Received: 9 June 2023 | Revised: 22 September 2023 | Accepted: 27 September 2023

DOI: 10.1111/geoj.12550

ARTICLE

WILEY

A neighbourhood Output Area Classification from the 2021 and 2022 UK censuses

Jakub Wyszomierski[1] | **Paul A. Longley**[1] | **Alex D. Singleton**[2] | **Christopher Gale**[3] | **Oliver O'Brien**[1]

[1]Department of Geography, University College London, London, UK

[2]School of Environmental Sciences, University of Liverpool, Liverpool, UK

[3]Homes England, London, UK

Correspondence
Paul A. Longley, Department of Geography, University College London, Gower Street, London WC1E 6BT, UK.
Email: p.longley@ucl.ac.uk

Funding information
Office for National Statistics; Economic and Social Research Council; Consumer Data Research Centre, Grant/Award Number: ES/L011840/1

Abstract

UK-wide multivariate neighbourhood classifications have been built using small area population data following every census since 1971, and have been built using Output Area geographies since 2001. Policy makers in both the public and private sectors find such taxonomies, typically arranged into hierarchies of Supergroups, Groups and Subgroups, useful across a wide range of applications in business and service planning. Recent and forthcoming releases of small area census statistics pose new methodological challenges. For example, the 2022 Scottish Census was carried out a year after those in other UK nations, and some of the variables now collected across different jurisdictions do not bear direct comparison with one another. Here we develop a methodology to accommodate these issues alongside the more established procedures of variable selection, standardisation, transformation, class definition and labelling.

KEYWORDS

cluster analysis, geodemographics, Output Area Classification, UK Census of Population

1 | INTRODUCTION

The accelerating complexity of advanced societies such as the UK brings renewed challenges for representing the nature and scale of residential differentiation. Underpinning data infrastructure for entire populations available for small area aggregations is provided in many countries, including the UK, by periodic (typically decennial) censuses, with the remit of ascertaining the general social, economic and demographic conditions of the population and the neighbourhoods in which they live. Such sources enable planners and policy makers in both the public and private sectors to make informed decisions about resource allocation and service delivery.

While direct or simple composite indicators are sufficient for many applications, many others benefit from summaries of multiple variables that include a fuller range of census data. Since the mid-twentieth century there has been a growing interest in the utilisation of census data in the creation of geodemographic 'analyses of people by where they live' (Harris et al., 2005). Such classifications use data reduction techniques to simplify multidimensional variables into qualitative typologies with distinctive characteristics at neighbourhood scales (Singleton & Longley, 2015).

Applied to the UK, such classifications have provided nationwide snapshots of neighbourhood conditions on census nights since the 1970s (Charlton et al., 1985; Vickers & Rees, 2007; Webber, 1977). These depictions have laid the foundations for recent Office for National Statistics (ONS) Output Area Classifications (Gale et al., 2016; Vickers & Rees, 2007), and have spawned extensions to bespoke geodemographic applications within a variety of fields including education (Singleton & Longley, 2009), ageing populations (Yang et al., 2023), health (Petersen et al., 2011), public safety (Anderson, 2010; Ashby & Longley, 2005; Gulma, 2022) and digital equity (Longley et al., 2008; Singleton & Longley, 2015). Although most application-specific geodemographic classifications make use of a wide range of administrative or consumer data sources, most remain heavily reliant upon census data for specification, estimation and testing of classifications (Harris et al., 2005; Stillwell, 2017).

The ONS commissioned research to produce a 2021 Output Area Classification (OAC) some time before the major disruption to the fulfilment of the 2021 Census that arose because of the COVID-19 pandemic. This led to additional requirements to accommodate the rescheduling of the Scottish part of the Census to 20 March 2022 and changed patterns of work, travel and residence (to differing degrees) in 2021 and 2022. Changed circumstances made some census data (such as mode of travel to work) less relevant to a classification designed to predict needs. It also led to delays in release of some 2021 outputs (e.g., for Northern Ireland) and all of those collected in Scotland in 2022.

This paper sets out the ways in which changed circumstances were accommodated, and describes technical measures used to improve the classification process. The resulting 2021/22 OAC uses a methodology to accommodate the asynchronous collection and release of Census data by constituent UK countries. It uses 2021 Census Output Area data for England and Wales; aggregate published 2021 Northern Ireland data, apportioned to 2011 small area census geography; and 2011 Scottish Census results carried forward in adjusted or unadjusted form. The latter two of these sources are used to create place-markers for distinctive Northern Irish and Scottish neighbourhood characteristics for replacement with observed estimates when these become available. In what follows, we begin by discussing general pandemic-related changes in neighbourhood composition and function, alongside the specific challenges introduced by asynchronous data collection and dissemination. We then set out our methodology to frame a 2021/2 OAC for the entire UK. We extend the underpinning methodology proposed by Gale et al. (2016) for the 2011 OAC, and present the results of the clustering and demonstrate the characteristics of the derived geodemographic groups. Lastly, we discuss the validity of the classification and its prospects following the availability of all 2021/2 Census data.

2 | METHODOLOGICAL CHALLENGES

When the COVID-19 virus began to take hold in the UK in March 2020, it was initially unclear what implications it held for the implementation of the Census scheduled for 21 March 2021. Only in July 2020 did ONS and the Northern Ireland Statistics and Research Agency (NISRA) confirm that they would each proceed on the planned Census Night, while National Records for Scotland (NRS) decided to postpone the Scottish Census until March 2022, citing requirements to preserve data quality. There is no doubt that the 2021/2 UK Census presents a snapshot of the population at an extremely unusual time, in terms of residence, employment and travel behaviour. In 2021 many people were temporarily resident in parental or holiday homes, furlough was distorting many sectors of the labour market, and homeworking became a norm for many groups. Particularly problematic was enumeration of the student population, part of which had returned to their family homes, or in the case of international students, had left the country. This issue was also amplified by a 2021 enumeration date that occurred at the end of the spring term. Nevertheless, the scope of the questions had been agreed in the Census (England and Wales) 2020 legislation, and it was not deemed possible to amend existing questions or add new ones so that the questionnaire might provide a more salient representation of either normal circumstances or those prevailing on Census night (Office for National Statistics, 2023).

