

Local Geographies of Digital Inequality

Grant Blank
University of Oxford

Mark Graham
University of Oxford

Claudio Calvino
Capgemini UK

Please cite as: Blank, G., M. Graham & C. Calvino. (2017). Local geographies of digital inequality. *Social Science Computer Review*. DOI: 10.1177/0894439317693332

Local Geographies of Digital Inequality

Abstract

Combining data from a sample survey, the 2013 Oxford Internet Survey (OxIS), with the 2011 UK census, we employ small area estimation to estimate Internet use in small geographies in Britain. This is the first attempt to estimate Internet use at any small-scale level. Doing so, allows us to understand the local geographies of British Internet use: showing that the area with least use is in the North East, followed by central Wales. The highest Internet use is in London and south eastern England. The most interesting finding is that after controlling for demographic variables, geographic differences become non-significant. The apparent geographic differences appear to be due to differences in demographic characteristics. We conclude by considering the policy implications of this fact.

Keywords: small area estimation (SAE); Internet use; geographical inequality; digital divide; Great Britain;

Acknowledgements:

This work was supported by the Economic and Social Research Council [grant ES/K00283X/1].

The data have been deposited in the UK Data Archive under the name "Geography of Digital Inequality".

Introduction

The Internet has fundamentally reorganised economic, social and political actions and relationships around the world. It has sparked what has been dubbed as an “information revolution,” and given rise to an “information economy” and “information society.” An unimaginable amount of content can now be accessed from almost anywhere on the planet: content that ultimately shapes our understandings of the world and the ways in which we interact with our surroundings. Social life is increasingly mediated, influenced, and augmented by online interactions that take place through the Internet (Graham et al., 2013)

Yet despite the importance of the Internet in everyday life, we know surprisingly little about the geography of Internet use and participation at sub-national scales. It is unclear whether citizens of Aberdeen, Manchester, Cardiff, or London are significantly more likely to be able to access our global network. This is important because, as we discuss below, there are widespread geographic differences in online activity. Because the differences may indicate digital inequalities on a geographic level it is important to investigate further. As a result, this paper proposes a novel method to calculate the local geographies of Internet usage. In this paper, we employ Britain as a case study, but, in future work, hope to be able to replicate the method in other parts of the world.

Britain has one of the largest Internet economies in the developed world. The Internet contributes an estimated 8.3 percent to Britain’s Gross Domestic product (GDP) (Dean et al., 2012), a share of the economy that is larger than health care, construction or education. Furthermore, researchers at the Boston Consulting Group predict the Internet’s share of Britain’s economy will grow at a rate of 11 percent per year over the next four years (Dean et al., 2012). This is a faster growth rate than China’s digital economy and more than double the growth rate of the US. The increasing embeddedness of the Internet in the British economy potentially provides a range of tangible benefits. The Internet strongly supports domestic job and income

growth by reducing a range of geographic frictions and allowing access to new customers, sales channels, markets and ideas. Shortening supply chains and online efficiencies produce lower prices online, saving an estimated £18 billion per year in 2009 (Kalapesi et al., 2010) People benefit from better communications, giving them more contact with friends and better access to jobs. Businesses are more likely to locate in areas with good digital access, thereby boosting local economies and providing jobs that reduce unemployment (Malecki and Moriset, 2008)

In addition to benefits to the UK as a whole, there are personal benefits to Internet use. These benefits include lower prices, access to wider and cheaper entertainment, better and easier access to information, superior access to news, capability to publish your own thoughts to a (potentially large) audience, and—above all—cheaper, easier, faster communication of all kinds. The advantages of the Internet are denied to people who are not online.¹ This has led to a stream of research on people who are not online. The initial work (e.g. Katz and Aspden, 1997; Hoffman and Novak, 1998; van Dijk, 2006) focused on who did or did not have physical access in the form of a computer and an Internet link. Work on this *first-level digital divide* led to the recognition that Internet access was not enough. There were differences in how effectively people were able to take advantage of the benefits of the Internet (DiMaggio et al., 2004) Effectiveness is related to the digital skills of users (Hargittai, 2008) and their attitudes toward technology (Blank and Dutton, 2012) This research stream has no standard name but it is variously called the *second-level digital divide* (Hargittai, 2002), *emerging digital differentiation*

¹ Although we focus on the benefits of the Internet, we recognize that the Internet is also has a negative side: the closure of high street stores, movement of jobs to lower-paid transport and courier occupations, enhanced possibilities for identity theft and fraud, and threats to privacy, among others. The impact of the Internet is far from wholly positive.

(Peter and Valkenburg, 2006), the *usage gap* (van Dijk, 2006) or the *participation gap* (Jenkins et al., 2006).²

These studies have used national-level data. In Britain we know a great deal about the Internet at a national level because of the Oxford Internet Survey (OxIS): a representative sample of more than 2000 people carried out biennially since 2003. The most recent OxIS report shows that Britain is strongly stratified by age: 100 percent of 14-17 year olds are connected, compared to only 39 percent of those over age 65; by income: 58 percent of people earning less than £12,500 per year are connected compared to 99 percent of people with incomes over £40,000; and by education: 95 percent of those with a university degree are online compared to 40 percent with no educational qualifications (Dutton and Blank, 2013). In all these cases, and many others, the difference between the lowest and highest categories is a factor of at least 2 and often 3. These are very large differences, but they have a major weakness: they tell us nothing about potentially significant geographic inequalities in Internet use at sub-national levels.

There are two major reasons to suspect that geographic differences may be important: apparent regional differences and the urban-rural divide. At the regional level we know that Internet use and benefits are not evenly distributed. London and the southeast are the most wired part of the country (Dean et al., 2012). OxIS 2013 shows that regional Internet use varies from 60 percent in the North East and 71 percent in Wales to 86 percent in London and 83 percent in the South East (Dutton and Blank, 2013). This difference—over 25 percentage points—suggests the possibility of strong geographic inequalities. These differences in Internet use mirror other,

² In the context of a discussion of being online, it is important to realize that people are online whether they want it or not. Bus passes, credit and debit cards, pictures other people have taken and put online, mobile phone mapping and geolocation services, and—soon with the Internet of things—all sorts of home devices put information about people online. It is almost impossible to avoid an involuntary Internet presence.

offline differences. Champion et al. (1987) document that southern England is better off than the rest of the UK on a host of characteristics, ranging from employment rates to health conditions. More narrowly, Dunford (1995) points out that many characteristics favor the Greater London metropolitan area compared to the rest of the UK. To this extent, the percentage differences in Internet use seem consistent with offline inequalities.

Inside a country the major geography-related divide is the urban-rural divide. Internet access in rural areas may improve economic opportunities by promoting the development of home businesses (LaRose et al., 2007), increase access to health care (Jenkins, 2003) and education (Horrigan and Murray, 2006). Indeed Parker (2000) argues that the payoff for Internet development is higher in rural areas than in urban areas. Urban areas are more connected than rural areas (e.g. Crang et al., 2006; Townsend, Sathiaselalan, Fairhurst and Wallace, 2013) and suburban teenagers have adopted the Internet faster than inner city teenagers (e.g. Lichy, 2011; Zhao, 2009).

Several articles argue that lack of broadband in rural areas is a significant problem (Ashmore et al, 2015; Townsend et al., 2013; Philip et al., 2015, Philip et al., 2017), but this research is based on a correlation: rural areas have less broadband access and lower rates of Internet use. The problem is the lack of multivariate analysis. Rural and urban areas differ on many characteristics.

Both regional differences and urban-rural differences are confounded by other factors that influence in Internet use, like age, education and occupation. In fact, several studies have found that, once demographic differences are statistically controlled, there is no difference between rural and urban areas (Government Accounting Office, 2006; Horigan and Murray, 2006; Mills and Whitacre, 2003). For example, rural populations tend to be much older than urban populations. As we will see in the results, Figure 2, there are very strong spatial patterns in age. The first law of the Internet is that everything is related to age. Older people are less likely

to be Internet users in Britain (Blank and Groselj, 2014; Dutton and Blank, 2013). Older people spend less time on the Internet and they do fewer activities (Blank and Groselj, 2014). So controlling for age is likely to reduce apparent urban-rural differences. The broadband research stream does not attempt to assess the relative strength broadband access compared to other variables that have an important impact on Internet use, such as age.

Education is a second major influence on Internet use. Better educated people are much more likely to be Internet users, and more likely engage in most types of Internet activity; the exception is that they are less likely to use instant messaging (Blank and Groselj, 2014; Bonfadelli, 2002; van Dijk and van Deursen, 2013; White and Selwyn, 2013). We will see strong spatial patterns in age in the results, Figure 4. Rural populations tend to be less well educated, so controlling for education would reduce urban-rural differences.

Finally, people with higher incomes or levels of economic well-being are more likely to be Internet users. This is a common finding throughout the literature (Dutton and Blank, 2013; van Deursen and van Dijk, 2013). People with higher incomes are typically early adopters of all innovations (Rogers, 2003). Zillien and Hargittai (2009) found better educated individuals were more likely to use the Internet for “capital-enhancing activities”. Again, rural incomes tend to be lower.

