# Using Census data to derive a new UK-wide area-based measure of deprivation

**Authors: Tej Nathwani, Siobhan Donnelly and Archie Bye**

**Acknowledgements:** The authors would like to thank anonymous reviewers for their helpful comments on earlier drafts of this paper. Any remaining errors are our own.

**Abstract:** Tackling regional inequality across the UK requires data that can identify where the most deprived areas can be found. Existing area-based measures of deprivation available for policy purposes, however, have their limitations. For example, the most common measure utilised in each of the four nations today – the index formed from the Indices of Deprivation – does not adequately capture deprivation in rural locations. This paper therefore uses Census 2011 output/small area data to develop a new UK-wide area based measure of deprivation [called the Socioeconomic Index for Small Areas (SEISA)] using information on the qualifications and occupations of residents. The theoretical rationale behind the choice of these two Census variables is discussed, alongside the methodology behind the derivation of our new measure. We go on to illustrate its correlation with low income and other factors associated with deprivation, before a comparison to the Indices of Deprivation is presented in each of the four nations to illustrate the added value our measure can bring.

## Section 1: Introduction

The fundamental purpose of this paper is to present a new UK-wide area-based measure of deprivation derived using Census 2011 called the Socioeconomic Index for Small Areas (SEISA). Across all nations of the UK, reducing regional inequality continues to be a key policy objective. To meet this goal, resource needs to be allocated to those parts of the country where they are most required. However, this opens up the question of how one defines and identifies the most deprived localities. We therefore begin with a discussion of how previous literature has interpreted the term deprivation, before looking briefly at what measures currently exist, including their limitations.

Townsend (1987) specified this concept as one in which an individual or household has clear disadvantage relative to other members of their local community or wider society. A similar relative construct was put forward by Gordon (1995) where those in deprived conditions were seen as having a level of resource that prevented them from fully participating in well-established activities and norms in public life. More recently in policy circles, the Scottish Government (2018) have set up the Fairer Scotland Duty, which requires public bodies to reflect on how they can reduce the extent of inequalities that result from socioeconomic disadvantage. As in the academic literature, a relative definition has been put forward, where those in deprived circumstances are seen to be living in inferior social and economic conditions compared with others in the same society.

With some general alignment over what is meant by deprivation, the next issue for consideration is how it can be measured. In the UK, material deprivation (i.e. whether an individual/household possesses certain material things) is currently captured through the Households Below Average Income dataset, which is itself derived using the Family Resources Survey. However, this only covers a sample of households across the UK, meaning it cannot be used to understand the levels of deprivation across all areas of the country. Additionally, exploring matters such as social deprivation is also not feasible, due to an absence of UK-wide information on the affordability of social activities. However, as an alternative, low income could be used as a way of identifying those experiencing deprivation. As Townsend (1987) noted, those facing one or more forms of deprivation have a very high probability of having little financial resource. However, income data too is not typically available in the public domain at either individual or small area level. This has left researchers having to use national collections such as the Census to formulate indices that are likely to be highly associated with deprivation.

Townsend (1987) highlighted the potential for four variables (available in the Census) to be utilised in the formation of an index relating to material deprivation. These were unemployment, overcrowding, as well as

car and home ownership. Yousaf and Bonsall (2017) have constructed the Townsend Deprivation Index using the 2011 Census at various geographic levels, including output area. The index designed by Carstairs and Morris (1989) also utilised similar Census variables, though social class was included as opposed to a variable on home ownership. Both of these measures though have been subject to criticism. For example, the inclusion of a variable on vehicle ownership when constructing an index has been questioned by Allik *et al.* (2016) on grounds that purchasing a vehicle may be necessary in rural areas with limited public transport, even among poorer households. The reduced validity of indicators of socioeconomic deprivation that are based partly on vehicle ownership among rural localities was also illustrated by Christie and Fone (2003).

These are not the only two area-based measures of deprivation to have been developed. The most common index drawn upon by government departments in the four nations today in their decision-making processes is based on the Indices of Deprivation, with the work of Townsend (1987) providing the conceptual framework for the generation of these indicators. Originally created to support policy design - such as the allocation of resources for the Labour government's neighbourhood renewal agenda in the early 2000s - the Indices of Deprivation initially covered domains such as education, employment and income. These were then weighted to generate a final index. Since then, they have continued to evolve every few years.

Currently, each UK nation creates their own final index and the different methodological approaches employed mean that this is not a UK-wide deprivation metric. For example, in Wales, eight domains are drawn upon to generate the composite measure, whereas Northern Ireland, Scotland and England utilise seven indicators to form their index. Weights assigned to similar domains also vary across countries. In Wales, community safety contributed 5% to the 2019 index, while in England, crime was given a weight of 9.3% when creating the final index. In their review of statistics in the post-16 education and skill sector within England, the Office for Statistics Regulation (OSR, 2019) noted the lack of a UK-wide deprivation metric was inhibiting the potential to publish country-level analysis, as well as assessing the progress being made on improving social mobility within society.

With regards to the size of the areas utilised in the Indices of Deprivation, lower layer super output areas (LSOAs) are presently used in England and Wales, which average approximately 1,500 inhabitants. Data zones are the geography level employed in Scotland, which have populations of 500 to 1,000, while in Northern Ireland, super output areas (SOAs) will generally consist of around 2,000 people. Areas are partitioned into deciles or quintiles, through which those living in the most deprived parts of a country are identified. It should be noted though that another limitation often raised about the Indices of Deprivation is

that the size of the areas used can lead to it being difficult to identify pockets of deprivation within less deprived areas. Additionally, a common critique of the final index developed in different countries is that they fail to adequately capture deprivation in more rural aspects of a nation, where it may not be as geographically concentrated (as is often the case in urban areas). Indeed, the Commission on Widening Access (2016) report acknowledged this weakness, as did the Welsh Government (2015) paper on evaluating deprivation in rural areas.

Consequently, there remains an opportunity to create a UK-wide indicator that can suitably proxy for deprivation and that addresses some of the drawbacks of existing measures. We therefore utilise the 2011 Census to create SEISA - a new index at output area level based on the educational qualifications and occupations of residents. Output areas are the base unit for Census data and therefore represent the smallest geographical level at which aggregated statistics are released into the public domain. They are more commonly referred to as 'small areas' in Northern Ireland. They generally contain less than 500 individuals, hence using areas of this size is designed to mitigate the limitation of the Indices of Deprivation of not being able to pick up pockets of deprivation in otherwise less deprived neighbourhoods. The methodology we employ to develop our measure of deprivation shares some similarities to the approach adopted by the Australian Bureau of Statistics (2016), who draw upon their Census results to create the Index of Education and Occupation (IEO). The geographical unit they utilise to generate this index is the Statistical Area Level 1, which typically contains between 200 and 800 people. This composite measure is frequently used in Australia for decisions relating to funding, as well as for research/statistical purposes. The relation our measure has to low income and other features associated with deprivation, such as poor health, is also illustrated. Furthermore, we undertake a detailed comparison of how the variable we have generated differs to the Indices of Deprivation, including its ability to capture more rural localities.

The rest of this paper proceeds as follows. Section 2 outlines the data sources we use, while section 3 discusses our methodology and presents some summary statistics on our measure. We then go on to look at whether our measure is likely to be correlated with factors associated with deprivation and also conduct a sensitivity analysis to examine whether it could be improved by incorporating housing tenure. This is followed by an assessment of the variable, alongside the Indices of Deprivation. The study closes with concluding remarks.

## Section 2: Data

To derive the new area-based measure, as well as carry out the subsequent analysis, it was necessary to access and link a variety of data sources.

The first of these was 2011 Census data that was available in the public domain. The Census is a UK-wide collection that occurs every ten years and is mandatory for all households to complete. A range of topics are covered as part of the questionnaire, including employment, education, as well as home and vehicle ownership. It is administered by the Office for National Statistics (ONS) in England and Wales, while the Northern Ireland Statistics and Research Agency (NISRA) and National Records of Scotland (NRS) gather the relevant data for Northern Ireland and Scotland, respectively. Alongside there being a very high level of coverage across the population, the UK Data Service (2022) Census forms illustrate that there is general consistency in the way questions are asked across all four nations. Indeed, a report by the ONS (2015) indicates that many of the published outputs on different aspects of the Census are either broadly or highly comparable.

As specified earlier, the smallest geographic domain at which data is subsequently released to the public is at output area level (or small areas in Northern Ireland). In England and Wales, ONS (2021a) highlight the aspiration was for output areas to contain approximately 125 households, while also being as homogenous as possible (based on tenure and dwelling type). In Scotland, no such requirement was set on homogeneity, with NRS (2015) indicating that output areas are expected to contain between 20 and 78 households. NISRA (2019) point out that small areas in Northern Ireland average around 160 households/400 individuals and are intended to be socially similar. Our starting point was to therefore obtain key statistics at output area level from the 2011 Census supplied by ONS (2021b), NRS (2021b) and NISRA (2021) relating to various indicators, which included;

a) Age structure (KS102)
b) Health and provision of unpaid care (KS301)
c) Household tenure (KS402)
d) Household composition (KS105)
e) Qualifications and students (KS501)
f) National Statistics Socioeconomic Classification (NSSEC) (KS611)

As stated in the introduction, qualifications (KS501) and occupation (KS611) were the two variables we used to create our index. Data on age, housing tenure, household structure and self-reported health were

utilised to enable us to develop some summary statistics on our index and the extent to which it may be correlated with low income/(material) deprivation. For example, The Health Foundation (2020) demonstrate that poorer self-reported health is correlated with lower household income, while HM Government (2014) note that lone parent households have a higher risk of experiencing long-term poverty. Meanwhile, a report by Welsh Government (2023) highlights that living in social housing, having poor health or being a single parent are all associated with a greater probability of facing material deprivation.