2.1 | Estimation and carry forward of small area data

Disruption to data collection and processing meant that as of March 2023, small area statistics were available only for England and Wales. National Scottish estimates are not due to be published until summer 2023, while the precise release date of small area estimates in 2024 is not yet known. Moreover, although the Census in Northern Ireland took place at the same time as in England and Wales, NISRA estimates for local government districts (LGDs) were the most granular outputs available as of spring 2023. These circumstances challenge the requirement for a timely geodemographic

classification applicable to the entire UK. Were the classification built only using England and Wales data, the resulting taxonomy would likely not be extensible to other parts of the UK, given their unique mixes of built environment and social conditions. Yet the value of geodemographics requires that a working classification is made available in a timely manner following release of small area statistics.

Despite the increased importance and policy engagement of devolved UK administrations, the UK remains a heavily centralised state with requirements for UK-wide benchmarking of social and physical conditions. Census data, and derivative classifications, are integral to fulfilling these requirements, but timeliness of raw and derived statistics is integral to the value proposition, since local conditions drift further from census estimates with elapsed time since census night.

Accordingly, the decision was made that the new classification would be built in a way that ensures a viable and contemporary solution for England and Wales that would also be extensible to the rest of the UK when further small area data for Northern Ireland and Scotland are published. Although Northern Ireland small area data would become available much sooner than 2022 Scottish data, no provisions were made for any interim update prior to availability of all UK data, in order to avoid multiple versioning of the classification and resulting confusion amongst prospective users. We utilise 2011 Census statistics and small area geographies to anticipate the recent neighbourhood population structures of Northern Ireland and Scotland. It is reasonable to anticipate stability of some social conditions since 2011 (e.g., those relating to housing structure, education level, religion), and predictability of others (e.g., incrementing mean adult resident ages in settled neighbourhoods), meriting inclusion of 2011 estimates where 2022 data are not yet available. Spatial and temporal inference were thus used to create geodemographic place-markers for small area data for Northern Ireland and Scotland, for replacement with observed data when these are both published.

Singleton et al. (2016) developed a change geodemographic by pooling 2001 and 2011 UK Census data using a consistent small area geography. They report that, between 2001 and 2011, 46,078 of the 171,372 2011 Output Areas (26.89%) were reassigned between clusters, 39,444 (85.6%) of which lay within urban areas as defined by ONS. They also identify that the pattern of cluster reassignment is uneven between clusters, and that the study period was characterised by greatest change in suburban areas surrounding large conurbations. Although their results are specific to the period 2001–11, the magnitude of change indicates that it would be unwise to rely upon the simulated results for Scotland in any capacity beyond providing place-markers for the observed data when they are published.

Prior to imputation of the 2021 data, it was first necessary to ensure consistency and comparability with available 2011 data. To this end, variables from the 2011 Census were manually matched with 2021 Census categories. Although the set of questions asked in the 2011 Census was not substantially different in 2021/2, there have been changes to the categorisation of the results. Specifically, while the 2021 Census provides better scrutiny on some of the domains (e.g., more categories in the legal partnership status, including a breakdown of same-sex relationships), it is also more restrictive with respect to age bands and country of birth. It is thus essential to aggregate some categories so that the 2011 and 2021 results are comparable. Such aggregation is not deemed to adversely affect the quality of the subsequent classification, since the more granular categories are not material to major differences in social structure.

The nomenclature of Northern Ireland's small area census geography has changed from small areas (SAs) in 2011 to data zones (DZs) in 2021, but small area data for the latter are not yet available. Each of these geographies nests within LGDs. We simulate the incidence of each 2021 Census variable at small area level by apportioning 2021 LGD totals to 2011 SAs in the proportions in which it was distributed between the SAs of each LGD in 2011. More sophisticated approaches to this problem are available: for example, Bracken and Martin (1989) use subjectively defined population-weighted census enumeration district centroids and inter-centroid distances to model within-zone distributions of census variables, while Lloyd (2010) uses grids to harmonise disparate census geographies. Such solutions permit more sophisticated, if unverifiable, modelling of local spatial distributions, but in geodemographic context may enable at best incremental improvement set against the shifting sands of changing population characteristics:

The standardised change in the score for each LGD i between 2011 and 2021 is first calculated as:

$$C_i = \frac{x_{i2021} - x_{i2011}}{x_{i2011}}$$

where C_i denotes the standardised change in LGD i, and x_i is the count of a given variable. To estimate the 2021 proportions at the SA level, the 2011 standardised values are multiplied by the relative C_i:

$$\mathrm{SA}_{i\ 2021} = \mathrm{SA}_{i2011} \times \left(1 + C_i\right)$$

where SA_i is a modelled value of x_i at the 2011 SA level. Although the method provides a good measure of the expected proportions for 2021 for the neighbourhood geography, the summary value is vulnerable to extreme values, particularly if the 2011 value is equal to, or is close to, zero. Thus where 2011 SA proportions were equal to 0 this value was carried forward to 2021 irrespective of any observed overall increase at the LGD level by 2021. Highly homogenous areas with very high 2011 proportions could also attain modelled 2021 proportions in excess of 1, but were capped at this value. Following these adjustments, modelled SA totals may not sum exactly to the observed 2021 LGD totals. Minor inconsistencies were not, however, observed to detract from the overall quality of the modelled SA proportions.

Not all variables for which 2021 LGD counts were available could be modelled. Specifically, although the 2021 LGD counts were released for ethnic groups, NISRA no longer provides information on the proportions of Bangladeshis—presumably because just 580 individuals in Northern Ireland identified with this group in the 2011 Census. We therefore carried forward the 2011 proportions of Bangladeshis at the SA level.