The effects of potential causal variables can only be sorted out with multivariate studies. Almost all of the multivariate scholarly research on urban-rural differences has been done in the US. It is not clear how well this research applies to the UK because the US has large areas, for example the Dakotas, Wyoming and Montana, that are far more rural than anything in the UK.

The bottom line is that, although there has been a great deal of work on who uses the Internet, very little of that research has had an explicit spatial focus. This stems from a simple reason: geographic data on Internet use have not been available. Sample surveys have been the basis for most of what we know about digital inequality. National sample surveys cannot work

with smaller, sub-national areas because the sample sizes would be too small to allow accurate generalization. Consequently, there has been little work on the geography of digital inequality.

It should also be pointed out that about a decade ago, there was a flurry of work which did look at small area data. Anderson (2007) used spatial microsimulation to estimate time spent online at small area levels: providing sample results for a few parts of the country. Longley and Singleton (2009) expanded on this work by looking at links between material deprivation and digital exclusion. Longley et. al. (2008) similarly developed a nationwide household classification of ICT use. The most detailed British work is Ofcom's reports on broadband speeds and penetration at the county level (Ofcom, 2011). This is an important but limited topic: there is much more of interest on the Internet in addition to broadband. Because the broadband data were obtained from commercial providers who hold them as a trade secret, data below the Local Authority level are not public.³ We know little more than that about the geography of digital inequality. Given the importance of the Internet in contemporary Britain, the lack of existing work on the local geographies of the Internet is surprising. To fill this gap our work aims to answer the following research question.

Can geographic differences in Internet use be explained by demographic characteristics or are spatial factors, e.g. regional differences, needed to account for them?

Methods

This paper estimates Internet usage in England, Scotland, and Wales by using small area estimation. We do this with two primary datasets: The 2013 Oxford Internet Survey and the 2011 UK national census.

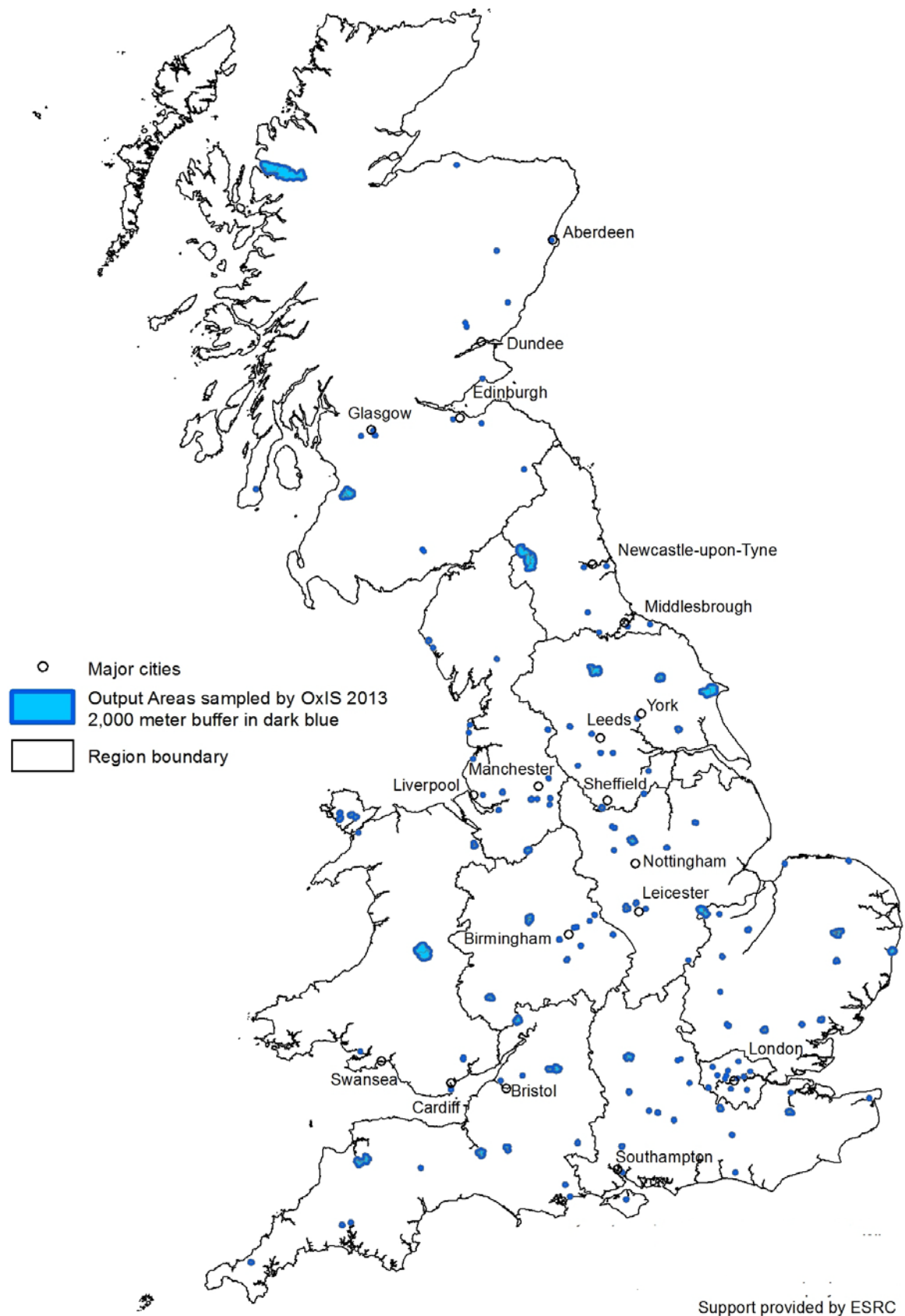
³ "Local Authorities" are a generic name for counties, metropolitan districts, Scottish council areas and Welsh unitary authorities, but this is too simple. British administrative geography is complicated and confusing. We explain it in more detail later in the paper.

The Oxford Internet Survey (OxIS), collects data on British Internet users and non-users. Conducted biennially since 2003, the surveys are random samples of more than 2,000 individuals aged 14 and older in England, Scotland, and Wales. Interviews are conducted face-to-face by an independent survey research company. The response rate for 2013 is 51 percent. The sample closely matches the census; we discuss this in detail when we describe individual variables, below. The data collection is a probability sample based on a two-stage cluster design. In the first stage, census output areas (OAs) were paired with the adjacent OA that was most similar in terms of its socio-demographic category.⁴ 134 pairs of OAs were randomly selected to form the Primary Sampling Units (PSUs). In the second stage, respondents were randomly selected within each PSU (see Dutton and Blank (2013) for details of the data collection and sample). Because not every OA contained respondents, there are a total of 260 OAs in the sample. The mean number of respondents is 10 per OA with a standard deviation of 5.9 and a range is from 1 to 20. The mean proportion of Internet users per OA is 0.78 with a standard deviation of 0.15 and a range from 0.17 to 1.00. Figure 1 shows the distribution of the sample of OAs included in OxIS. This illustrates the distribution of 260 OAs out of the 227,759 OAs in the 2011 Census.⁵

⁴ We measure socio-demographic category using the ACORN classification system. This categorizes OAs in the United Kingdom into one of 62 different socio-economic types, see CACI (2014) for details.

⁵ OAs are defined based on population so the OAs that are geographically large are in rural areas with few people. OAs in densely populated urban areas like London or Manchester cover much smaller areas; to make them big enough to be visible at this scale all sampled OAs are surrounded by a dark blue buffer of 2,000 meters. In the small urban OAs only the buffer is visible, which is the reason for including it.

Figure 1. Output Areas sampled by OxIS 2013



OAs are the fundamental building block of the British Census. They are the smallest geographical area for which the British 2011 census reports data. OAs are intended to have a minimum size of 40 households and 100 people and a maximum size of 250 households and 625 people, and are designed to be as socially homogeneous as possible (based on household tenure, dwelling type and urban/rural category) Scotland conducts a separate census, although the Scottish and English/Welsh censuses are coordinated so that the questionnaires are similar, the enumeration is on the same day, and the reporting is similar. OAs in Scotland are smaller, with a minimum size of 20 households and 50 people, and homogeneity is not a criterion in their construction (ONS 2014). In the actual census data, the largest OA has 4,092 people and there are 108 OAs with populations greater than 1,000. The census documentation provided by the Office of National Statistics (ONS) does not explain why these large OAs exist. The OAs are central to this project since they exist in both the census and OxIS. We use the information on Internet use in OxIS to estimate Internet use (and other variables) for all census OAs across Britain.

We create our estimates using a technique called small area estimation (SAE). SAE addresses two fundamental problems. First, survey data is not available for sub-national areas. No one knows, for example, the proportion of Internet users in Glasgow, Manchester, or Cardiff because national sample surveys don't have enough respondents in those cities to make reliable statistical estimates. National surveys like OxIS are not designed to estimate small areas. The sample size of the survey is too small to make detailed inferences about the geographies of digital inequality below level of the 11 British regions, but OxIS uses a random sample of OAs and this opens up a range of possibilities.

