sample size	**1 percentage point**	**2 percentage points**	**3 percentage points**	**4 percentage points**	**5 percentage points**	
25	45.5	9.1	18.2	18.2	9.1	Internet use
50	40.9	13.6	22.7	13.6	9.1	Internet use
25	36.4	22.7	9.1	9.1	4.5	Financial difficulties
50	27.3	31.8	9.1	9.1	4.5	Financial difficulties
25	9.1	18.2	4.5	13.6	18.2	Feeling unsafe
50	18.2	9.1	4.5	13.6	18.2	Feeling unsafe
25	18.2	9.1	31.8	22.7	13.6	Satisfaction with WG
50	18.2	13.6	18.2	31.8	13.6	Satisfaction with WG
25	50.0	18.2	13.6	9.1	4.5	GP satisfaction
50	50.0	13.6	18.2	9.1	4.5	GP satisfaction
25	22.7	13.6	4.5	13.6	13.6	Area satisfaction
50	9.1	22.7	4.5	18.2	22.7	Area satisfaction