Given the probable non-availability of full 2022 Scottish Census small area data until 2024, it was necessary to use small area data from the 2011 Census so that the clustering process would instate distinctively Scottish placeholders in the UK-wide classification. The intention is that these modelled 2021 estimates may be replaced with observed 2022 small area data when these become available, with the realistic expectation that they will fit the overall classification, albeit that some DZs—such as areas of rapid redevelopment or new building—may be assigned to different clusters. Age bands deemed likely to characterise sedentary residents in ageing neighbourhoods were incremented from their 2011 values: residents aged 15–24 were modelled to 25–44; those aged 35–54 were modelled to 45–64; those aged 55–74 were modelled to 65–84; and those aged 75–84 were modelled to age 85 and above. Those aged 85 and above in 2011 were deemed deceased in 2021. Two other age bands (0–4 and 5–14) were carried forward without incrementation in order to retain these age cohorts in child-rearing neighbourhoods. All other small area demographic and built environment characteristics in 2011 were carried forward.

The credibility of these assumptions was verified by comparing 2011 and 2021 neighbourhood profiles for England and Wales, where both are available. Moderate positive correlations were observed for the youngest age cohorts and, with the exception of the youngest adults, these correlations increase steadily and cumulatively towards the oldest age groups. Older residents thus appear more likely to age with their neighbourhoods.

2.2 | Variable selection

Successive OACs (Gale et al., 2016; Vickers & Rees, 2007) have been motivated by the challenge to provide a shorthand summary of the full range of census data in the form of a general purpose classification that supports a range of policy applications. The UK Census has a statutory remit under the 1920 Census Act to collect data on demographics (specifically name, sex and age), employment (occupation, profession, trade or employment), ethnicity and origins (nationality, race, birthplace, language), usual residence (place of abode, character of dwelling), living arrangements (condition as to marriage, relation to head of family, issue born in marriage) and a catch-all 'any other matter with respect to which it is desirable to obtain statistical information with a view to ascertaining the social or civil conditions of the population' (Dewdney, 1981). In practice, the catch-all clause has often been used by successive governments to reflect their concerns and preoccupations at the time of data collection (The Economist, 2021), and the remit of general purpose geodemographic classifications has extended to encompass many of the 'catch-all' topics that have been added since the 1921 Census.

However, the circumstances of the pandemic are such that it is expedient to pare back some of these additional topics (e.g., travel-to-work or short-term residential relocation) and revise others (e.g., replacing specific occupations that might have been furloughed with more general employment characteristics). Still other characteristics might highlight neighbourhood differentiation (e.g., secular decline in religious affiliation, or demographic ageing). Our intention was thus to pare back inclusion of Census questions likely to elicit pandemic-dependent circumstances, while foregrounding data relating to secular trends in society likely to underpin residential differentiation. A related point is that although the remit and content of the most recent Census is fundamentally unchanged from its recent forbears, the content of small area statistics has been amended in some instances—specifically, the emphasis upon employment rather than occupational structure.

The overall selection of variables was guided by the unique circumstances of the COVID-19 pandemic and its aftermath, cognisant that inclusion of circumstances that have changed since 2021 (England, Wales and Northern Ireland) or 2022 (Scotland) might limit the applicability of the classification. Variable selection was otherwise guided by the 2011

OAC procedures. Adjustments were also made to accommodate changes in population baselines—as, for example, with definitions of full-time employment and the provision of unpaid care.

The complexity of geodemographic classifications in terms of numbers of variables and statistical transformations has tended to increase over time—for example, the 2001 OAC used 41 variables while 2011 OAC used 60 (Gale et al., 2016). This might be seen as mirroring the increasing complexity of societal organisation across neighbourhoods. However, given that the pandemic renders use of some 2021 and 2022 Census data redundant to applications of the classification, we anticipated consolidation rather than any further escalation of the number of input variables used. Some previous classifications have reduced larger numbers of input variables to principal components as inputs (Harris et al., 2005) or have used principal components analysis as integral to automated variable selection (Liu et al., 2019). These approaches were not adopted here because the underlying relationships between input variables may not be linear across space, and there was no compute constraint on reducing a larger number of variables into a set of clusters representing latent structure.

Variable selection was undertaken cognisant of the scope of the available data, the statistical properties of identified measures, the structure of past OACs and the influence of selected variables on potential cluster solutions. To test their performance, variables were examined for multicollinearity and temporal transferability: a strong correlation between two or more variables can unduly bias the cluster assignment. Consistent with the 2011 OAC, 'weak' correlation was defined as in the interval 0–0.4, 'moderate' as lying between 0.41 and 0. 65, 'strong' as between 0.66 and 0.8 and 'very strong' as between 0.81 and 1. Final selection of variables was guided by avoiding strong and very strong correlations and maintaining representation of all OAC variable domains.

Table 1 lists 60 of the 188 candidate variables, and associated census domains as set out above, that were selected for the 2021/2 OAC—coincidentally the same number of variables featured in the 2011 OAC. The most significant change in the variable specification when compared with the 2011 OAC was inclusion of residential mobility/migration and religious observance, and replacement of industry variables used in 2011 with 2021 occupation statistics. The lattermost of these reflected a change in what was initially available in 2021/2 small area statistics, while the first two were included in recognition of mobility and accelerating secularisation of society.

The process of using 2011 Census data as place markers for Scottish data assumed that broad patterning of 2021/2 data follows that observed in 2011. This assumption was tested by correlating the known Output Area data for each of the 60 variables in England and Wales (Figure 1). The variables relating to accommodation type and tenure had very strong correlations. The lowest correlations are between observations for intermediate- and low-skilled occupations, perhaps because of revisions to the detail of their definition, as well as changes in the employment structure of the population. Although no correlation thresholds were set, the overall pattern was deemed supportive of use of 2011 Scottish data as classification place-markers.

Other aspects of the methodology remain essentially unchanged from the 2011 OAC. Prior to the cluster analysis, data were range standardised and transformed using the same manipulation procedures: data were first converted to percentages (to control for non-uniformity of areal units), then transformed using an inverse hyperbolic sine function to account for the non-normality of data distribution and then, finally, were range standardised.