Second, the census enumerates everyone and census data are available for very small areas. However, the census is very expensive and it collects few variables. It collects no data whatsoever on Internet use or Internet activity or mobile use of the Internet. This is a problem

because detailed information on the Internet use of the population in local areas is valuable to identify areas that would benefit most from policy intervention. Thus small area data is essential to understand where to target efforts to tackle problems like limited Internet access or limited ability to use the Internet. OxIS, on the other hand, is a very rich dataset with more than 550 variables measuring all kinds of Internet activity.

By combining the richness of OxIS survey data with the comprehensive small area coverage of the census we can use the strengths of one to offset the gaps in the other. Specifically, we follow a two-step process. First, we use the information that is reliably available in OxIS to create a model that estimates the proportion of Internet users in an OA. Second, we use the parameters from this model combined with census data to estimate the proportion of Internet users each OA in Britain. Once these estimates are available, we aggregate the estimates up to higher levels of geography. In this way we can estimate Internet use in Glasgow, Manchester, and Cardiff as well as other small areas in Britain. This procedure is called indirect, model-based or synthetic estimation. In recent years such SAE techniques have been widely used throughout Europe and North America (e.g. Molina and Rao 2010; Simpson and Tranmer 2005; Twigg and Moon 2002)

This procedure assumes that if the demographic characteristics of individuals influence their Internet use, then small areas will be different from each other to the extent that the demographic characteristics of their respective populations differ.⁶ For example, age is a fundamental variable that influences most Internet variables. If age were the only predictor of

⁶ The ONS explored separate models for the different nations (England and Wales) but concluded that a single model was appropriate (White et al. 2011), but does not explain why. In the OxIS data, sample sizes for Wales and Scotland or any of the other nine regions are too small for separate models. We know of no reason why the relationship of demographic variables and Internet use should be different in England compared to Wales or Scotland.

Internet use, then small areas would only differ in Internet use if the age profile of their population differed. So a strong argument can be made for assuming that we can use differences in the area demographic profile to predict area differences in Internet use.

This process has the advantage that it allows us to directly test our research question about spatial effects vs demographic effects. There are two possibilities. First, if there are spatial effects—contextual effects—on Internet use then they will show up as statistically significant after we introduce our demographic controls. If the apparent spatial effects are simply the result of differences in the demographic profile of different geographic areas then the spatial variables will not be significant after we control for demography.

The methodology works because OxIS data collection was based on small geographic areas; that is, OAs. Britain is virtually alone in using this approach. Almost all other countries collect Internet data using random-digit-dialing telephone interviews. This includes most of the countries that are part of the World Internet Project like the USA, Canada, Sweden Australia, New Zealand, and Switzerland (Russia is an exception) (Cole, 2012; Cole, 2013) It also includes the Pew Internet and American Life Project data. Eurostat assembles Internet data collected from national household surveys. But many countries collect their household data from telephone, mail, or web-based surveys (or a combination). This includes Germany, France, Austria, Sweden, Finland, the United Kingdom, and others (Eurostat, 2014) Telephone, mail and web-based surveys cannot be used for small area estimation because respondents are randomly scattered across the country and do not have the necessary geographic concentration. This means that most large countries cannot use small area estimation to estimate Internet use in small areas.

There are many types of SAE models, see Rao (2003) for a summary. The fact that we have OxIS data on relatively few OAs limits our choice of methods. We can directly estimate the proportion of Internet users in only the OAs sampled by OxIS, so we can't use any technique that relies on direct estimation. Other techniques that combine model-based estimates with direct

estimates—such as generalised regression synthetic estimators, composite estimators, or Fay-Herriot estimators—cannot be used for the same reason; the direct estimates are available for too few OAs. Instead we use a model-based procedure.

We estimated the parameters of our model using a linear, 2-level mixed model. The first level is the 134 OA pairs; the second level is the 260 OAs. We chose this procedure because it allows for inclusion of demographic variables while taking into account the fact the respondents in each OA and each OA pair will be more alike than a random sample of the population as a whole. Taking into account the structure of the clustering of respondents in OAs improves the accuracy of confidence intervals and significance tests. These confidence intervals will generally be smaller and more conservative than confidence intervals that ignore the clustered data (Goldstein, 2003). We include 5 fixed effects variables—region, age, lifestage, gender, and education—and a single random effects variable, OA pairs. These are random intercept models, to account for the fact that different OA pairs will have different overall levels of Internet use.

We weighted each OA according to the number of respondents, so that OAs with more respondents (which have smaller standard errors) are weighted more heavily. Respondents themselves are counted based on post-stratification weights that were created so that the sample matched the British population on gender, age, region, rurality, ACORN code, and household size (see Dutton and Blank (2013) for details). Since these are weighted mixed models, simple R^2 s are not available and we report BIC.

Variables

The data are proportions at the OA level; in the jargon of mixed models this is an area-level model. We cannot use individual level covariates because the census does not provide individual-level microdata; it only makes public summary data for each OA. The dependent variable is the proportion of Internet users in each of the 260 OAs. Internet use is measured by responses to the OxIS item “Do you, yourself, personally use the Internet on whatever device at

home, work, school, college or elsewhere or have you used the Internet anywhere in the past?” This broad question asks about any kind of Internet use. It encompasses Internet use out of the home, such as in a library or community centre, mobile use, and in remote rural areas, possible use in another town. We measure spatial effects using *Region*, a collection of 10 dummy variables; Scotland is the omitted region. *Gender* omits males. *Age* is a 7-category variable. Preliminary work showed that the relation between age and Internet use is piecewise linear. The proportion of Internet users has a single shallow slope between age 14 to age 51. There is a discontinuity beginning at age 52 where the slope changes abruptly; it remains linear but it is much steeper. Since age is always a strong explanatory variable we wanted to capture as much variation as possible. We divided age into three categories between age 14-51, with approximately equal number of cases in each category. Because the slope was steeper for people age 52 and above we created four categories each containing approximately equal numbers of cases (see Table 1 below). The four categories balance our desire for as much fine-grained result as possible with the need for a minimum number of cases in each category. The disadvantage of this approach is that the age categories are not nice round numbers ending in zero or five. *Lifestage* is a four-category variable where we combined census categories to match the OxIS categories: student, employed, unemployed, and retired. For the four-category *education* variable, we also combined census categories to match the OxIS categories: no qualifications, secondary school, further education, and higher education. All variables are measured as proportions in all models. Table 1 shows the characteristics of our sample and of the census as a whole. Notice that the sample closely matches the census, particularly for age, gender and region. In education, OxIS matches the percent of people with secondary school or university qualifications, but it is different for people with no qualifications or further education. For lifestage the percent of students and percent of unemployed matches while the OxIS estimates

fewer employed and more retired people. These differences are mostly due to the difficulty of matching census categories to OxIS categories.

Table 1. Comparison of OxIS with census data
(percents)

Variable	OxIS 2013	Census 2011
Age		
14-22	12.8	13.1
23-37	24.3	23.5
38-51	23.7	24.1
52-59	11.8	11.5
60-67	10.7	11.0
68-75	8.4	7.8
76+	8.3	8.6
Education		
No qualification	24.6	10.9
Secondary degree	35.4	36.1
Further education	14.0	22.5
Higher education	25.7	30.3
Lifestage		
Students	8.4	8.0
Employed	47.8	60.5
Retired	24.7	14.5
Unemployed	18.6	16.8
Gender		
Male	49.3	49.0
Female	50.7	51.0
Region		
Scotland	8.6	8.7
North West	11.4	11.4
South West	8.7	8.7
Wales	5.0	5.0
South East	14.1	14.0
London	12.9	13.1
East of England	9.7	9.5
East Midlands	7.4	7.4
West Midlands	9.0	9.0
Yorkshire & the Humber	8.8	8.6
North East	4.3	4.2

British administrative geography is a complicated structure filled with inconsistent names. Although there are additional administrative units, we will report data on two levels: regions and subregions. At the top level are 11 regions: nine English regions plus Wales and Scotland. Each region has a subregional structure with different names. Scotland is subdivided

into 32 council areas. Wales is divided into 22 unitary authorities. England has a complex structure that subdivides the regions into 32 London boroughs (plus the City of London, which is not a borough), 27 counties, 56 unitary authorities and 36 metropolitan districts.⁷ To avoid this confusing nomenclature, we refer to all of the sub-regional divisions by a single name, “local authority districts” (LADs) OAs are much smaller than LADs and they are contained entirely within LADs. LADs vary widely in area and population; the number of OAs in a LAD ranges from 31 in the City of London to 5,486 in Glasgow, with a total of 227,759 OAs in Britain.

Results

We begin by mapping three key independent variables to give readers a sense of the geographic variation in Britain. Figure 2 shows a map of the percent of elderly people in each LAD. Since we are interested in Internet use, we use the point where use begins its precipitous drop to define elderly as people over age 51.⁸ The darker black lines are regions; LAD boundaries are in grey. The six category scale used for shading emphasizes the extreme values. The lowest and highest categories each have 5 percent of LADs. The categories next to the extreme categories each contain 15 percent of LADs, and the middle categories each contain 30 percent of the LADs. Notice the strong spatial patterns, which seem to be associated with an urban-rural divide. Large cities, especially London, but also Manchester, Birmingham, Cardiff,

⁷ There are actually more English administrative units than we identify here. We omitted parishes, wards, middle- and lower-layer super output areas, non-metropolitan districts, metropolitan counties, and unitary authority districts. These are not relevant for our purposes. This relationship between these units is so complicated that it is frequently represented incorrectly in official British government websites and papers.