This was followed by ingesting look-up files for each of the home nations. ONS (2018a) locates output areas to an English region, while ONS (2018b) indicates how output areas in England and Wales map to LSOAs, middle layer super output areas (MSOAs) and local authority districts. NRS (2011) supply the output area to data/intermediate zone look-up file for Scotland. It should be noted though that 2011 output areas in Scotland only match perfectly into council areas, with best fit aggregations having to be applied at other levels of geography. For Northern Ireland, NISRA (2013a) disseminate information that highlights how small areas map to larger geographical domains, such as wards and local government districts (LGDs). However, small areas do not nest properly into LGDs, so assignment is commonly determined by the location of the majority of households. As this data source provides the 1992 LGDs/wards, we also utilise an additional file supplied by NISRA (2013b) to obtain the updated 2014 LGDs. Within this, there is supplementary detail on how 2014 District Electoral Areas (DEAs) match up to the 2014 LGDs. The rationale behind linking these look-up files to our Census data was that we would then have the codes needed to bring in the Indices of Deprivation and/or urban-rural classifications, which are formed at a higher level of geography.

One of the drawbacks of the Indices of Deprivation is that they are less useful in capturing deprivation in rural areas. Consequently, we wanted to ensure that our dataset contained information on the urban-rural classification for each nation to assess the extent to which SEISA addresses this limitation. In England and Wales, a grouping has been developed by the Department for Environment, Food and Rural Affairs (DEFRA, 2021) and we examine their detailed 10-fold categorisation when conducting our analysis for these two countries. For Scotland, the Scottish Government (2019a) have developed a file that outlines how data zones map to higher level geographies, with this also containing an urban-rural classification at varying levels of granularity.

The Indices of Deprivation (both the composite measure and individual domains) were then ingested for all four countries. The Ministry of Housing, Communities and Local Government (2019) was the government department responsible for publishing the most recent version for England. As they also release supplementary data relating to the income deprivation affecting children index (IDACI) and given the

relevance of this variable to our work, this was also brought into our dataset. 2019 was also the year that the latest version of the Welsh Index of Multiple Deprivation (WIMD) was disseminated by Welsh Government (2019), while the Scottish Government (2020) circulated their updated Scottish Index of Multiple Deprivation (SIMD) data a year later. In Northern Ireland, NISRA (2017) distributed the Northern Ireland Multiple Deprivation Measure (NIMDM), which also included some extra data on the urban-rural nature of an area, as well as an Income Deprivation Affecting Children (IDAC) indicator.

Additionally, as we stated in the introduction to our paper, interest lies in understanding how indices developed from the Census correlate with income. Concerns around sensitivity and the potential impact on non-response have precluded questions on income from emerging in the Census. However, one file we can ingest to enhance our knowledge of how a derived index may be associated with income is the ONS (2020) small area income estimates for England and Wales, with these figures having been derived using a model-based approach based on a dataset comprising of both survey and administrative sources. Data at MSOA level is available for the financial year 2011/12 and contains four different income measures – total and net weekly household income, as well as equivalised net weekly household income (before and after taking into account housing costs). However, similar data is not available for ingestion into our master dataset in either Scotland or Northern Ireland.

## Section 3: Deriving a new UK-wide area-based measure based on Census 2011

To form our composite measure, we utilise data at the output area level on the qualifications and occupations of residents.

Earlier in this paper, we noted why ownership of a vehicle was not suitable to include in developing a measure for deprivation. While Townsend (1987) incorporates housing quality into the construction of his index, a synthesis of the evidence around the drivers of poverty by HM Government (2014) illustrates that it is low income that is more likely to result in a family living in poor quality housing, rather than the other way around. The same study, however, highlights the important role that education (and subsequently occupation) plays in deprivation. ONS (2018c) illustrates the relationship between education and earnings, with those possessing a degree or equivalent having the highest gross hourly median pay. Meanwhile, Table 29 in ONS (2022) demonstrates how those in professional occupations tend to have the largest earnings. As noted in HM Government (2014), the transmission mechanism linking education, occupation and income is therefore as follows. Adults with no/few recognised qualifications are more likely to experience spells of unemployment/underemployment and have a higher likelihood of finding low-paid insecure work (e.g. routine occupations like being a waiter or member of bar staff), thus increasing their chances of experiencing one or more forms of deprivation. In contrast, those with greater levels of education often obtain employment in high wage work (e.g. professional jobs such as accountancy and law) – a pattern illustrated by ONS (2001).

Across the UK, there are 232,296 output areas. For each of these, we therefore generate the following two variables;

- Proportion of residents in the output area aged 16 and over with below level 4 qualifications
- Proportion of residents in the output area aged 16 to 74 in NSSEC groups 3 to 8 (those that couldn't be classified were excluded from the calculation)

Table 1 presents some initial summary statistics relating to the population sizes of the two variables we draw upon to create SEISA. We see from this that the number of residents aged 16 to 74 in NSSEC groups 1-8 averages (median) 193 across the output areas of the UK. The median value is slightly higher at 233 when considering the education levels of those aged 16 and over (from which the proportion of individuals with below level 4 qualifications is calculated). Though there are a few particularly small/large

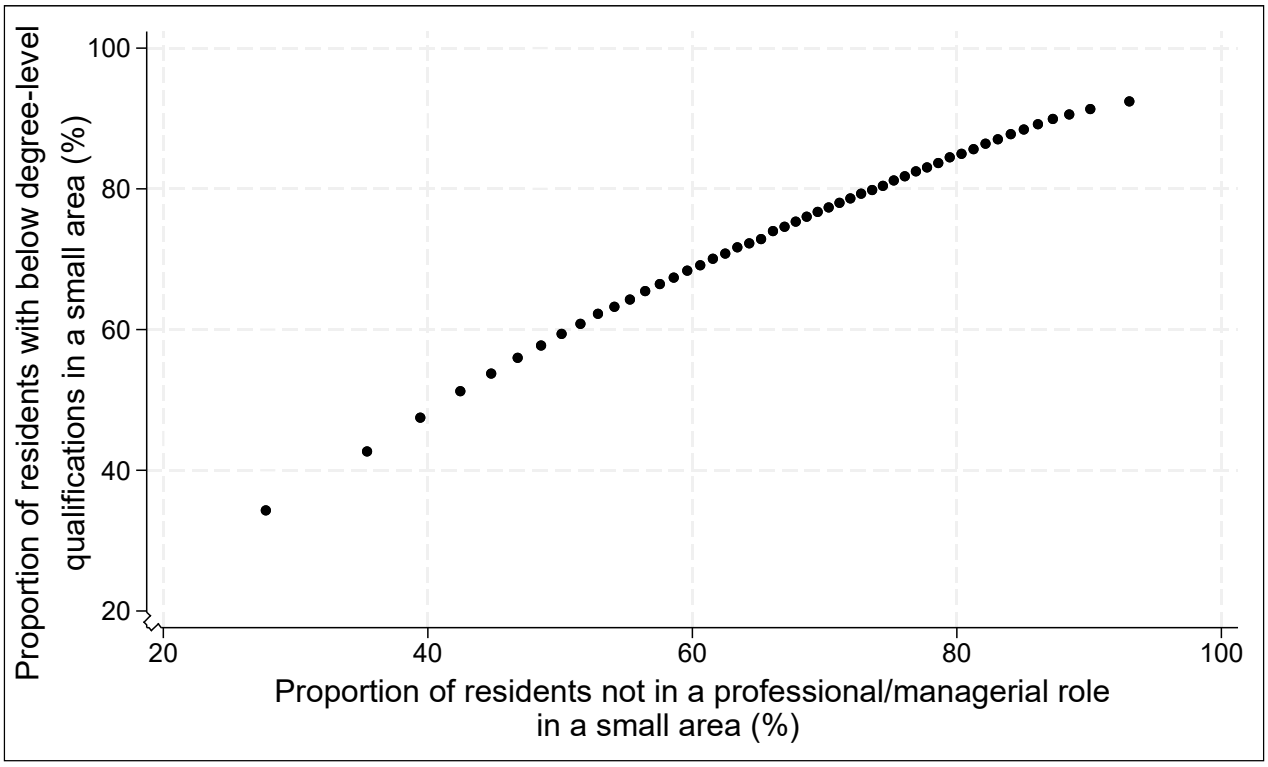
populations at either end of the scale, the proportions that make up our measure are generally based on totals ranging from 40 to 450. Smaller population sizes are more common in Scotland where the number of households included in an output area can be as low as 20. In Australia, two of the criteria adopted by the Australian Bureau of Statistics when deciding whether an area will be given a score is if the population level is under 11 or if the denominator of a variable in the index is less than 6. In such areas, confidentiality and data quality concerns would lead to the value being suppressed. Only 1 output area in our file would fail to meet such criteria. Indeed, only in 0.1% of cases do we encounter an output area where the occupation proportion is based on a total of less than 30. We therefore do not impose any form of suppression on our data in this paper.