2.3 | Clustering methodology

The methodology utilised for the creation of the 2021/2 OAC is based on the successful previous work of Gale et al. (2016). The method of classification retained the widely implemented technique of *k*-means clustering and was performed in the R statistical software. Following OAC 2011, the method was initiated through assignment of a set of *k* randomly generated seed locations within a multidimensional data array comprising a row for every small area in the UK, and a column for each of the selected standardised variables. The *k*-means algorithm allocates areas to their nearest seed locations to create an assignment of initial clusters. Means of the clusters are then computed, and these centroid locations are then used to re-assign each area to its nearest cluster centroid. This process of mean cluster centroid calculation and reassignment continues until no further reassignments occur. The statistical quality of the final assignment at the end of this iterative process was assessed by calculating a total within sum of squares statistic, which is composed of the summed squared Euclidean distance between each area and its assigned cluster centroid. This total within sum of squares statistic can be used to compare cluster solutions, with the most optimal solution having a smaller total within cluster sum of squares. The quality of a solution is stochastic and depends on the specific location of the initial seeds. As such, and also following OAC 2011, multiple *K*-means runs were implemented, using total within sum of squares to select an optimal result from 10,000 runs of the model.

TABLE 1 Variables selected for the 2021/2 OAC and associated domains from the 1920 Census Act.

No.	Variable name	Domain
v01	Usual residents per square kilometre	Demographic
v02	Aged 4 years and under	
v03	Aged 5–14 years	
v04	Aged 25–44 years	
v05	Aged 45–64 years	
v06	Aged 65–84 years	
v07	Aged 85 years and over	
v08	Country of birth: Europe: United Kingdom	Ethnicity and origins
v09	Country of birth: Europe: EU countries	
v10	Country of birth: Europe: Non-EU countries	
v11	Country of birth: Africa	
v12	Ethnic group: Bangladeshi	
v13	Ethnic group: Chinese	
v14	Ethnic group: Indian	
v15	Ethnic group: Pakistani	
v16	Ethnic group: Other Asian	
v17	Ethnic group: Black	
v18	Ethnic group: Mixed or Multiple ethnic groups	
v19	Ethnic group: White	
v20	Cannot speak English well or at all	
v21	No religion	
v22	Christian	
v23	Other religion	
v24	Never married and never registered a civil partnership	Living arrangements
v25	Married or in a registered civil partnership	
v26	Separated or divorced	
v27	One-person household	
v28	Families with no children	
v29	Families with dependent children	
v30	All household members have the same ethnic group	
v31	Lives in a communal establishment	Usual residence
v32	Address 1 year ago is the same as the address of enumeration	
v33	Detached house or bungalow	
v34	Semi-detached house or bungalow	
v35	Terraced (including end-terrace) house or bungalow	
v36	Flat, maisonette or apartment	
v37	Ownership or shared ownership	
v38	Social rented	
v39	Private rented	
v40	Occupancy rating of rooms: +1 or more	
v41	Occupancy rating of rooms: −1 or less	

TABLE 1 (Continued)

No.	Variable name	Domain
v42	Standardised Illness Ratio	Other
v43	Provides unpaid care	
v44	2 or more cars or vans in household	
v45	Highest level of qualification: Level 1, 2 or Apprenticeship	
v46	Highest level of qualification: Level 3 qualifications	
v47	Highest level of qualification: Level 4 qualifications or above	
v48	Hours worked: Part-time	Employment
v49	Hours worked: Full-time	
v50	NS-SeC: L15 Full-time students	
v51	SOC: 1. Managers, directors, and senior officials	
v52	SOC: 2. Professional occupations	
v53	SOC: 3. Associate professional and technical occupations	
v54	SOC: 4. Administrative and secretarial occupations	
v55	SOC: 5. Skilled trades occupations	
v56	SOC: 6. Caring, leisure, and other service occupations	
v57	SOC: 7. Sales and customer service occupations	
v58	SOC: 8. Process, plant and machine operatives	
v59	SOC: 9. Elementary occupations	
v60	Economically active: Unemployed	

An important decision in any geodemographic classification is choice of the number of clusters that best represents the latent structure of the collection of selected input measures. Following OAC 2011 a top-down clustering approach was adopted, identifying the most aggregate clusters first, then using these to split the input data to create further tiers of a hierarchy. As with both previous OAC solutions, three tiers were created, consisting of Supergroups, Groups and Subgroups, with Subgroups nested within Groups that in turn are nested within Supergroups. An innovation relative to the previous OAC methodology was the use of clustergrams (Fleischmann, 2023) to inform the choice of an optimal number of partitions for each level of hierarchy. This graphical technique plots a weighted mean of the first component of principal components analysis (PCA) for each individual area across a range of different potential k solutions (Singleton et al., 2022).

Given that the classification is intended to serve a range of resource allocation decisions, it was vital to design a product that could meet specific end user requirements. For this reason, an expert panel including local authority end users and technical experts from ONS and the Greater London Authority (GLA), was convened to evaluate the empirical results and agree provisional cluster names and pen portraits. Initial consultations considered choices of variables and the outline clustering methodology—including the strategy to accommodate Scottish and Northern Irish data when available. Detailed group discussion then addressed: the numbers of Supergroups, Groups and Subgroups; the provisional naming of the clusters; and the detailed pen portraits that would describe the classification. Further waves of consultation were undertaken following completion of the classification, through commentary by ONS topic leads on: Demography; Ethnicity, National Identity, Language & Religion; Health and Disability; Housing; Migration; and Unpaid Care.