⁸ Internet use declines from 91 percent among respondents age 38-51 to 26 percent among respondents over age 75.

Glasgow, Aberdeen, and others stand out because they have relatively few elderly people. Areas like western rural Scotland, central Wales, and Cornwall have a much higher proportion of older people. In general, regional differences seem to be less strong than the urban-rural differences.

Figure 3 shows a map of the percent of people employed in each LAD. Again there are distinct geographic differences between place, but the patterns are different from the geography of age. With some exceptions, like Birmingham, there is no apparent urban-rural difference. London, for example, has boroughs in both the highest and second-lowest categories of percent employed. The absence of an urban/rural divide may be because cities are home to large numbers of young people who are students and not employed. The lowest percentages of employed people are in western Wales, the North East of England, and southwestern Scotland. The highest percent of employed are in the south centre, and northeastern Scotland.

The third map (Figure 4) displays the percentage of people with a higher education degree. We again see a diverse spatial pattern. London has some LADs in the highest category and some in the second lowest. Edinburgh also stands out in the highest category. Some university cities like Aberdeen, Cardiff, and York stand out from their surrounding countryside, but Leicester and Swansea, also home of a university are in the second or third lowest category. The South East tends to have more highly educated people, whilst the East Midlands, southern Wales, and southwestern Scotland tend to have a smaller percent of people with a university degree.