**Table 1: Summary statistics on the population sizes for the two variables used to generate SEISA**

| | Occupation | Qualifications |
|---|---|---|
| Minimum | 5 | 33 |
| 1st percentile | 41 | 52 |
| 50th percentile | 193 | 233 |
| 99th percentile | 355 | 436 |
| Maximum | 1,895 | 4,087 |

We find that in the median output area, the proportion of residents with below level 4 qualifications stands at 76%, while the corresponding figure for the percentage in NSSEC groups 3 to 8 is 68%. The standard deviation for both fields is very similar (around 15). Indeed, when exploring the correlation between the two, we found there to be a highly linear and positive relationship, with the Pearson's correlation coefficient being 0.91. Figure 1 demonstrates this graphically. Earlier, we discussed the transmission mechanism between education and occupation. Reassuringly, the expected pattern emerges in the data, with areas in which residents have higher levels of qualification more likely to be working in professional jobs and hence being classified in NSSEC groups 1 or 2.

**Figure 1: The correlation between the qualification and occupation indicators in Census small areas used to create SEISA**

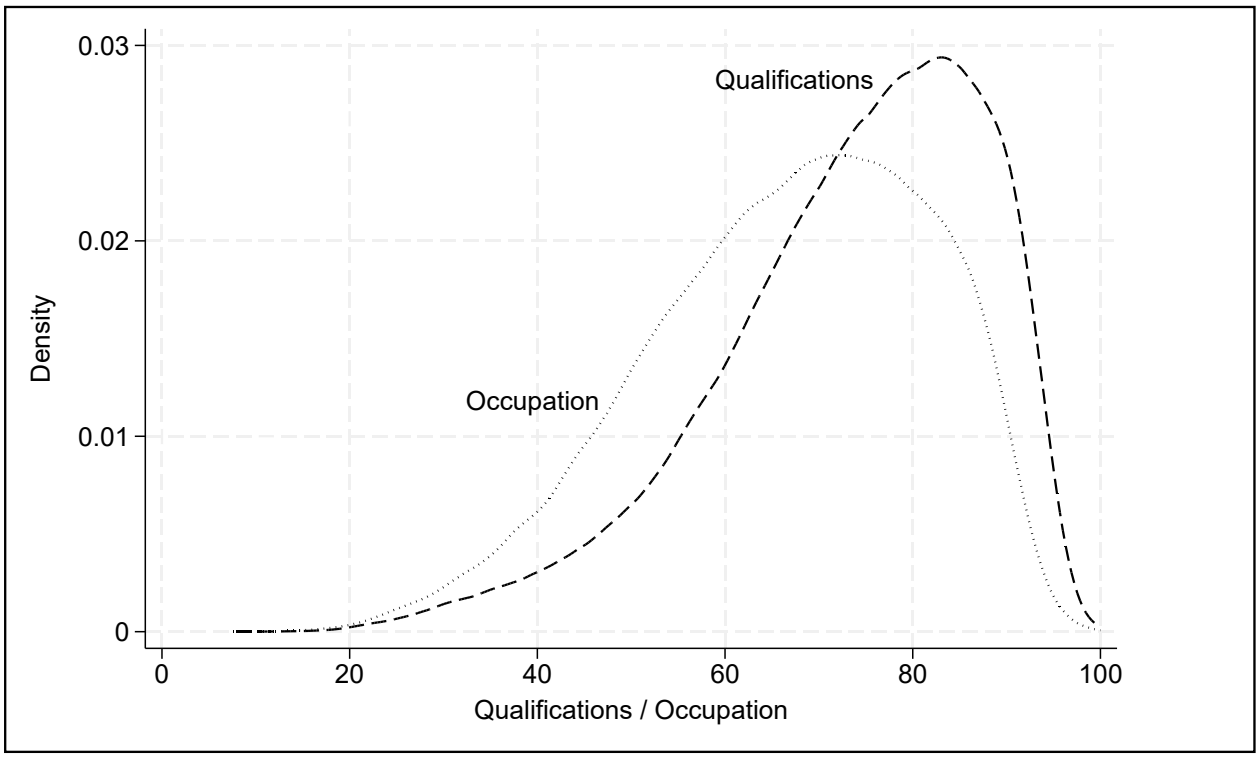



In the Townsend index, each of the four variables are standardised before being added together (they all carry equal weight). The unemployment and overcrowding indicators are log transformed prior to standardising, due to the skewed nature of these fields. Standardising a variable is often carried out when we have several variables that are not all measured on the same scale. That is not the case here though. Both the variables we are looking to work with are percentages that can range from 0 to 100.

Furthermore, we have two variables that display a strong linear association (it is not too distant from being the case that the qualification and occupation variables have equal values across output areas, which would be represented by a correlation of 1). Both exhibit quite similar distributions (see Figure 2 below) and yield comparable summary statistics. Here therefore, the simplest method by which a composite measure can be created is either through taking an average of the two values for each output area or proceeding with using one of the indicators only. We choose the former approach. Utilising a technique such as principal component analysis would not produce any material difference in the final composite measure and would also bring the disadvantage of being less easy to interpret for the end user of the

statistics. Indeed, we find the correlation between our preferred approach to creating the index and an alternative method such as principal component analysis to be 1.

**Figure 2: The distribution of the qualification and occupation indicators in the Census used to create SEISA**




For each of our 232,296 output areas therefore, we assign a value equal to the average of the two variables we originally started with on education and occupation. Hence, the median output area would have a figure of 72% for this new composite measure we have created. All output areas are then ranked based on this numeric value, with those areas displaying the highest percentages considered to be the most deprived, as they have the largest proportions of residents without level 4 qualifications/working in occupations that fall into NSSEC groups 3 to 8.

The appendix accompanying this paper provides supplementary material on the methodology, including further detail on why this measure can be utilised for a UK-wide analysis, as well as how it can be used for comparing statistics across nations. Additionally, we highlight how similar results would emerge even if we use only one of the two variables to create the measure, as opposed to both.

## Section 4: How does SEISA correlate with income and other factors associated with deprivation?

Although we do not hold individual or small area level data on income, this does not completely prevent us from examining the extent to which our measure may be correlated with it. As we do have household income estimates at the MSOA domain for England and Wales, we can see whether there is any clear pattern between our measure and net equivalised weekly household income after housing costs. We utilise this measure of income, as the level of resource available to a family to participate in society (and hence the extent of deprivation they are facing) can only be truly determined once fixed costs, such as mortgage or rent payments, are paid. The correlation between SEISA (applied as a continuous measure) and income estimates in England was found to be -0.70 (i.e. more deprived areas are associated with lower income levels), with the equivalent figure in Wales being -0.58. Figures 3 and 4 provide a visualisation of the relationship for both nations, where SEISA is categorised into country-specific deciles, so that all deciles within the nation contain the same proportion of output areas (i.e. 10%). As the appendix demonstrates, at a UK-wide level, the distribution of output/small areas in any given nation would not be equal across deciles/quintiles.

**Figure 3: The relationship between SEISA in England and net equivalised weekly household income after housing costs (at MSOA level)**