3 | RESULTS

3.1 | Clustergram solutions and cluster geography

Figure 2 presents the clustergram for the first partitioning of the input data for the range of k values from 1 to 11. Clustergrams provide a useful diagnostic tool for selecting k means cluster frequency (Fleischmann, 2023). For each potential cluster solution (x axis), the y axis plots each cluster's weighted mean (red dots) of the first component of a PCA that is applied to the input

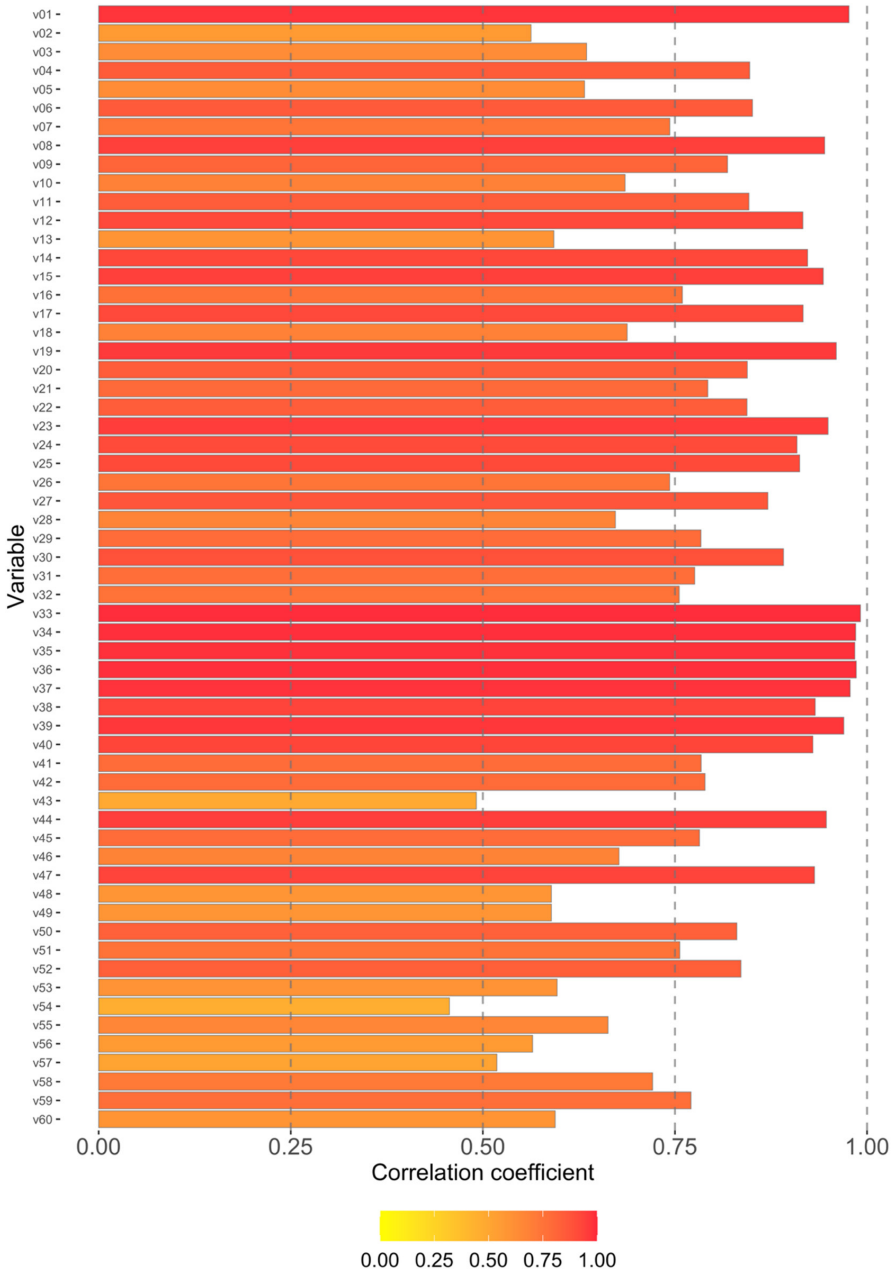

FIGURE 1 Pearson correlation coefficients for the selected variables between 2011 and 2021 at the Output Area level in England and Wales (for key to variables, see Table 1).

dataset. An optimal solution is deemed to be that in which clusters show greater separation on the *y* axis, identifying that they have substantially different weighted PCA scores. A feature of clustergrams is that they enable identification of the areas assigned to clusters under different *k* cluster solutions. For smaller values of *k*, lines are thicker as all areas are aggregated into a smaller number of clusters. Visual inspection along the *x* axis allows identification of clusters which are more stable, since they appear more consistently across successive *k* values.

Interpretation of clustergrams to select a value of *k* are therefore informed empirically, but are ultimately a qualitative choice by the classification builder. A more conventional elbow plot (Gale et al., 2016) was also produced, but no clear break point was discernible. Solutions involving between six and nine clusters were mapped, and the distinctiveness of resulting residential structure noted, along with statistical properties such as relative sizes of the clusters. Following this, the eight cluster solution was confirmed as the most appropriate solution.

Clustergrams were then developed, first to partition each Supergroup into a parsimonious number of Groups, and then to partition each Group into Subgroups. Illustrative clustergrams are presented in Figure 3. Where the best solutions

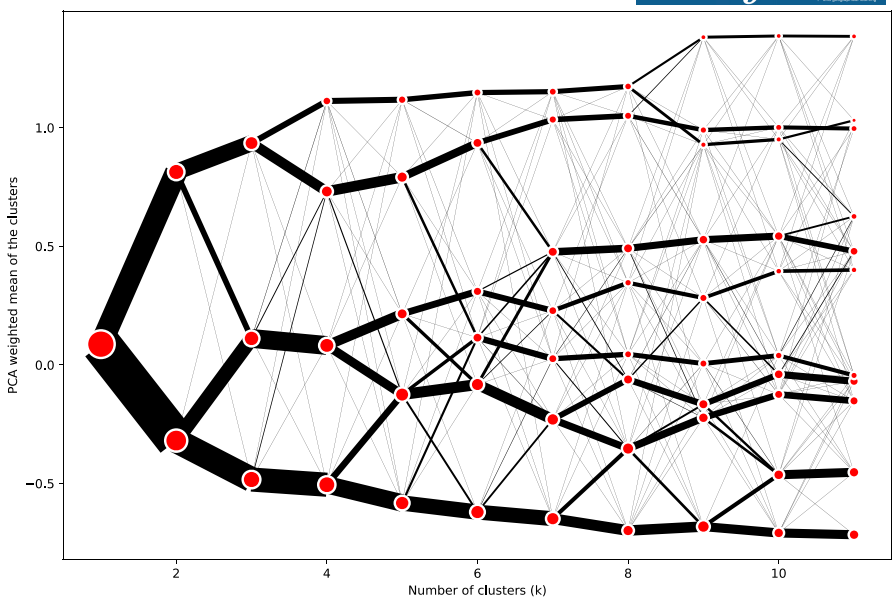

FIGURE 2 Clustergram for 1 to 11 cluster (Supergroups) solutions.