Figure 2: Percent Elderly

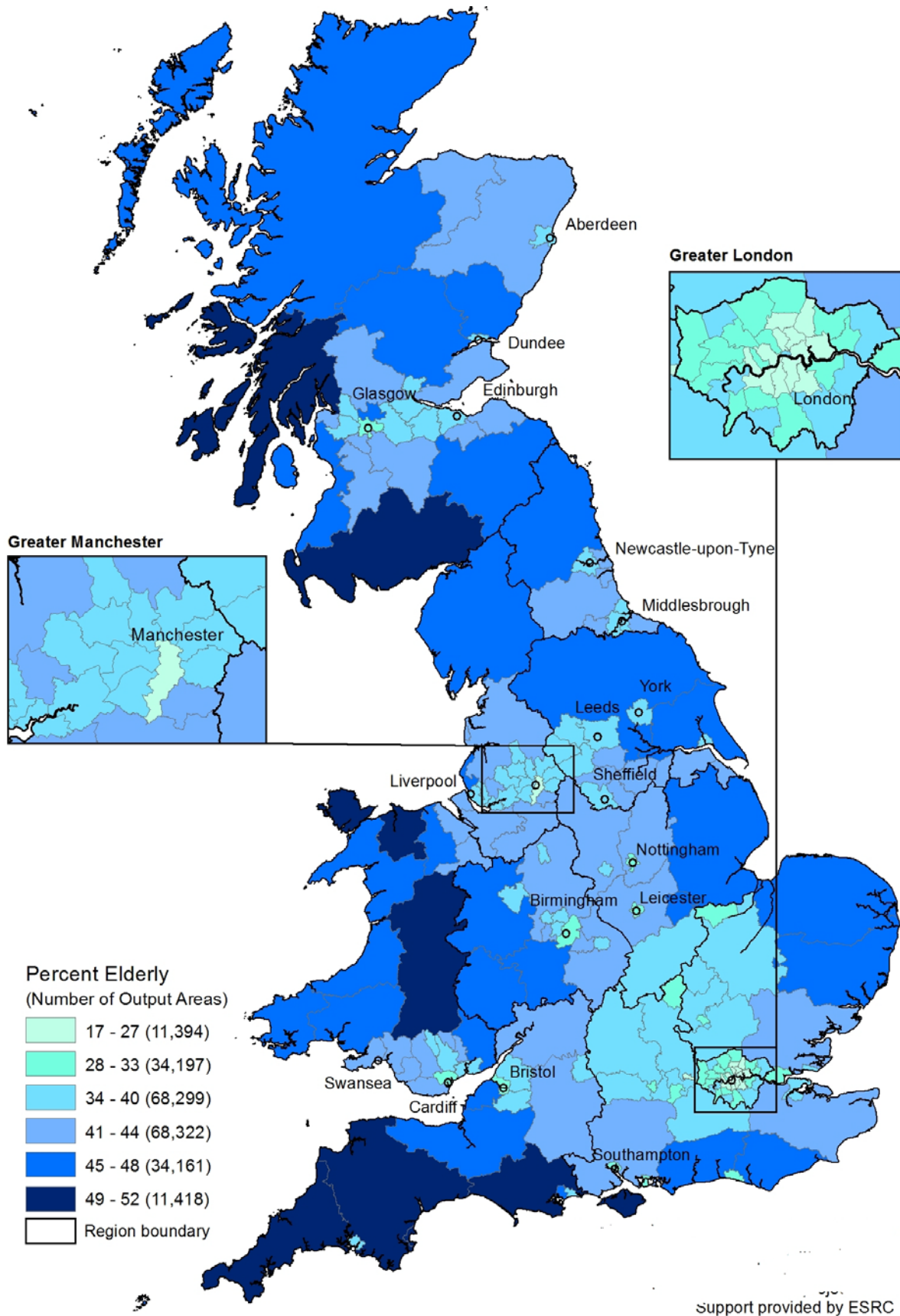


Figure 3: Percent employed

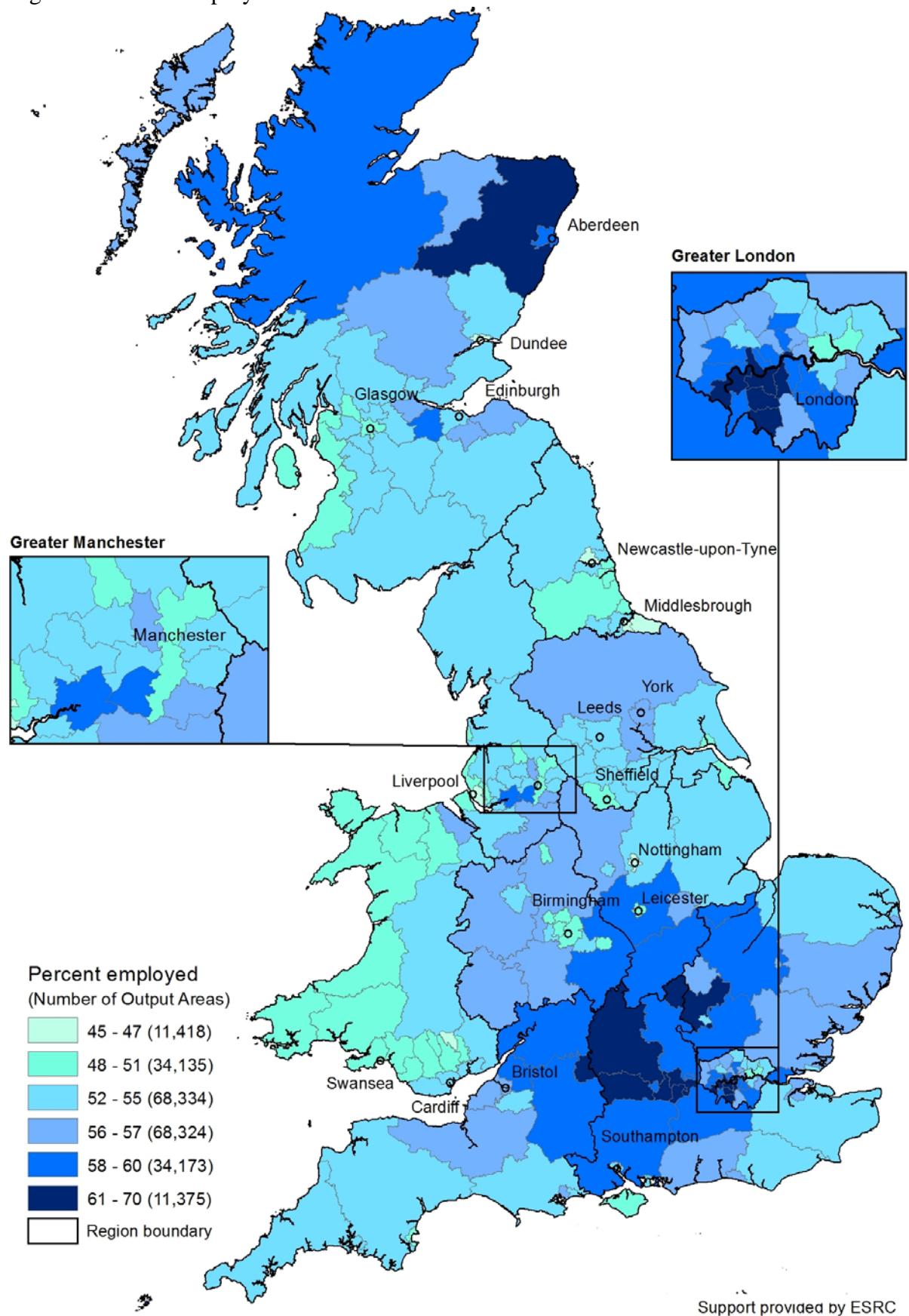
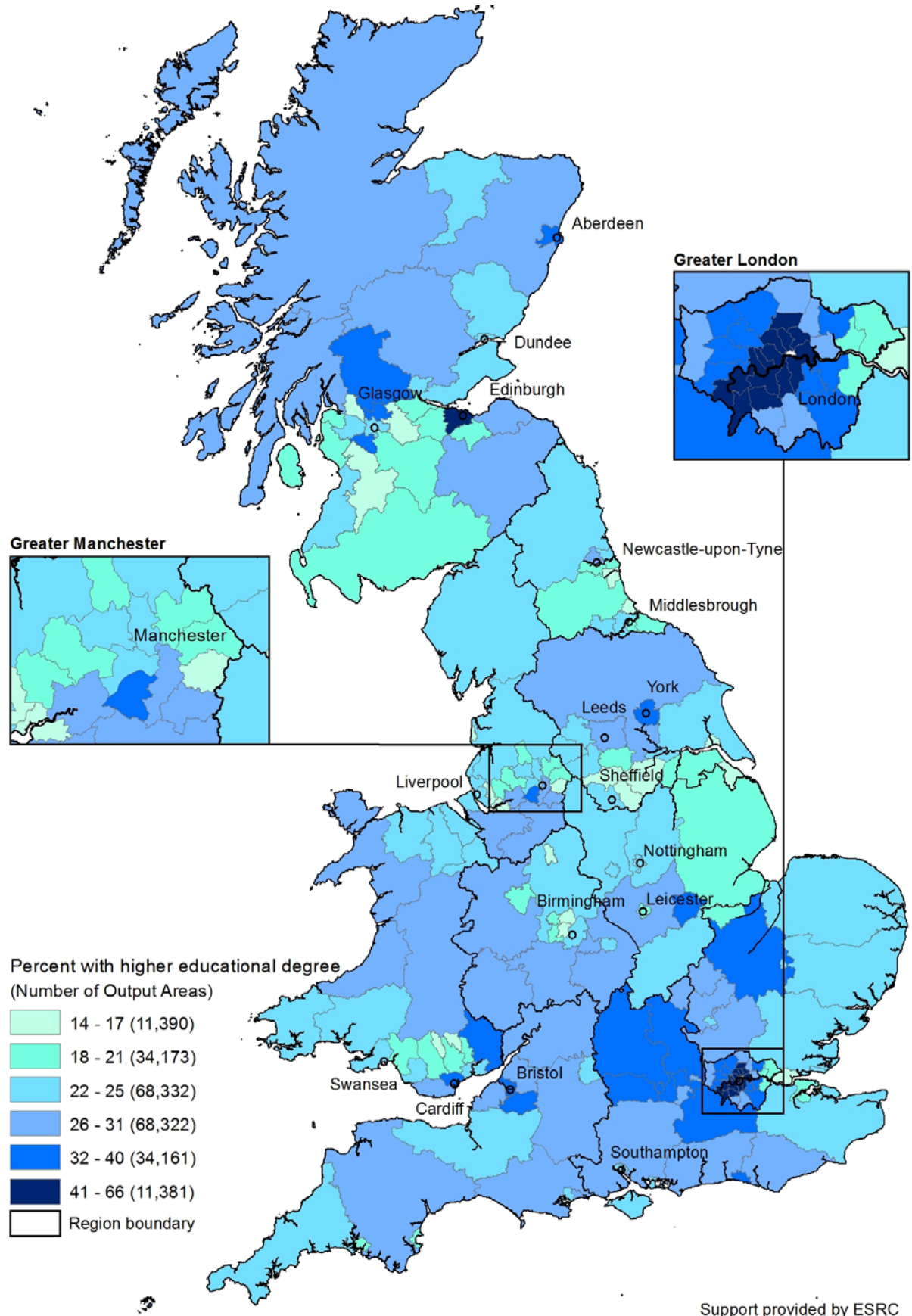


Figure 4. Percent with higher education degree



The first stage of our procedure was the estimation of the coefficients from the OxIS survey (see Table 2). Table 2 shows three mixed models with different independent variables. Model (1) includes age, education and lifestage; model (2) adds gender; and model (3) adds region. The data have been aggregated to the OA level; thus, N = 260 for all models.

Table 2: Mixed Models Estimating the Proportion Internet Users

	Model (1)	Model (2)	Model (3)
Age			
23-37	0.140	0.125	0.140
38-51	0.102	0.089	0.104
52-59	-0.138	-0.144	-0.138
60-67	-0.091	-0.125	-0.094
68-75	-0.031	-0.061	-0.034
75+	-0.365	-0.392	-0.368
Education			
No qualifications	-0.229**	-0.237**	-0.227**
Secondary school	-0.056	-0.053	-0.054
Further education	-0.160**	-0.163**	-0.159**
Lifestage			
Employed	-0.12	-0.108	-0.118
Retired	-0.29	-0.252	-0.286
Unemployed	-0.271*	-0.243*	-0.271*
Gender		-0.068	
Region			
North West			-0.019
South West			0.087
Wales			0.027
South East			0.047
London			-0.011
East of England			-0.023
East Midlands			0.053
West Midlands			-0.004
Yorkshire			0.007
North East			-0.006
Constant	1.071***	1.102***	1.054***
BIC	-5747	-5767	-5706
N	260	260	260

Notes: * $p < .05$; ** $p < .01$; *** $p < .001$.

Omitted categories are age 17-22, higher education degree, students, males and Scotland.

The signs and sizes of coefficients are consistent with previous research on Internet use.

The age coefficients turn negative and are largest in oldest age category; older users are more

likely to be non-users. Since higher education is the omitted category, all the education coefficients are negative and they are largest for least educated people; less educated people are less likely to use the Internet. The lifestage coefficients are all negative because all lifestage groups use the Internet less than omitted category, students.

Age alone accounts for about 37 percent of the reduction in the residual variance, so why are none of the age dummy variables significant and why did we leave age in the model? The reason is collinearity. The largest condition index for the model is 27. Auxiliary regression shows that every age dummy variable is collinear with both education and lifestage dummy variables. Collinearity also influences the education and lifestage variables, which accounts for the fact that they are not highly significant. Collinearity increases the standard errors of the coefficients, which is why they are not significant, but the coefficients are not biased by the collinearity and predictions are not damaged by collinearity. Since our goal is to build a prediction model we have no hypotheses about the size or direction of any coefficients, and we do not interpret the coefficients, so we did not attempt to mitigate the collinearity. We left the age, lifestage, and education dummy variables in the model. Gender in model (2) is also not statistically significant, and it increases BIC. Auxiliary regressions show that it is not collinear with anything; it is just not significantly related to Internet use. We therefore elected to not include it in any further analysis. None of the region dummy variables are statistically significant in model (3). Again, collinearity is not the issue. The largest condition index for model (3) is virtually identical to the largest condition index for model (1). Auxiliary regressions showed that none of the other independent variables are significantly related to any of the region dummy variables. We elected not to include region in our predictive model. Because model (1) includes all statistically significant variables, we used the coefficients from that model for the SAE.

The results of the SAE are in the map in Figure 5. We present the data at the LAD level. LAD use is a population-weighted average of all the OAs in the LAD. Darker shading on the

map indicates higher Internet use in a LAD. We begin by discussing the areas with above average Internet use, the three darkest shades. The highest Internet use is concentrated in the south east, and London particularly stands out. Estimated Internet use in the darkest areas is 88-89 percent, concentrated in London plus Reading to the west of London and Leicester. The LADs with 84-87 percent use stretch in a 'C'-shaped band anchored in London and sweeping north. Bristol, Southampton, and Nottingham also have high levels of estimated use. The rest of the south, interestingly including Cornwall, has above average usage levels of 78-83 percent. Cornwall may be above average because of its above average level of broadband use (Cornwall 2007) Leeds, York and Manchester are also in this category.

When we look at the below average LADs, the middle category, 75-77 percent Internet use, they form two blocks. One surrounds southern England, covering East of England, along the Welsh border, and in the Midlands surrounding Leeds, York, and Manchester. The other block is in Scotland, in the belt between Glasgow and Edinburgh and up to Dundee, and surrounding Aberdeen. Wales is composed of both of the two lower categories. 59-70 percent and 71-74 percent, with the exception of Cardiff. Scotland has no OAs in the lowest category. The entire North East region is in the lowest 59-70 percent category.