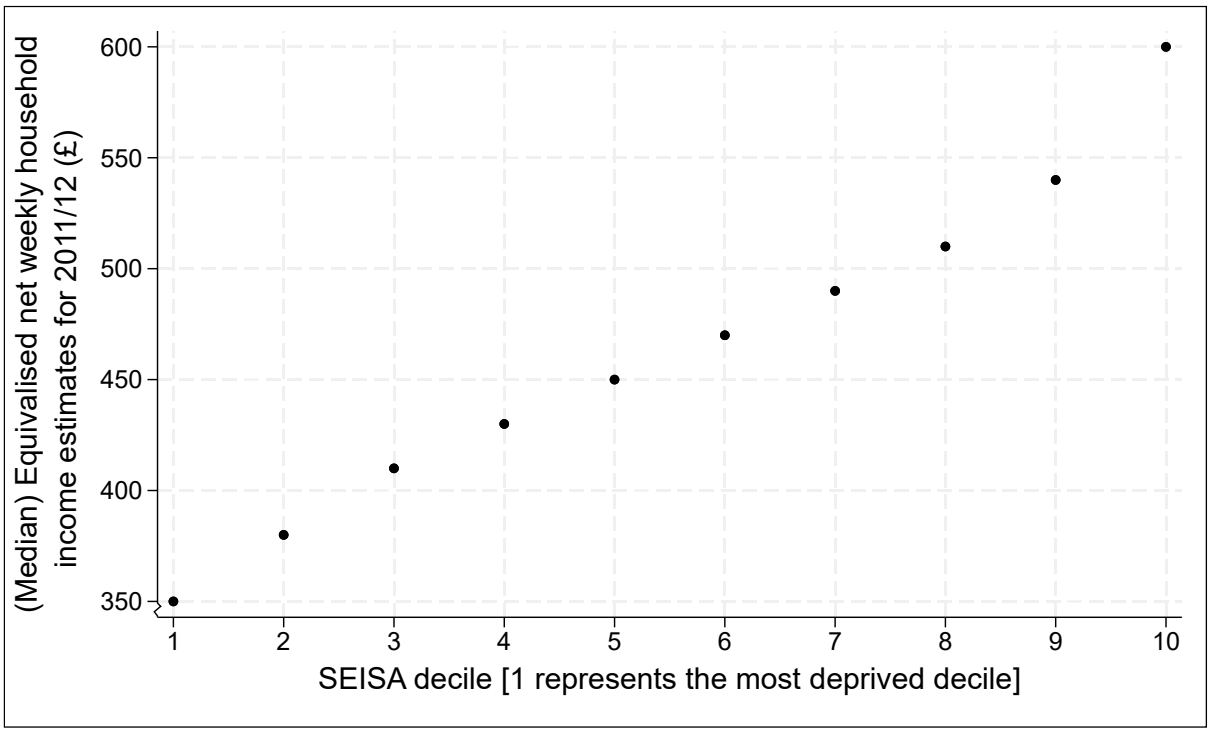
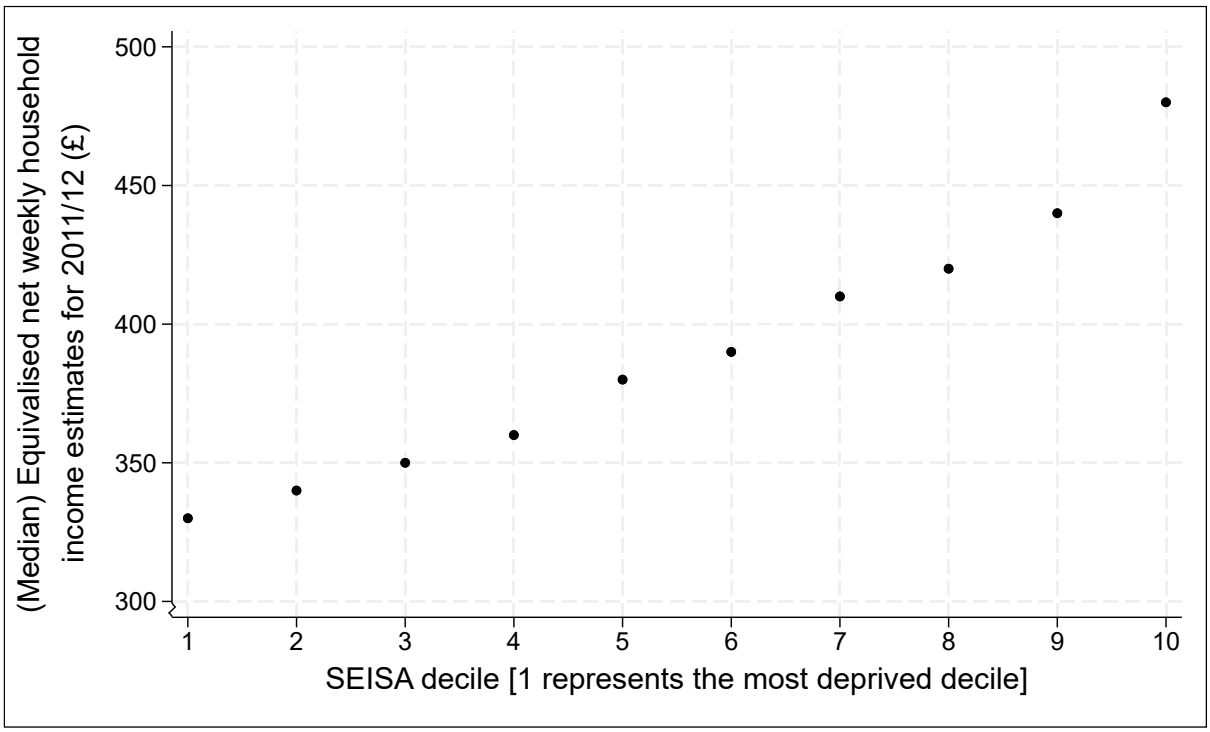
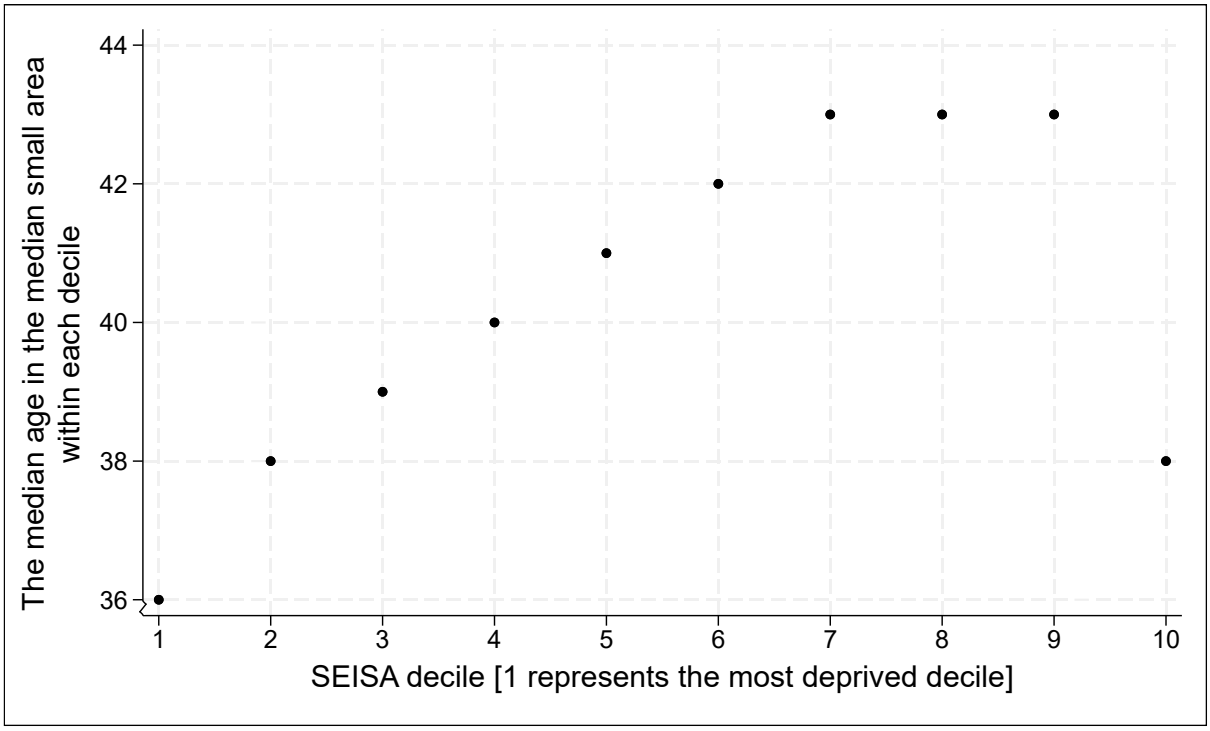


**Figure 4: The relationship between SEISA in Wales and net equivalised weekly household income after housing costs (at MSOA level)**




While we do not have similar income data for Scotland and Northern Ireland, we can utilise some of the other Census variables in our dataset to get a better understanding of whether our measure is likely to be associated with low income and/or material deprivation across the whole of the UK. We begin by considering the link between our measure of deprivation and the median age. Francis-Devine (2021) illustrates the quadratic relationship that exists between income and age in the UK, with individuals experiencing their highest earnings in their forties. The Census provides us with information on the median age of residents in each output area. Figure 5 explores the relationship between our deprivation index and the age of residents, with the output areas falling into the most deprived deciles tending to display a younger age profile. Based on research evidence on how age and income relate, we would therefore expect deprived localities according to our measure to have residents with lower levels of income.

**Figure 5: The relationship between SEISA and age across the UK**




Furthermore, we utilise information on household composition to develop a variable that identifies the proportion of lone parents (with either dependent or non-dependent children) in each output area. As stated earlier, single parent households are more likely to experience poverty and material deprivation, with reasons for this potentially including lone parents having to leave the labour market following separation. Figure 6 demonstrates that the output areas in the most deprived deciles of our measure have a higher average proportion of lone parent families.