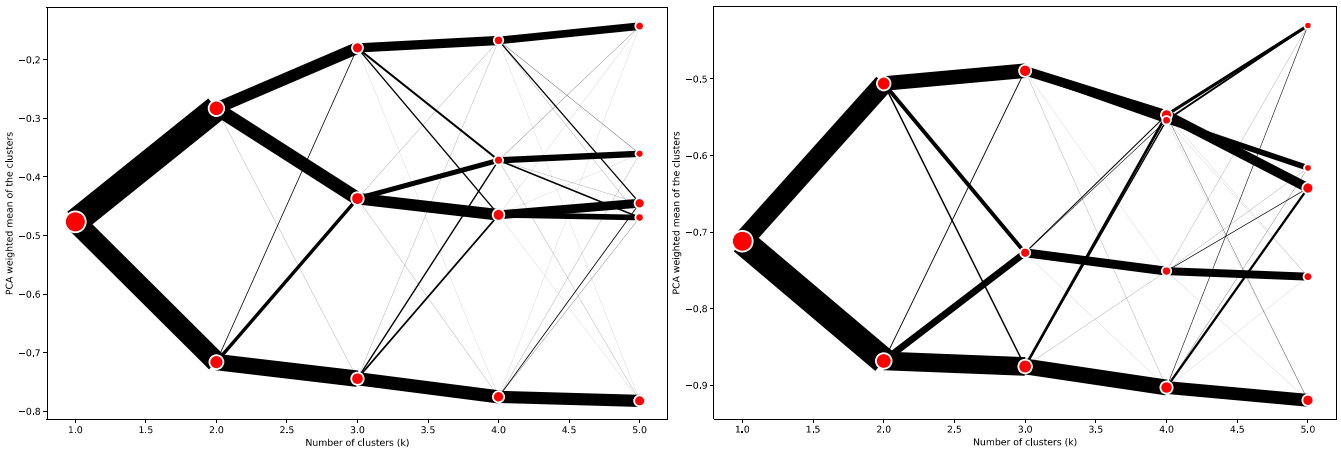

FIGURE 3 Illustrative clustergrams for 1 to 5 cluster solutions for Supergroup 1 and for Group 1b.

were not immediately apparent, we opted for higher numbers of Groups and lower numbers of Subgroups, as experience has shown that most applications are developed at the Supergroup and Group levels.

The eight cluster Supergroup solution is mapped in Figure 4a, using both the observed 2021 data for England and Wales and modelled data for Scotland and Northern Ireland. The distribution of these Supergroups in London is shown in Figure 4b. Although there is distinctiveness in allocations between Supergroups within England, Northern Ireland, Scotland and Wales, there is little to suggest that the national outcomes are systematically determined by use of simulated or observed data.

Although the initial formal ONS public release of UK OAC only comprises lookups for England and Wales, the advantage of including both Northern Ireland and Scotland within the model is that the assignment of cluster centroids is informed by the UK rather than England and Wales average values across the selected measures. It is anticipated that this procedure will enable the inclusion of Northern Ireland and Scotland lookups through either the direct assignment of areas to cluster centroids, or the reclustering of the UK national data from all constituent countries, when these data are available. The precise updating methodology should only be finalised after the Scottish and Northern Irish small area data are available and the impact of using modelled estimates for Scotland and Northern Ireland can be fully evaluated.