Figure 6 adds estimates for the largest cities to the OA estimates. Much of the story is identical. Cities in southern England have the highest estimated Internet use. The second-highest levels are in Wales, Scotland, and the Midlands. The North Eastern cities, Newcastle and Middlesbrough, have the lowest Internet use at 64 percent and 61 percent of the population respectively. What is notable is how cities tend to stand out with higher Internet usage than their surrounding regions: Wales, Scotland, East Midlands, Nottingham, Leicester, Bristol and London are all examples of this. The key exceptions are Liverpool, Sheffield and Birmingham, possibly reflecting their relatively urbanized surrounding areas.

Figure 5. Estimated percent of Internet users

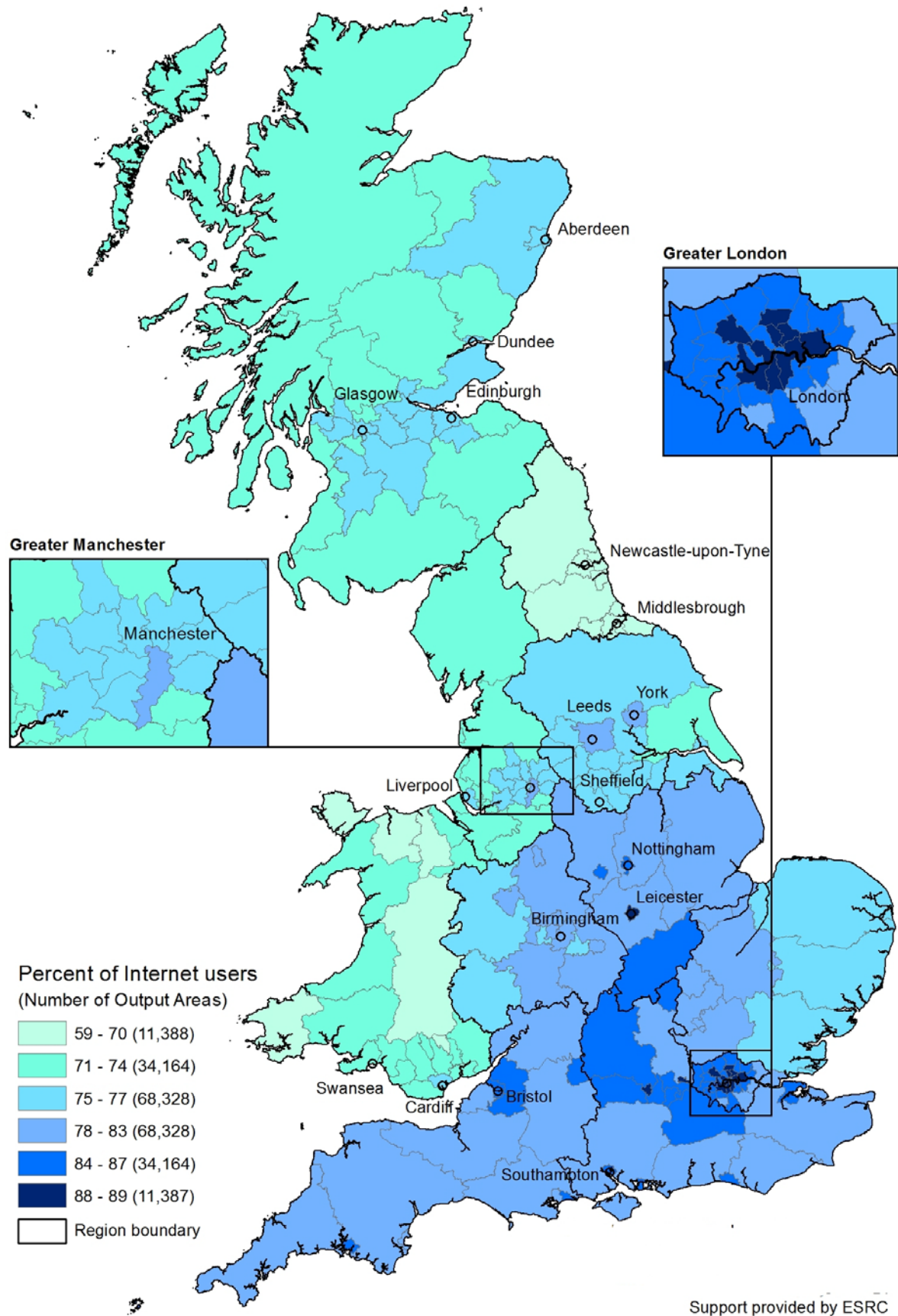
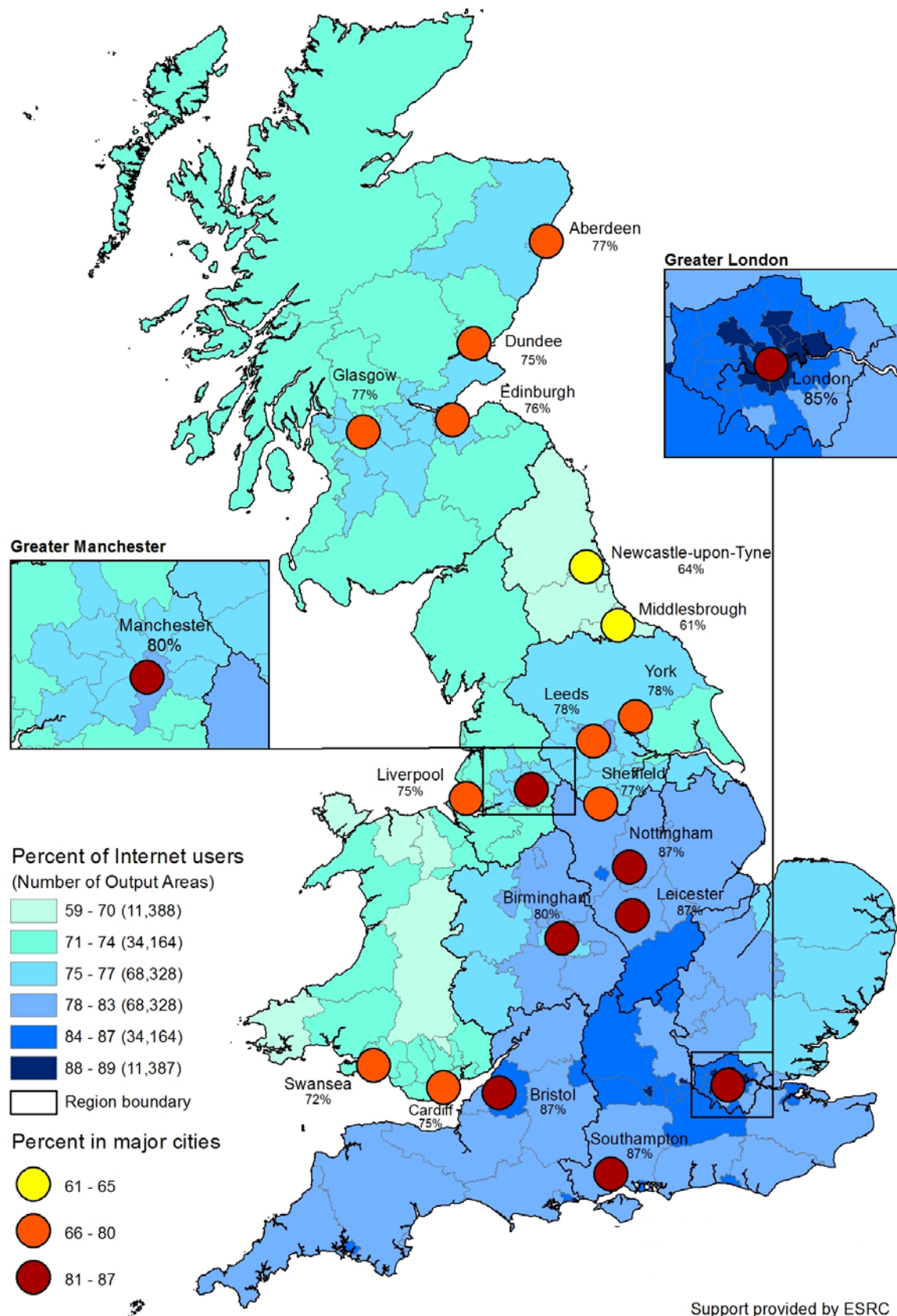


Figure 6. Estimated percent of Internet users with largest cities



Discussion

These results have offered a never-before seen look at the local geographies of Internet usage in Great Britain. However, the results do need to be interpreted with caution. There is no evidence of any issues other than ordinary sampling error in OxIS (Table 1 shows OxIS is consistent with the census) but the estimates are subject to the problems of all SAEs. Because they are estimated based on a few variables the variance of the estimates is small. The results will have an error of exceeding the ± 3 percent—the 95 percent confidence interval for OxIS—and possibly much larger. This suggests that the difference between the top two categories on the map is too small to matter, although it produces intuitively reasonable results for London.

Several other methodological issues deserve brief mention. We produced estimates at the OA level, but this report has aggregated everything to the LAD level. We do this because incorporating a large number of OAs into a single average estimate of a LAD is more likely to produce reliable values. Some of the error in the estimates from smaller OAs should wash out in the aggregated, larger LADs. We entered variables like lifestage, age or region as a single conceptual unit. Because they are a single concept, then when one variable is part of the model the other variables in the concept should also be part of the model. One result is that a number of variables are not statistically significant (e.g. the age variables or secondary school education), but they are conceptually important and so they remain in the model.