**Figure 6: The relationship between SEISA and household composition across the UK**

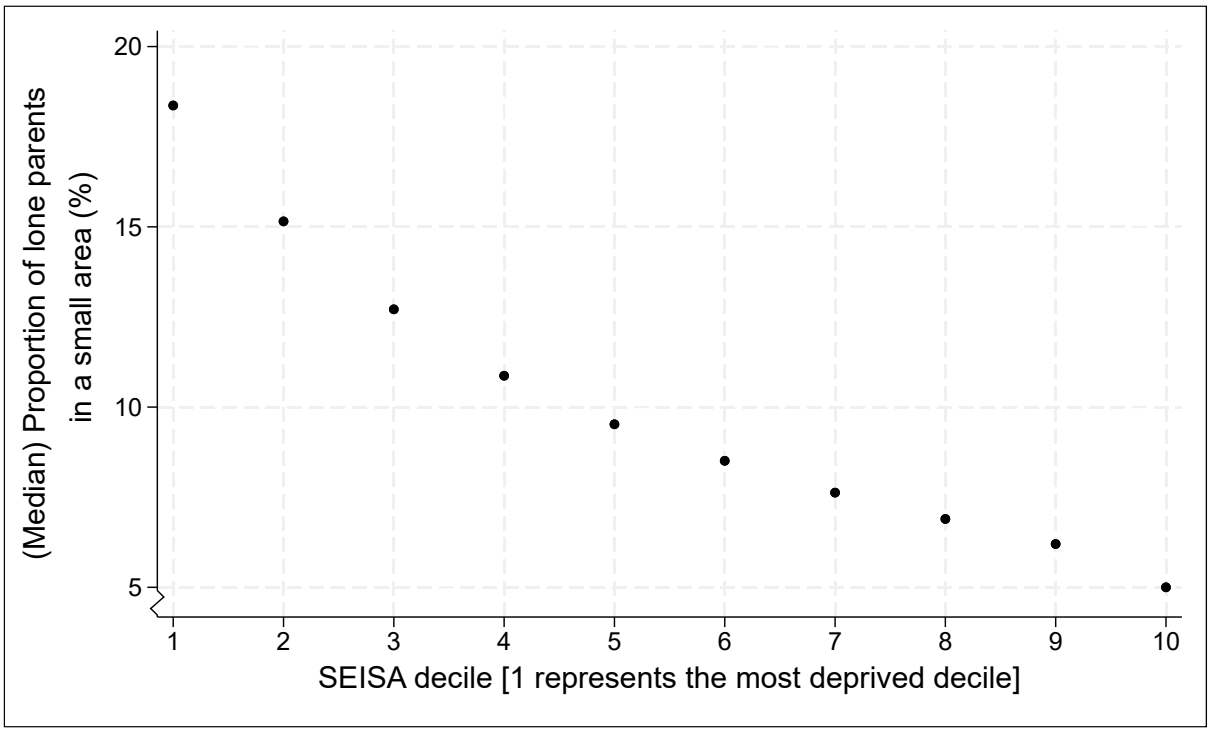
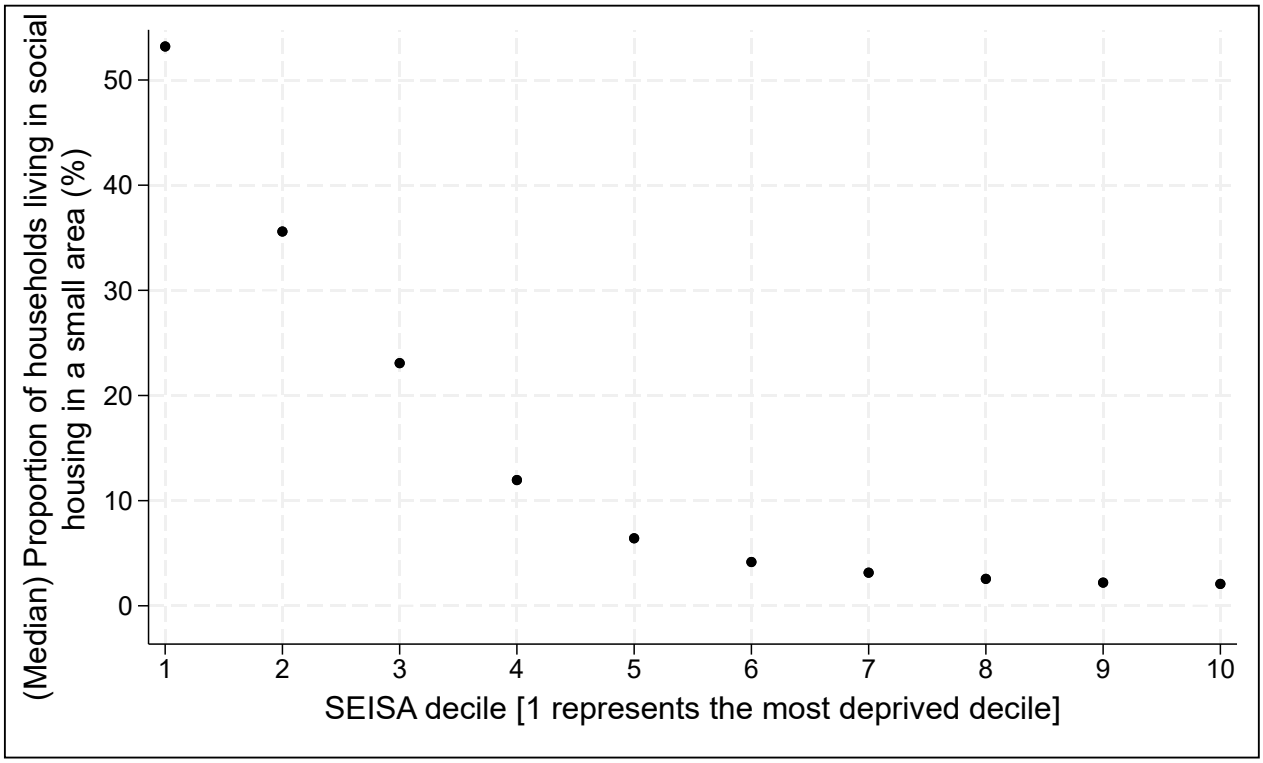


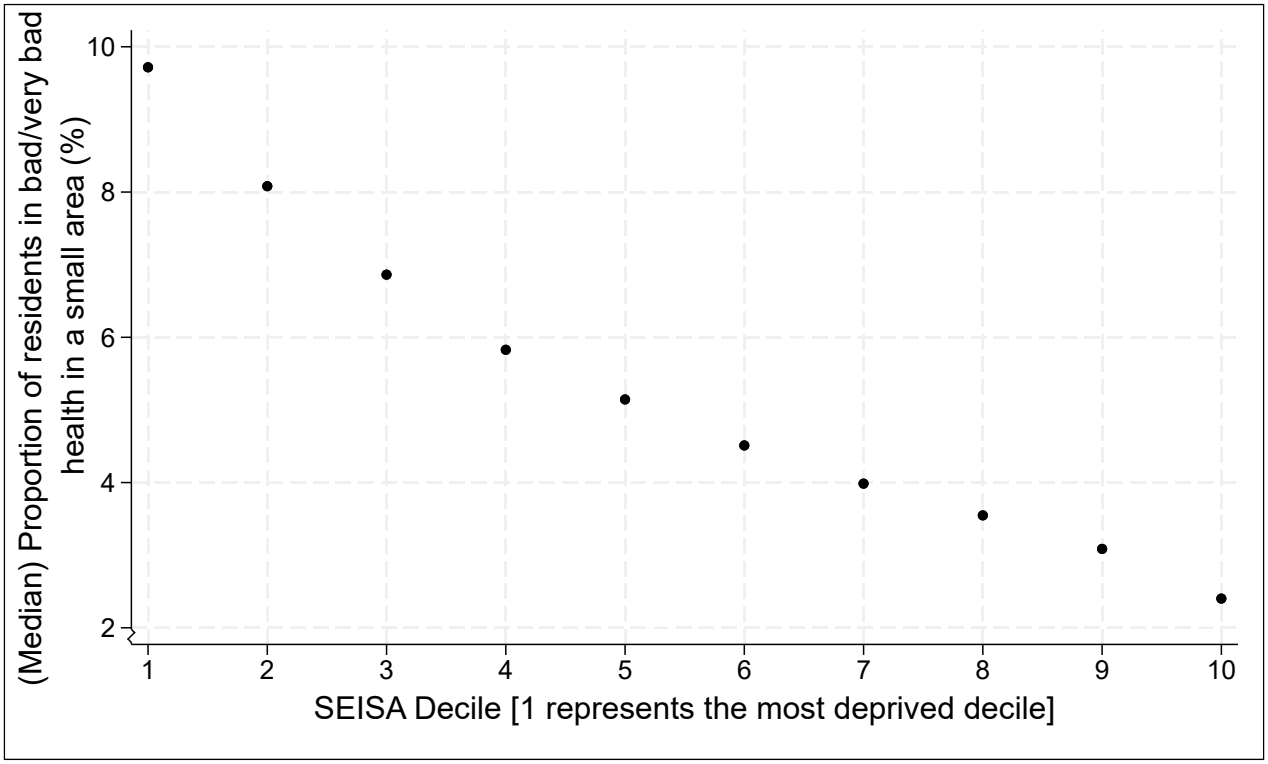

Table 26 in ONS (2022) highlights that those in social housing often have the lowest disposable incomes, while reports by the Ministry of Housing, Communities and Local Government (2020) and Scottish Government (2019b) indicate that access to internet among home owners and those privately renting is quite similar, but evidently lower among those in social housing. As the Welsh Government (2020) study points out, lack of internet connection is associated with a higher probability of being in material deprivation. Our Census variable on housing tenure allows us to create a variable that corresponds to the proportion of households in an output area that are in social housing. In Figure 7 below, we see that there is a clear link between our deprivation measure and housing tenure, with the average proportion of households in social housing rising as we move from the least deprived deciles to those that represent the most deprived output areas.

**Figure 7: The relationship between SEISA and housing tenure across the UK**




Finally, we already pointed out The Health Foundation (2020) and Welsh Government (2023) reports on the correlation between poor health and income/material deprivation, so we use the Census indicator on self-reported health to generate a field that relates to the proportion of residents in an output area that are in bad or very bad health. Figure 8 illustrates that those individuals living in output areas that are classified as most deprived according to our measure have a higher likelihood of reporting poor self-reported health outcomes.

**Figure 8: The relationship between SEISA and self-reported health across the UK**



## Section 5: Should housing tenure form part of a composite measure based on the Census?

As we discussed earlier, the relationship between income and housing generally indicates that it is low levels of income that drive housing outcomes, rather than vice-versa. This is also noted in a report by the Joseph Rowntree Foundation (2015). However, one of the key features of the UK housing market over recent decades has been the growing proportion of households in the private rented sector. The continued rise in rental costs combined with stagnating wages has meant that housing arrangements can lead to more people entering poverty once living costs are taken into account, with this particularly being the case for those in (private) rental accommodation.

We therefore carried out a sensitivity analysis to see how the inclusion of a variable relating to housing tenure in our composite measure altered our findings. Hence, we began by creating a field that indicated the proportion of households in an output area that were either privately or socially renting their accommodation. A principal component analysis was then carried out using this variable, alongside our two indicators on education and occupation. The correlation between a composite index including housing tenure and one created using only education/occupation (i.e. SEISA) was 0.97. Typically, when attempting to determine which areas to target, organisations will tend to concentrate first on those in the bottom quintile. We therefore also assessed how the bottom quintile of our composite measure inclusive of housing tenure differed to one based on only education and occupation. A two-by-two matrix of the findings is illustrated in the table below.