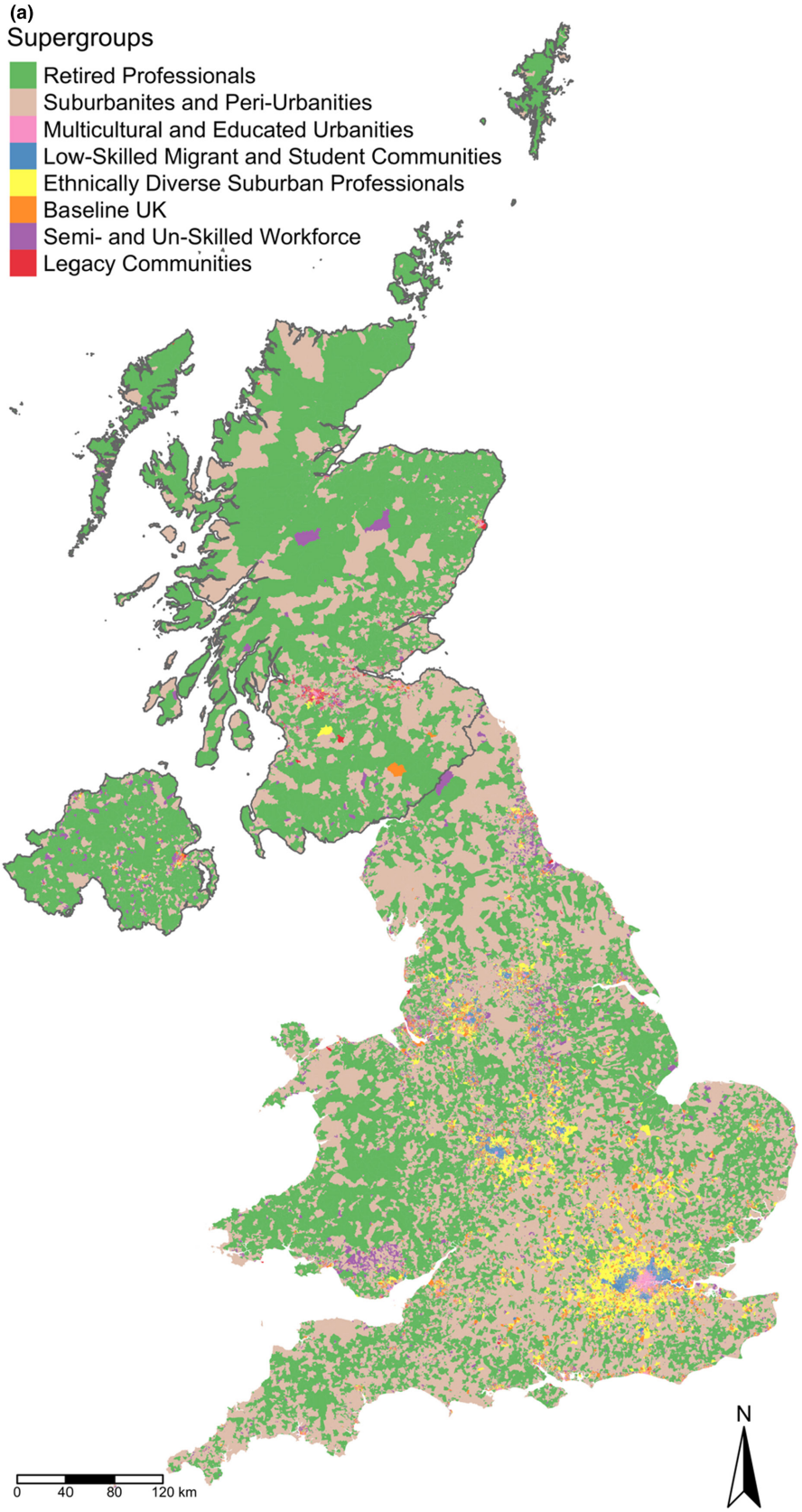

FIGURE 4 Cluster Supergroup distribution of 2021/2 OAC: (a) across the UK (including simulated Scottish and Northern Irish results); and (b) within Greater London (Supergroup labels as defined in Section 3).

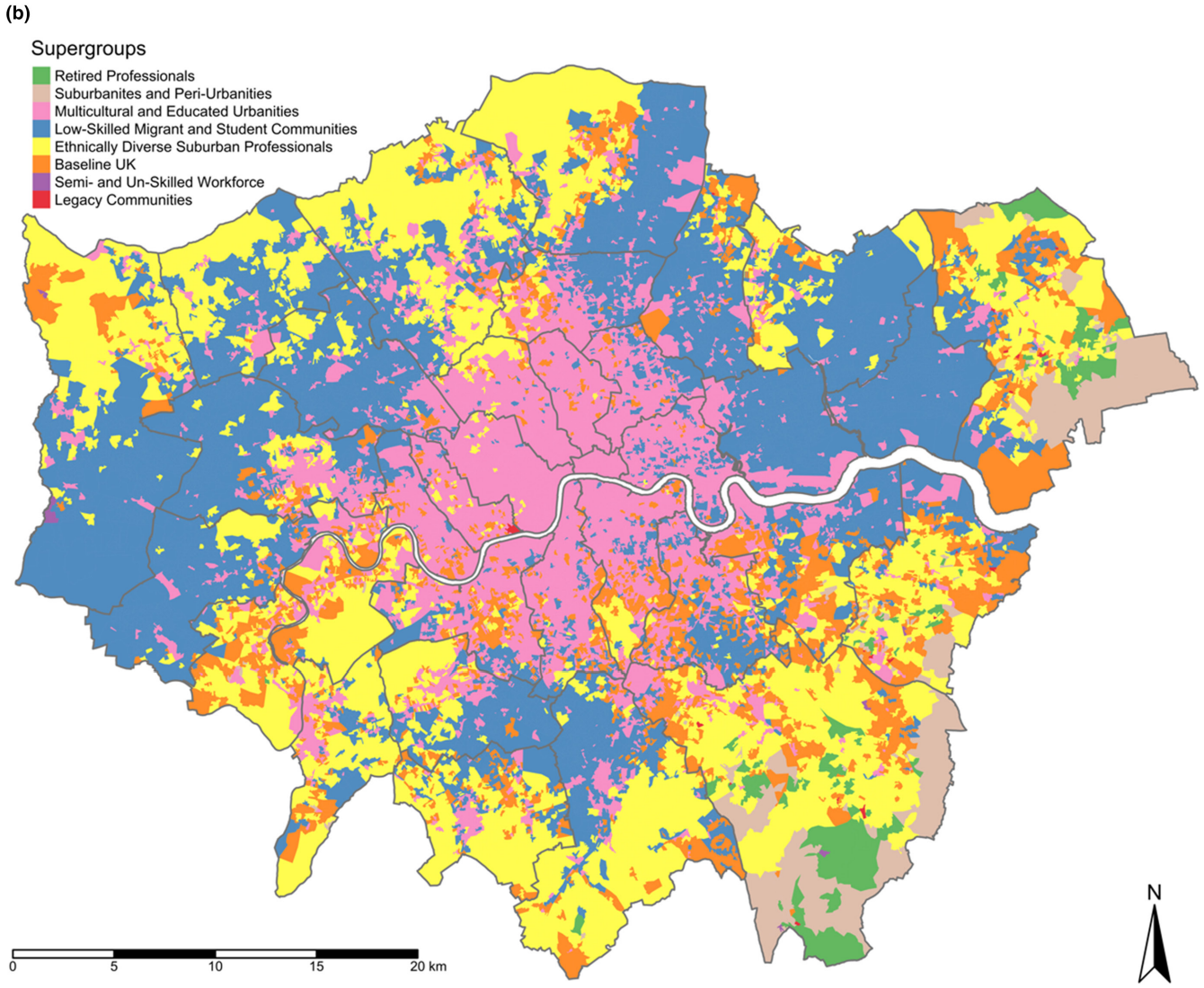

FIGURE 4 (Continued)

3.2 | Cluster labelling and pen portraits

The process of interpreting, naming and describing clusters was conducted by a panel of academic experts and end users from the ONS and local government. Much of the description was informed by comparing cluster mean values of variables with the UK mean (for Supergroups), the Supergroup mean (for Groups) and Group mean (for Subgroups). Cluster names were developed in consultation with the Advisory Group, taking particular account of the requirements of the UK's national statistical agency (Office for National Statistics, 2015).