Beyond methodology, there are major policy implications from this work. Previously we described the importance of the Internet to core economic, social, and political facets of contemporary Britain, but we know relatively little about who is, and isn't connected.

First, it points to several areas where there are very clearly low levels of Internet use. The North East is striking. Northeastern cities have lower Internet use than Welsh cities, and the difference is not small; it is over 10 percentage points. Furthermore, unlike other regions Northeastern cities do not seem to be very different from the surrounding rural areas. This

suggests that allocating resources to improve Internet use in the North East would be valuable. The North East needs it more than anywhere else. Although not as low as the North East, rural Wales, rural North West and rural Scotland are the next lowest areas. They could be a secondary priority. Cornwall is actually above average, despite the fact that it has been regarded as one of the poorest areas in Britain, which was part of the reason for a major effort to extend broadband, the Superfast Cornwall funded by the European Regional Development Fund, BT and the Cornwall Council (www.superfastcornwall.org)

The most unexpected result is in Table 2. We thought that controlling for demographic variables would weaken the geographic effects but at least some geographic differences would remain. The insignificant coefficients for the region variables in Table 2 suggest that this is not the case. The regional differences that we see in the maps in Figures 5 and 6, seem to reflect the fact that demographic characteristics are unevenly distributed across Great Britain's regions. The apparent regional differences are an epiphenomenon of demographics. This conclusion is consistent with older studies in the US (Horigan & Murray, 2006; Mills & Whitacre, 2003).

Others (e.g. Townsend et al., 2013; Philip, Cottrill & Farrington, 2015; Salemink, Strijker & Bosworth, 2015) argue that the challenges of connecting rural communities to the Internet are due to a combination of factors, including the technological difficulties of deploying broadband away from population centers, the cost of providing digital infrastructure to sparse populations in remote areas and the demographic characteristics of the rural population which influence levels of Internet use. The results in Table 2 suggest that the demographic factors are the most important factor. If it is true that demographics are crucial, then even if digital infrastructure were improved to be equal to urban areas, rural Internet take-up and use would still lag. This evidence directly contradicts the perspective of some studies, e.g. Townsend et al. (2013) or Philip et al. (2015). It is worth repeating, as we noted in the literature review, these studies are based on a correlation and not on multivariate analysis.

The key role played by age and education raises several policy issues. First, differences in education or age are not going to disappear. To the extent that they are important, then programs based on training or technology access may do little to reduce digital inequalities. Second, although the differences in broadband access between rural and urban areas are real (Farrington et al., 2015; Townsend et al., 2013), they may make less difference than policy-makers hope. Various qualitative studies (e.g. Peronard and Just, 2011) provide evidence that broadband would make a difference to many individuals, but it is not clear how these demographic realities might be overcome. Ashmore, Farrington & Skerrat (2015, p. 274) make the reason clear when they say that “the Internet, and specifically the inclusion of superfast broadband, is heavily dependent on personal, individual contexts.” The personal contexts of rural areas include comparatively large numbers of people who are elderly and/or with less education. These demographic categories have comparatively low proportions of Internet users. Broadband access will not change this demographic fact.

Of course, this is an individual-level conclusion. It does not speak to the value of broadband for business. Broadband access may encourage business investment leading to rural economic development and jobs (Galloway and Mochrie, 2005). Furthermore, the British government’s aspirations for a digital society with services delivered ‘digital by default’ is not possible without rural broadband. For both of these reasons, expanding rural broadband remains an important policy.

These differences in Internet use are not just a matter of a technology. The Internet has made an enormous difference in social life, culture, and our economy (Graham and Dutton, 2014) The evidence is that those who have access are able to live very different everyday lives from those who don’t: communicating with friends and family, shopping, political engagement, and job searches are just some of the types of activities for which lack of Internet access can be inherently exclusionary. Internet access provides formal and informal education opportunities,

and it promises to provide better access to health care. Digital inclusion is an individual choice, but the uneven geographies of access point to exclusionary factors preventing some who would like to be connected from ever getting online. Our hope is to contribute to this effort by more closely pinpointing the areas of greatest need.

What's next?

We have used a straightforward measure of Internet use versus non-use as our dependent variable. Similar techniques could predict and map a variety of other variables. For example, we could take a more nuanced view of how people use the Internet. The patterns of mobile use vs fixed-line use may differ geographically, and could be mapped. We could separate work-only users, teenagers using social media, or other subsets. In particular, the 10 major Internet activities (Blank and Groselj, 2014) could be mapped. These include such things as entertainment use, information gathering, commerce and content production. In addition the amount of use and the variety of uses could be mapped. All these are major issues and their geographic distribution has never been tracked.

We are limited by the variables available in the British census, but we have not yet used all variables that are available. A similar technique could be used to examine possible urban-rural differences in more depth, the index of multiple deprivation, occupation, and socio-economic status measured by National Statistics Socio-economic Classification (NS-SEC). However, in the prior work with OxIS these variables have not been strongly predictive, which is why we did not use them in this article. Exploration of interaction effects is also desirable.

We have aggregated the OAs to LADs but this begs the question what is the appropriate spatial unit? City? Except for the largest cities we have not explored intra-urban variation. A future paper could highlight intra-urban variation in Internet use. In a time of widespread commuting where people sleep, which is what census measures, is not the only way to capture

the location of Internet use. Where they work or go to school may also be relevant. One reason why a city like Leicester may be relatively low despite its university is that some of the students who attend the university may not sleep in Leicester. The increasingly dispersed settlement patterns accompanied by long commutes to work may be influencing the results. Of course this explanation is speculative but we hope that others will build on this paper to explore travel to work area and other levels of aggregation.

We have been working with census data but other geo-coded data are available. The relationship between geo-coded data like tweets, Wikipedia edits, social network site use, and social structural variables has never been explored. Without these sort of small area estimation techniques it is impossible to describe the characteristics of users of these social media in terms of gender, age, occupation, politics, marital status and other vars. Having the capacity to do this opens a new area for social media research. More theoretically, ask how we might expand some of the insights from OxIS using traditional (the census) and unconventional (social media data shadows of places) means. Ultimately this work leads us to ask what methods are appropriate to measure local-scale geographies of digital participation.

References

- Anderson, B. (2007). Estimating time spent on-line at small area levels: a spatial microsimulation approach." *Chimera Working Paper number 2007-01*. Ipswich: University of Essex.
- Ashmore, F.H., J. H. Farrington, & S. Skerrat. (2015). Superfast broadband and rural community resilience: Examining the rural need for speed. *Scottish Geographical Journal*, 131, 265-278.
- Blank, G. and Dutton, W. (2012). Age and trust in the Internet: The centrality of experience and attitudes toward technology. *Social Science Computer Review*, 30, 135-151 DOI: 10.1177/0894439310396186
- Blank, G. & D. Groselj. (2014). The dimensions of Internet use: Amount, variety and types. *Information, Communication & Society*, 17, 417-435. DOI: 10.1080/1369118X.2014.889189.
- Bonfadelli, H. (2002). The Internet and knowledge gaps: A theoretical and empirical investigation. *European Journal of Communication*, 17, 65–84. doi: 10.1177/0267323102017001607
- CACI. (2014). *The Acorn User Guide*. <http://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf> (last accessed 20 October 2016).
- Champion A.G., Green, A.E., Owen, D.W., Ellin, D.J., Coombes, M.G., (1987). *Changing places: Britain's demographic, economic and social complexion*. London: Edward Arnold.
- Cole, J.I. (2012). *World Internet Project International Report*. (4th ed.) Los Angeles: USC Annenberg School, Center for the Digital Future.
- (2013). *World Internet Project International Report*. (5th ed.) Los Angeles: USC Annenberg School, Center for the Digital Future.
- Cornwall (2007). *Convergence Programme for Cornwall and the Isles of Scilly: Operational Programme 2007-13*. <http://www.erdfconvergence.org.uk/investments/next-generation-broadband-infrastructure-for-cornwall-and-the-isles-of-scilly/> (last accessed 11 June 2014).
- Crang, M., T. Crosbie, & S. Graham. (2006). Variable geometries of connection: Urban digital divides and the uses of information technology. *Urban Studies*, 43, 2551-2570.
- Dean, D., S. DiGrande, D. Field., V. Lundmark, J. O'Day, J. Pineda, & P. Zwillenberg. (2012). *The connected world: The Internet economy in the G-20*. Boston: Boston Consulting Group.
- DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). Digital inequality: From unequal access to differentiated use. In K. Neckerman (Ed.), *Social inequality* (pp. 355–400). New York, NY: Russell Sage Foundation.
- Dunford, M. (1995). Metropolitan polarization, the north-south divide and socio-spatial inequality in Britain. *European Urban and Regional Studies*, 2, 145-170. DOI: 10.1177/096977649500200204
- Dutton, W. H., & G. Blank. (2013). *Cultures of the Internet: The Internet in Britain*. Oxford: Oxford Internet Institute. <http://oxis.oii.ox.ac.uk/reports> (last accessed 1 February 2016).
- Eurostat. (2014). *Methodological Manual for Statistics on the Information Society*. <https://circabc.europa.eu/sd/a/d8e9a475-9436-474f-a495-acba34465f51/Meth%20Manual%20for%20statistics%20on%20Information%20society%202014.zip> (last accessed 1 February 2016).
- Farrington, K., L. Phillip, C. Cottrill, P. Abbott, G. Blank & W. H. Dutton. (2015). *Two-Speed Britain: Rural Internet Use*. Aberdeen: University of Aberdeen. <http://www.dotrural.ac.uk/two-speed-britain/> (last accessed 10 January 2016).
- Galloway, L. & R. Mochrie. (2005). The use of ICT in rural firms: A policy-orientated literature review. *Info: The Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media*, 7, 33-46.