**Table 2: Cross-tabulation of SEISA against a measure using occupation, education and household tenure. Figures represent output areas.**

| | Quintile 2-5 (including tenure) | Quintile 1 (including tenure) | Total |
|---|---|---|---|
| **Quintile 2-5 (excluding tenure)** | 178,977 | 6,860 | 185,837 |
| **Quintile 1 (excluding tenure)** | 6,860 | 39,599 | 46,459 |
| **Total** | 185,837 | 46,459 | 232,296 |

However, when using MSOA income data for England and Wales to evaluate the incomes of those in the bottom quintile of the two composite measures, we found median incomes in the bottom quintiles to be equivalent. Furthermore, the average incomes in output areas that emerge in the bottom quintile of SEISA, but a higher quintile when including housing tenure, are slightly lower (at £380) when compared with

output areas that are in the bottom quintile of a measure that uses housing tenure, but not SEISA (£390). There is therefore no strong theoretical or empirical evidence to suggest that including housing tenure can materially improve the measure.

**Table 3: Cross-tabulation of SEISA against a measure using occupation, education and household tenure. Figures represent median net equivalised weekly household income estimates after housing costs in England and Wales (£)**

| | Quintile 2-5 (including tenure) | Quintile 1 (including tenure) | Total |
|---|---|---|---|
| **Quintile 2-5 (excluding tenure)** | £470 | £390 | £470 |
| **Quintile 1 (excluding tenure)** | £380 | £350 | £360 |
| **Total** | £470 | £360 | £450 |

## Section 6: How does our measure compare to the Indices of Deprivation?

For our measure to be valuable and useful to end users, it should also ideally be able to overcome some of the known limitations of the Indices of Deprivation. The aim of this section is to therefore explore whether this is the case across the various nations of the UK.

As we have income estimates at the MSOA level in England, one of our first lines of analysis in this country was to look at the average net equivalised weekly household income (after accounting for housing costs) of those living in areas that fall into the bottom quintile of the English Index of Multiple Deprivation (IMD). The median value was found to be £350, which is only slightly lower than the figure one obtains if they consider the lowest quintile of SEISA when applied in England only (£360). Here, SEISA is formulated by splitting English output areas only into five equal quintiles (hence the output/small areas of all other nations were excluded when creating this categorical variable). Similar averages were also found if focus was placed on particular domains of IMD, such as income and education, as well as for IDACI.

34,274 output areas were located within the bottom quintile of SEISA when applied to England only, with 12,063 of these not being found in the corresponding quintile of IMD – representing just over a third of the total. When looking at where those output areas only found in the bottom quintile of our composite measure were located (and hence were situated within a higher quintile of IMD), we observed that almost 30% were in central England, with just over a quarter found in East of England or Yorkshire and The Humber.

Furthermore, when examining the types of areas that the bottom quintile of IMD tends to capture, only 3% are classified as rural areas. Rather, IMD appears particularly effective in catching localities in major urban areas, such as London and Birmingham. Indeed, one of the key differences we find is that 12% of output areas within the bottom quintile of IMD are in London, with the equivalent figure being 3% for SEISA. In contrast, 7% of areas in the lowest quintile of our measure are rural locations, with our measure tending to pick up a greater proportion of medium and large towns in England. Table 4 illustrates the urban/rural classification of the output areas that fall into the bottom quintile of IMD/SEISA.

**Table 4: Urban/Rural classification of areas that fall into the bottom quintile of IMD and SEISA (when applied to England only)**

| Urban/rural classification | IMD (%) | SEISA (%) |
|---|---|---|
| Rural hamlets and isolated dwellings | 0.3 | 0.3 |
| Rural hamlets and isolated dwellings in a sparse setting | 0.0 | 0.0 |
| Rural town and fringe | 2.4 | 5.6 |
| Rural town and fringe in a sparse setting | 0.1 | 0.4 |
| Rural village | 0.3 | 1.0 |
| Rural village in a sparse setting | 0.0 | 0.1 |
| Urban city and town in a sparse setting | 0.2 | 0.4 |
| Urban city and town | 40.5 | 48.7 |
| Urban major conurbation | 49.7 | 36.4 |
| Urban minor conurbation | 6.4 | 7.1 |
| Total number of output areas | 34,229 | 34,274 |

When assessing how our measure compared to some of the individual domains of IMD/IDACI, we found that similar overall conclusions were reached when considering IDACI, as well as the income and employment domains. That is, around one-third to two-fifths of output areas in the bottom quintile of our measure did not appear in the equivalent quintile of these variables. Typically, IDACI and these components of IMD were again found to be picking up predominantly major urban conurbations and very few rural areas. The indicator that showed greatest similarity with our own measure with regards to the urban/rural distribution was the education domain. However, even in this instance, approximately 30% of

the output areas that emerged in the bottom quintile of our measure were located in a higher quintile of the education domain, with these localities typically found in northern and central England.

In Scotland, modelled estimates of income and poverty relating to the period 2008/09 have been published by the Scottish Government (2014). We utilise the weekly net equivalent household income after housing costs below to assist us in contextualising our findings. Please note that these estimates could not be directly incorporated into our dataset as the geography codes utilised are based on the 2001 Census. As has been highlighted in statistics released on SIMD, not all council areas are covered in the bottom quintile of this index, despite these areas being known to have deprivation. For example, Na h-Eileanan Siar is a place that does not feature in the bottom quintile of SIMD, despite Scottish Government (2014) estimates indicating that it has a median weekly income of £334, which is below the national average of £349. Applying SEISA in Scotland only (so that each Scottish SEISA quintile contains the same number of Scottish output areas), we find that 3,629 of the 9,270 Scottish output areas that fall within the bottom quintile of our composite measure are located in a higher quintile of SIMD, which represents almost 40% of the total. One of the key differences between our measure and SIMD is that all council areas of Scotland are covered through the bottom quintile of our index.

As we pointed out earlier, one of the main criticisms of SIMD is its inability to capture more rural parts of the country. Table 5 demonstrates that only 3% of areas in the bottom quintile of SIMD are listed as rural areas. In contrast, of the localities in the lowest 20% of our measure, 7% are categorised as accessible/remote rural vicinities.

Even if one were to compare our measure with the income, employment and education domains of SIMD, the overall conclusions we report here do not change. That is, around 40% of output areas that are in the bottom quintile of our measure are in a higher quintile of the SIMD domain being considered, with our measure able to capture a larger proportion of rural areas.

**Table 5: Urban/Rural classification of areas that fall into the bottom quintile of SIMD and SEISA (when applied to Scotland only)**

| Urban/rural classification | SIMD (%) | SEISA (%) |
|---|---|---|
| Large urban areas | 52.5 | 39.8 |
| Other urban areas | 37.6 | 41.9 |
| Accessible small towns | 5.2 | 8.5 |
| Remote small towns | 1.6 | 2.7 |
| Accessible rural | 2.2 | 5.1 |
| Remote rural | 1.0 | 2.0 |
| Total number of output areas | 10,025 | 9,270 |

A similar approach to that taken for England was applied for our analysis in Wales. That is, we began by utilising the MSOA income estimates to understand how the net weekly household income levels (equivalised and after housing costs) compared between the bottom quintile of our composite measure (when applied to Welsh output areas only) and WIMD. We found the median value for WIMD to be £330, with an equivalent estimate emerging if one were to concentrate on the income, employment or education domains of WIMD. The same median figure was found for SEISA when applied to Wales only.

2,007 output areas fall into the bottom quintile of our composite measure, with 763 (38%) of these being based in a higher quintile of WIMD. While these 763 output areas are spread across all parts of Wales, we

do find that almost a quarter of these are found in two local authorities in south Wales (Rhondda and Caerphilly).

As in the other nations assessed so far, we find that the bottom quintile of our composite measure comprises of a larger proportion of rural areas than WIMD. One-fifth of areas in the bottom quintile of our measure are categorised as rural compared with approximately 14% for WIMD. Table 6 illustrates this. Similar findings emerge for Wales if we were to compare our measure with the income, employment or education domains of WIMD.

**Table 6: Urban/Rural classification of areas that fall into the bottom quintile of WIMD and SEISA (when applied to Wales only)**

| Urban/rural classification | WIMD (%) | SEISA (%) |
|---|---|---|
| Rural hamlets and isolated dwellings | 0.4 | 0.3 |
| Rural hamlets and isolated dwellings in a sparse setting | 0.1 | 0.1 |
| Rural town and fringe | 11.8 | 15.4 |
| Rural town and fringe in a sparse setting | 1.0 | 2.4 |
| Rural village | 0.6 | 1.5 |
| Rural village in a sparse setting | 0.1 | 0.7 |
| Urban city and town in a sparse setting | 2.2 | 1.8 |
| Urban city and town | 84.0 | 77.9 |
| Urban major conurbation | 0.0 | 0.0 |
| Urban minor conurbation | 0.0 | 0.0 |
| Total number of output areas | 1,988 | 2,007 |

Within Northern Ireland, 907 small areas fall into the bottom quintile of our composite measure (when applied in Northern Ireland only), with nearly 42% (380) of these localities found in a higher quintile of NIMDM. Those places that emerge in the bottom quintile of our measure only are generally found in north and/or eastern parts of Northern Ireland (e.g. Belfast, Mid and East Antrim, Antrim and Newtownabbey, as well as Armagh City, Banbridge and Craigavon).

Table 7 highlights the urban/rural classification of areas that fall into the bottom quintile of our measure and the equivalent figures for NIMDM. In contrast to what we observe in other nations, we do not see a great deal of difference by urban/rural classification. 85% of output areas in the bottom quintile of our measure are classified as urban, with the figure being marginally lower (84%) for the lowest 20% of areas according to NIMDM. So far, when we have extended our work to include comparisons between individual domains and our measure, the conclusions we have drawn have not been altered. In Northern Ireland, however, an assessment of our measure against the income or IDAC domains does lead to different results to what we see for the overall index. In particular, around 30% of output areas in the bottom quintile of either domain are classified as rural, which bucks the general trend we have seen across nations (i.e. that our measure captures a greater proportion of rural spots).

**Table 7: Urban/Rural classification of areas that fall into the bottom quintile of NIMDM and SEISA (when applied to Northern Ireland only)**

| Urban/rural classification | NIMDM (%) | SEISA (%) |
|---|---|---|
| Mixed urban/rural | 2.6 | 3.5 |
| Rural | 13.2 | 11.5 |
| Urban | 84.2 | 85.0 |
| Total number of small areas | 930 | 907 |

## Section 7: Discussion and concluding remarks

Our aim in this paper was to firstly outline the creation of a new UK-wide area-based measure of deprivation using Census 2011. Having done so, we then assessed its correlation with factors associated with deprivation and the extent to which it differs to the Indices of Deprivation. Overall, the findings provide evidence in favour of our measure being related to deprivation as expected and also offering added value to users of deprivation statistics through overcoming some of the known limitations of the Indices of Deprivation. Furthermore, given it is UK-wide, our measure can also enable comparable statistics to be developed across nations – a mark of quality as highlighted in the UK Code of Practice for Statistics.

In England, Wales and Scotland, our measure captures a greater proportion of rural areas in the bottom quintile when compared with a multiple deprivation index. Within Scotland, all council areas emerge in the bottom quintile of SEISA, which is not the case for SIMD, where some known areas of deprivation are not found in the bottom quintile. For England, the lowest quintile of the variable we have created also has the advantage (over the same quintile of IMD) of picking up medium/large towns in northern and central England. As in Scotland, the bottom quintile of our measure covers all local authorities/LGDs in Wales and Northern Ireland, respectively. In Wales, one of the key differences between our measure and the bottom quintile of WIMD is the greater ability of our variable to catch deprivation across two local authorities in the south. Meanwhile, for Northern Ireland, the benefit of our measure relative to NIMDM (based on our assessment of the lowest quintile) appears to lie in picking up parts of the north and/or east of the country.

Although taking part in the Census is obligatory, it should be acknowledged that the data is still self-reported and could therefore be subject to measurement error, as individuals may be unwilling or unable to give an accurate response to questions on their qualifications and occupation. Indeed, the ONS (2014) completed a Census Quality Survey between May and August 2011, in which volunteers were requested to participate in a face-to-face interview. Those who agreed were asked the same questions that they responded to approximately two to five months earlier, with the key difference being the mode of survey completion. Agreement rates were generally found to be lower (around two-thirds) for the questions on occupation and education, which tended to be for reasons such as respondents giving different job titles or struggling to recall the qualifications they had attained. Additionally, in this paper, our evaluation of the degree to which our measure is correlated with income relies upon estimates at a relatively aggregated level of geography. We recognise that within such areas, there is likely to be variation in the incomes of households. In the case of Scotland and Northern Ireland, we had no relevant data that we could draw upon.

The 2021 Census has now been administered across all nations of the UK. Once this data is released into the public domain, we will seek to produce an updated version of the measure we have presented here.

## References


**Allik, M., Brown, D., Dundas, R. & Leyland, A. (2016).** Developing a new small-area measure of deprivation using 2001 and 2011 census data from Scotland. *Health and Place*, 39, pp 122-130 doi: https://doi.org/10.1016/j.healthplace.2016.03.006

**Australian Bureau of Statistics (2016).** *Socio-Economic Indexes for Areas (SEIFA) Technical Paper.* Available at https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/756EE3DBEFA869EFCA258259000BA746/$File/SEIFA%202016%20Technical%20Paper.pdf

**Carstairs, V & Morris, R. (1989).** Deprivation: Explaining differences in mortality between Scotland and England and Wales. *British Medical Journal*, 299(6704), pp 886-889 doi: 10.1136/bmj.299.6704.886

**Christie, S. & Fone, D. (2003).** Does car ownership reflect socioeconomic disadvantage in rural areas? A cross-sectional geographical study in Wales, UK. *Public Health*, 117(2), pp 112-116 doi: https://doi.org/10.1016/S0033-3506(02)00027-6

**Commission on Widening Access (2016).** *Technical paper on measures and targets.* Available at https://www.gov.scot/binaries/content/documents/govscot/publications/progress-report/2016/03/blueprint-fairness-final-report-commission-widening-access/documents/00496620-pdf/00496620-pdf/govscot%3Adocument/00496620.pdf

**Department for Environment, Food & Rural Affairs (2021).** '2011 Rural Urban Classification lookup tables for all geographies'. Available at https://www.gov.uk/government/statistics/2011-rural-urban-classification-lookup-tables-for-all-geographies

**Francis-Devine, B. (2021).** *Average earnings by age and region*. Available at https://commonslibrary.parliament.uk/research-briefings/cbp-8456/

**Gordon, D. (1995).** Census based deprivation indices: their weighting and validation. *Journal of Epidemiology and Community Health*, 49(Suppl 2), pp S39-S44 doi 10.1136/jech.49.suppl_2.s39

**HM Government (2014).** *An evidence review of the drivers of child poverty for families in poverty now and for poor children growing up to be poor adults*. Available at

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/285389/Cm_8781_Child_Poverty_Evidence_Review_Print.pdf

**Joseph Rowntree Foundation (2015).** *Housing and poverty*. Available at https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/housing-poverty-roundup-full.pdf

**Ministry of Housing, Communities and Local Government (2019).** 'English indices of deprivation 2019'. Available at https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019

**Ministry of Housing, Communities and Local Government (2020).** *English Housing Survey: Headline report, 2019-20*. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/945013/2019-20_EHS_Headline_Report.pdf

**National Records of Scotland (2011).** 'Output area 2011 to Data Zones and Intermediate Zones 2011'. Available at https://www.nrscotland.gov.uk/statistics-and-data/geography/our-products/census-datasets/2011-census/2011-indexes

**National Records of Scotland (2015).** *Census geography comparison with England and Wales.* Available at https://www.nrscotland.gov.uk/files/geography/2011-census/geography-bckground-info-comparison-of-thresholds.pdf

**National Records of Scotland (2021b).** '2011 Census table data: Output Area 2011'. Available at https://www.scotlandscensus.gov.uk/documents/2011-census-table-data-output-area-2011/

**Northern Ireland Statistics and Research Agency (2013a).** 'Northern Ireland Super Output Areas'. Available at https://www.nisra.gov.uk/support/geography/northern-ireland-super-output-areas

**Northern Ireland Statistics and Research Agency (2013b).** 'Small area Look-up Tables and Guidance Documents'. Available at https://www.nisra.gov.uk/publications/small-area-look-tables-and-guidance-documents

**Northern Ireland Statistics and Research Agency (2017).** 'Northern Ireland Multiple Deprivation Measure 2017 (NIMDM 2017)'. Available at https://www.nisra.gov.uk/publications/nimdm17-soa-level-results

**Northern Ireland Statistics and Research Agency (2019).** *NISRA Geography Fact Sheet.* Available at https://www.ninis2.nisra.gov.uk/public/documents/NISRA%20Geography%20Fact%20Sheet.pdf

**Northern Ireland Statistics and Research Agency (2021).** '2011 Census Key Statistics'. Available at https://www.nisra.gov.uk/statistics/2011-census/results/key-statistics

**Office for National Statistics (2001).** 'NS-SEC by highest level of qualification'. Available at https://www.nomisweb.co.uk/datasets/cs114/view?geography=2092957699&c_nssec=0,1,2,3,4,5,6,7,8,9&c_hlqpuk11=0,1,2,3,4,5,6&rows=c_nssec&cols=c_hlqpuk11&tools=hidden

**Office for National Statistics (2014).** *Assessing accuracy of responses (Census Quality Survey)*. Available at https://webarchive.nationalarchives.gov.uk/ukgwa/20160108085257/http:/www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/assessing-accuracy-of-responses--census-quality-survey-/index.html

**Office for National Statistics (2015).** *2011 Census UK Comparability*. Available at https://ukdataservice.ac.uk/app/uploads/2011censusesukcomparabilityreport.pdf

**Office for National Statistics (2018a).** 'Output area to region (December 2011) lookup in England'. Available at https://geoportal.statistics.gov.uk/datasets/ons::output-area-to-region-december-2011-lookup-in-england/explore

**Office for National Statistics (2018b).** 'Output area to lower layer super output area to middle layer super output area to local authority district (December 2011) lookup in England and Wales'. Available at https://geoportal.statistics.gov.uk/datasets/ons::output-area-to-lower-layer-super-output-area-to-middle-layer-super-output-area-to-local-authority-district-december-2011-lookup-in-england-and-wales/about