- Ghosh, M. & J. N. K. Rao K. (1994). Small Area Estimation: An Appraisal. *Statistical Science*, 9, 55-76.
- Goldstein, H. (2003). *Multilevel statistical models*. New York: Halstead Press.
- Government Accounting Office (GAO) (2006). *Broadband deployment is extensive throughout the United States, but it is difficult to assess the extent of deployment gaps in rural areas*. (GAO-06-426). <http://www.gao.gov/assets/260/250064.pdf>. (last accessed 1 February 2016).
- Graham, M & W. H. Dutton. (eds) (2014). *Society and the Internet: How Networks of Information and Communication are Changing our Lives*. Oxford: Oxford University Press.
- Graham, M., M. Zook & A. Boulton. (2013). Augmented Reality in the Urban Environment: contested content and the duplicity of code. *Transactions of the Institute of British Geographers*, 38, 464-479.
- Hargittai, E. (2002) 'Second level digital divide: differences in people's online skills', *First Monday*, 7, no. 4, http://www.firstmonday.org/Issues/issue7_4/hargittai/.
- Hargittai, E. (2008). The digital reproduction of inequality. In D. Grusky (Ed.), *Social stratification* (pp. 936-944). Boulder, CO: Westview Press.
- Hoffman, D. L., & Novak, T. P. (1998). Bridging the racial divide on the Internet. *Science*, 280(5362), 390-391. doi:10.1126/science.280.5362.390
- Horrigan, J., & K. Murray. (2006). Rural broadband Internet use. Pew Internet and American Life Project. http://www.pewinternet.org/pdfs/PIP_Rural_Broadband.pdf (last accessed 4 February 2015).
- Jenkins, T. (2003). Community-based, community pride: Telcos enrich rural way of life. *Rural Telecommunications*, 22, 14.
- Jenkins, H., Purushotma, R., Clinton, K., Weigel, M. & Robison, A. J. (2006) *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*, Chicago, IL: John D. and Catherine T. MacArthur Foundation.
- Kalapesi, C, S. Willersdorf, P. Zwillenberg. (2010). *The connected Kingdom: How the Internet is transforming the U.K. economy*. Boston: Boston Consulting Group.
- Katz, J., & Aspden, P. (1997). Motives, hurdles, and dropouts. *Communications of the ACM*, 40(4), 97-102. doi:10.1145/248448.248464
- LaRose, R., J. L. Gregg, S. Strover, J. Straubhaar, & S. Carpenter. (2007). Closing the rural broadband gap: Promoting adoption of the internet in rural America. *Telecommunications Policy*, 31, 359-373.
- Lichy, J., (2011). Internet user behavior in France and Britain: exploring socio-spatial disparity among adolescents. *International Journal of Consumer Studies*, 35, 470-475.
- Longley, P. A., R. Webber, & C. Li. (2008). The UK Geography of the e-Society: A National Classification, *Environment and Planning A*, 40, 362-382.
- Longley, P. A. & A. D. Singleton. (2009). Linking Social Deprivation and Digital Exclusion in England. *Urban Studies*, 46, 1275-1298.
- Malecki, E.J. & B. Moriset. (2008). *The digital economy: Business organization, production processes and regional developments*. London: Routledge.
- Mills, B. F., & B. E. Whitacre. (2003). Understanding the non-metropolitan-metropolitan digital divide. *Growth and Change*, 34, 219-243.
- Molina, I & J. Rao. (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistics*, 38, 369-385.
- Ofcom. (2011). *Communications infrastructure report: Fixed broadband data*. http://stakeholders.ofcom.org.uk/binaries/research/broadband-research/Fixed_Broadband_June_2011.pdf (last accessed 23 September 2015).

- ONS. (2014). A beginner's guide to UK geography. Office of National Statistics. <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/index.html> (last accessed 19 March 2014).
- Parker, E. B. (2000). Closing the digital divide in rural America. *Telecommunications Policy*, 24, 281–290.
- Peronard, J-P. & F. Just. (2011). User motivation for broadband: A rural Danish study. *Telecommunications Policy*, 35, 691–701.
- Peter, J. & Valkenburg, P. M., (2006) 'Adolescents' internet use: testing the "disappearing digital divide" versus the "emerging digital differentiation" approach', *Poetics*, 34, 293–305.
- Philip, L., C. Cottrill, J. Farrington. (2015). 'Two-speed' Scotland: patterns and implications of the digital divide in contemporary Scotland. *Scottish Geographical Journal*, 131, 148–170.
- Philip, L., C. Cottrill, J. Farrington, F. Williams, F. Ashmore. (2017). The digital divide: Patterns, policy and scenarios for connecting the 'final few' in rural communities across Great Britain. *Journal of Rural Studies*. DOI: <http://dx.doi.org/10.1016/j.jrurstud.2016.12.002>
- Rao, J. N. K. (2003). *Small area estimation*. New York: Wiley & Sons Inc.
- Rogers, E.M. (2003). *Diffusion of innovations*. New York: Free Press.
- Salmink, K., D. Strijker, G. Bosworth. (2015). Rural development in the digital age: A systematic literature review on unequal ICT availability, adoption and use in rural areas. *Journal of Rural Studies*, doi: 10.1016/j.jrurstud.2015.09.001.
- Simpson, L. & M. Tranmer. (2005). Combining sample and census data in small area estimates: iterative proportional fitting with standard software. *The Professional Geographer*, 57, 222–234.
- Townsend, L., A. Sathlaseelan, F. Fairhurst & C. Wallace. (2013). Enhanced broadband access as a solution to the social and economic problems of the rural digital divide. *Local Economy*, 28, 580–595.
- Twigg, L. & G. Moon. (2002). Predicting small area health-related behaviour: a comparison of multilevel synthetic estimation and local survey data. *Social Science & Medicine*, 54, 931–937.
- Van Deursen, A. J. A. M., & J. A. G. M. van Dijk. (2013). The digital divide shifts to differences in usage. *New Media & Society*, 16, 507–526. doi:10.1177/1461444813487959
- van Dijk, J. A. G. M. (2006) 'Digital divide research, achievements and shortcomings', *Poetics*, 34, 221–235.
- White, N., A. Dent, P. Clarke, D. Silva, & J. Naylor. (2011). Model-based estimates of income for Middle Super Output Areas, 2007/08 technical report. http://www.neighbourhood.statistics.gov.uk/HTMLDocs/images/TechnicalReport2007_08final_tcm97-112199.pdf (last accessed 19 March 2015).
- White, P., & N. Selwyn. (2013). Moving On-Line ? An Analysis of Patterns of Adult Internet Use in the UK, 2002–2010. *Information, Communication & Society*, 16, 1–27. doi:10.1080/1369118X.2011.611816
- Zhao, S. (2009) Teen adoption of MySpace and IM: inner-city versus suburban difference. *Cyberpsychology & Behavior*, 12, 55–58.
- Zillien, N. & Hargittai, E. (2009) 'Digital distinction: status-specific types of Internet use', *Social Science Quarterly*, 90(2), 274–291.

Author bios

Grant Blank (PhD, University of Chicago) is the survey research fellow at the Oxford Internet Institute, University of Oxford, United Kingdom. He is a sociologist specializing in the political and social impact of computers and the Internet, the digital divide, statistical and qualitative methods, and cultural sociology. He is currently working on a project asking how cultural hierarchies are understood in online reviews of cultural attractions. His other project studies the effects of search engines on politics. He can be reached at grant.blank@oii.ox.ac.uk; visit also <http://www.oii.ox.ac.uk/people/blank/>.

Mark Graham is the Professor of Internet Geography at the Oxford Internet Institute in the University of Oxford. He is also a Faculty Fellow at the Alan Turing Institute, a Visiting Fellow at the London School of Economics and Political Science, an Associate in the University of Oxford's School of Geography and the Environment, and a Senior Research Fellow at Green Templeton College. His current research focuses on digital labour, ICTs and development, digital geographies, and inequalities in our digital world.

Claudio Calvino is Senior Data Scientist at Capgemini UK. After finishing his Ph.D in Economic Geography at the University of Rome La Sapienza, he spent a year at the Oxford Internet Institute, University of Oxford analysing the geographic distribution of Internet activity. He then worked for Kantar Media as Lead Data Scientist before his current position.