**Office for National Statistics (2018c).** 'The mean and median gross weekly and gross hourly earnings measured by highest education qualification.' Available at https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/adhocs/008042themeanandmediangrossweeklyandgrosshourlyearningsmeasuredbyhighesteducationqualification

**Office for National Statistics (2020).** 'Income estimates for small areas, England and Wales.' Available at https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/smallareaincomeestimatesformiddlelayersuperoutputareasenglandandwales

**Office for National Statistics (2021a).** *Census geography*. Available at https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography

**Office for National Statistics (2021b).** '2011 Census: Bulk Data Download (Release 2)'. Available at https://www.nomisweb.co.uk/census/2011/bulk/r2_2#KeyStatistics

**Office for National Statistics (2022).** 'The effect of taxes and benefits on household income, disposable income estimate – Financial year ending 2021'. Available at https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/datasets/householddisposableincomeandinequality

**Office for Statistics Regulation (2019).** *Exploring the public value of statistics about post-16 education and skills in England*. Available at https://osr.statisticsauthority.gov.uk/wp-content/uploads/2019/04/post-16-education-and-skills-in-England-SR-April-2019.pdf

**Scottish Government (2014).** 'Income and poverty modelled estimates'. Available at https://statistics.gov.scot/data/income-and-poverty-modelled-estimates

**Scottish Government (2018).** *Equality Statement: Scottish Budget 2019-20*. Available at https://www.gov.scot/binaries/content/documents/govscot/publications/impact-assessment/2018/12/equality-fairer-scotland-budget-statement-scottish-budget-2019-20/documents/scottish-budget-2019-20-equality-fairer-scotland-statement/scottish-budget-2019-20-equality-fairer-scotland-statement/govscot%3Adocument/00544000.pdf

**Scottish Government (2019a).** '2011 Data Zone Lookup'. Available at https://statistics.gov.scot/data/data-zone-lookup

**Scottish Government (2019b).** *Scottish Household Survey 2018: annual report.* Available at https://www.gov.scot/publications/scotlands-people-annual-report-results-2018-scottish-household-survey/pages/7/

**Scottish Government (2020).** 'Scottish Index of Multiple Deprivation 2020v2 – ranks'. Available at https://www.gov.scot/publications/scottish-index-of-multiple-deprivation-2020v2-ranks/

**The Health Foundation (2020).** *Living in poverty was bad for your health long before covid-19*. Available at https://www.health.org.uk/publications/long-reads/living-in-poverty-was-bad-for-your-health-long-before-COVID-19

**Townsend, P. (1987).** Deprivation. *Journal of Social Policy*, 16(2), pp 125-146 https://doi.org/10.1017/S0047279400020341

**UK Data Service (2022).** *Census Forms*. Available at https://ukdataservice.ac.uk/learning-hub/census/about-uk-censuses/census-forms/

**Welsh Government (2015).** *Welsh Index of Multiple Deprivation 2014: A guide to analysing deprivation in rural areas*. Available at https://gov.wales/sites/default/files/statistics-and-research/2019-05/welsh-index-of-multiple-deprivation-2014-a-guide-to-analysing-deprivation-in-rural-areas.pdf

**Welsh Government (2019).** 'Welsh Index of Multiple Deprivation (full index update with ranks): 2019'. Available at https://gov.wales/welsh-index-multiple-deprivation-full-index-update-ranks-2019

**Welsh Government (2020).** *What factors are linked to people living in households that are in material deprivation?* Available at https://www.gov.wales/sites/default/files/statistics-and-research/2020-03/what-factors-are-linked-to-people-living-in-households-that-are-in-material-deprivation.pdf

**Welsh Government (2023).** *Poverty and deprivation (National Survey for Wales): April 2021 to March 2022*. Available at https://www.gov.wales/sites/default/files/pdf-versions/2023/2/4/1677139488/poverty-and-deprivation-national-survey-wales-april-2021-march-2022.pdf

**Yousaf, S. & Bonsall, A. (2017).** *UK Townsend Deprivation Scores from 2011 census data.* Available at https://s3-eu-west-1.amazonaws.com/statistics.digitalresources.jisc.ac.uk/dkan/files/Townsend_Deprivation_Scores/UK%20Townsend%20Deprivation%20Scores%20from%202011%20census%20data.pdf

## Appendix: A further look into the application of SEISA

We argue that a key advantage of our measure is that it can be applied in a UK-wide analysis and for comparing statistics across the four nations. In this section, we summarise additional sensitivity checks we carried out regarding these features of SEISA. Specifically, we illustrate that where there are differences by nation in the way Census data is collected/disseminated, this does not seem to be resulting in the patterns we observe in the distribution of output/small areas. Furthermore, we note the consistency in the methodology used in generating SEISA across all four nations.

The ONS (2015) report highlights that key statistics on qualifications are broadly comparable, while they are highly comparable in the case of occupation. In the case of qualifications, although not discussed in the ONS (2015) publication, we do note that HNC/HND qualifications are placed in the 'Level 3' category within Scotland, but 'Level 4' in all other nations. To the best of our knowledge, the way in which HNC/HND qualifications have been grouped was determined based on user need from Census outputs.

Below, we provide a breakdown of how output/small areas are distributed across quintiles in each nation. We can see that, in the case of Scotland, 29% of output areas emerge in quintile 1 of the UK-wide measure. When utilising SEISA for a UK-wide analysis (i.e. a study which includes individuals from all nations), one may question whether the bottom quintile picks up a greater proportion of Scottish output areas due to the way qualifications have been banded across the various levels in Census outputs. However, to illustrate that this is not driven by the difference in the manner in which HNC/HND qualifications are grouped, we also provide a table illustrating how output/small areas are dispersed across the quintiles when based on the occupation variable only (which is highly comparable).

Evaluating the two tables (A1 and A2) demonstrates that the findings are very similar. As stated in the main paper, the very high linear correlation between qualifications and occupation means that there will not be a great deal of difference between a measure of deprivation based on both compared to one formed using one of the two variables only.

**Table A1: The distribution of output/small areas across quintiles by nation for SEISA**

| | England | Northern Ireland | Scotland | Wales |
|---|---|---|---|---|
| Quintile 1 | 17.2 | 27.2 | 28.8 | 23.2 |
| Quintile 2 | 20.1 | 24.8 | 18.5 | 23.3 |
| Quintile 3 | 20.9 | 22.9 | 15.9 | 22.1 |
| Quintile 4 | 21.1 | 15.1 | 16.6 | 18.9 |
| Quintile 5 | 20.7 | 10.0 | 20.1 | 12.5 |
| Total | 171,372 | 4,537 | 46,351 | 10,036 |

**Table A2: The distribution of output/small areas across quintiles by nation based on the Census occupation variable only**

| | England | Northern Ireland | Scotland | Wales |
|---|---|---|---|---|
| Quintile 1 | 17.2 | 29.2 | 28.8 | 23.4 |
| Quintile 2 | 19.8 | 26.5 | 19.1 | 24.8 |
| Quintile 3 | 20.8 | 20.8 | 16.3 | 22.6 |
| Quintile 4 | 21.3 | 13.8 | 16.2 | 17.3 |
| Quintile 5 | 20.9 | 9.7 | 19.6 | 11.9 |
| Total | 171,372 | 4,537 | 46,351 | 10,036 |

One may also hypothesise that the higher proportion of Scottish output areas that are observed in the bottom quintile could be due to Scottish output areas being smaller than those of other countries. To assess this, we carried out an additional analysis whereby we formulated our occupation variable (i.e. the most comparable Census field) using data zones in Scotland, but output/small areas in all other nations. In this scenario, it is Scotland that has the areas with the highest population sizes. The distribution by nation

for the occupation field is demonstrated below. Even in this instance, a quarter of Scottish data zones fall within the bottom quintile. Consequently, it does not seem to be the case that output area size is the reason behind why a higher proportion of Scottish areas emerge in the bottom quintile.

**Table A3: The distribution of output/small areas (data zones in Scotland) across quintiles by nation based on the Census occupation variable only**

| | England | Northern Ireland | Scotland | Wales |
|---|---|---|---|---|
| Quintile 1 | 19.1 | 31.8 | 24.7 | 26.4 |
| Quintile 2 | 19.5 | 25.8 | 21.6 | 24.0 |
| Quintile 3 | 20.0 | 19.7 | 19.1 | 21.2 |
| Quintile 4 | 20.5 | 13.0 | 17.9 | 16.5 |
| Quintile 5 | 20.9 | 9.7 | 16.7 | 11.9 |
| Total | 171,372 | 4,537 | 6,976 | 10,036 |

Please note that given this occupation variable was created using a different geographic domain in Scotland compared to the other nations, we have not supplied this field as part of the 'Research dataset' available through our interactive map webpages. Users who wish to recreate Table A3 could do so using the links/data provided in the 'Research dataset' for output/small areas in England, Wales and Northern Ireland. This would need to be supplemented with occupation information for Scottish data zones (available at https://www.scotlandscensus.gov.uk/documents/2011-census-table-data-sns-data-zone-2011/).

Finally, for a measure of deprivation to be suitable for comparing statistics across nations, a key criteria that it should meet is that it is created using the same methodology across all countries. This is the case with SEISA. Indeed, as at a UK-wide level, whether one uses both qualifications and occupation to create the measure in a particular country or just the occupation variable is not likely to make any material difference to the analysis/results, given the strong linear correlation between the two fields at nation-level too. The highest correlations are observed in Northern Ireland and Wales (0.95 and 0.94, respectively), while the figures were 0.92 in England and 0.89 in Scotland.