Cluster names and descriptions were subjected to an additional review by ONS to ensure impartiality and appropriateness. The resulting taxonomy is presented in Table 2. Each of the Supergroup cluster labels was accompanied by a 'pen portrait' description and descriptions of the Groups and Subgroups that the Supergroup contains. As an illustration, Table 3 shows the pen portraits for Supergroup 2, one of the Groups within it (Group 2a) and one of the Subgroups within the Group (Subgroup 2a1).

4 | COMPARING OAC 2021 AND 2011 IN ENGLAND AND WALES

The motivation for this work is to develop a general-purpose geodemographic classification that extends the well established tradition of Open UK-wide segmentations published following successive censuses (Gale et al., 2016;

TABLE 2 The 2021/2 OAC taxonomy, with the percentage of UK small areal units assigned to each cluster.

Supergroup	Group	Subgroup
1—Retired Professionals (17.7%)	1a—Spacious Rural Living (6.55%)	1a1—Pre-Retirement Spacious Living (4.37%)
		1a2—Retirement Spacious Living (2.19%)
	1b—Small Town Suburbia (6.35%)	1b1—Younger Established Suburban Communities (3.61%)
		1b2—Older Established Suburban Communities (2.74%)
	1c—Established Mature Families (4.85%)	1c1—Affluent Mature Families (2.97%)
		1c2—Burgeoning Mature Families (1.88%)
2—Suburbanites and Peri-Urbanities (20.3%)	2a—Inner Suburbs and Small Town Living (5.99%)	2a1—Younger Suburban Family Renters (1.94%)
		2a2—Settled Owner-Occupied Suburbs (2.35%)
		2a3—Terraced Communities (1.69%)
	2b—Rural Amenity (7.43%)	2b1—Ageing Rural Communities (4.2%)
		2b2—Rural Mix (3.23%)
	2c—Ageing Communities (6.92%)	2c1—Communal Retirement Living (1.32%)
		2c2—Ageing Independent Living (5.6%)
3—Multicultural and Educated Urbanites (8.24%)	3a—Student Living and Professional Footholds (2.26%)	3a1—University Centric (0.37%)
		3a2—Professional Progression (0.47%)
		3a3—Urbanite Mix (0.78%)
		3a4—Affluent Graduate Living (0.64%)
	3b—Ethnically Diverse Young Families (2.8%)	3b1—Private Rental Ethnic Minority Families (1.65%)
		3b2—Young Ethnic Minority Families (1.15%)
	3c—Diverse Educated Urban Singles (3.17%)	3c1—Centrally Located Professionals (0.95%)
		3c2—Career Progression (2.23%)
4—Low-Skilled Migrant and Student Communities (9.14%)	4a—Ethnically Diverse Families in Less Connected Locations (3.47%)	4a1—Semi Detached, Service Workers and Students (1.19%)
		4a2—City Service Workers (1.38%)
		4a3—Multi-Child Young Families (0.9%)
	4b—Established Multi-Ethnic Communities (3.62%)	4b1—Multi-Generational Migrants (0.77%)
		4b2—European Skilled Workforce (1.2%)
		4b3—Inner Suburb Ethnic Group Mix (0.76%)
		4b4—Ethnic Minority Routine Service Workers (0.89%)
	4c—Challenged Multicultural Communities and Students (2.05%)	4c1—African and Asian Influences (1%)
		4c2—European or Asian Heritage (1.05%)

TABLE 2 (Continued)

Supergroup	Group	Subgroup
5—Ethnically Diverse Suburban Professionals (8.75%)	5a—Outer Suburbs (3.97%)	5a1—Outer Suburb Asian Mix (1.18%)
		5a2—Suburban Empty Nesters (1.24%)
		5a3—Young Suburban Families (1.55%)
	5b—Suburban Professionals (4.78%)	5b1—Families in Multi-Ethnic Terraces (2.59%)
		5b2—Established Multi-Ethnic Suburbs (2.19%)
6—Baseline UK (13.8%)	6a—Challenged Communities (5.68%)	6a1—Suburban Housing Starters (1.43%)
		6a2—Semi-Detached Strivers (1.98%)
		6a3—Younger Ethnic Minority Families in Flats (2.28%)
	6b—Legacy Industrial and Coastal Communities (4.46%)	6b1—Retired Seniors (1.88%)
		6b2—Traditional Terraces (1.33%)
		6b3—EU Singles (1.25%)
	6c—Multicultural Inner Suburbs (3.67%)	6c1—Transient Communities (1.36%)
		6c2—Semi-Detached Family Renters (2.32%)
7—Semi- and Un-Skilled Workforce (14.3%)	7a—Established but Challenged (8.29%)	7a1—Ageing Established Urban Communities (2.92%)
		7a2—Industry Associations (5.38%)
	7b—Young Families in Industrial Towns (5.96%)	7b1—Terraces in Transitional Towns (3.28%)
		7b2—Families and Later Life (2.69%)
8—Legacy Communities (7.71%)	8a—Routine Occupations or Retirement (2.42%)	8a1—Retirement Residences (0.52%)
		8a2—Flats and Routine Occupations (1.9%)
	8b—Legacy and Demographically Mixed Communities (5.29%)	8b1—Challenged Families (2.74%)
		8b2—Retirement Pockets (1.51%)
		8b3—Young Families and Neighbourhood Turnover (1.04%)

TABLE 3 Illustrative 'pen portrait' of one of the OAC 2021/2 Supergroups, typical occurrences, and the Groups and Subgroups it contains. (This Supergroup comprises a total of three Groups and six Subgroups, but only one of each is shown here.)

2: Suburbanites and Peri-Urbanites

Pervasive throughout the UK, members of this Supergroup typically own (or are buying) their detached, semi-detached or terraced homes. They are also typically educated to A-Level/Highers or degree level and work in skilled or professional occupations. Typically born in the UK, some families have children, although the median adult age is above 45 and some property has become under-occupied after children have left home. This Supergroup is pervasive not only in suburban locations, but also in neighbourhoods at or beyond the edge of cities that adjoin rural parts of the country.

2a: Inner Suburbs and Small Town Living

Families with resident dependent children (but not students) are common. Long-established family groups and White ethnicity predominates, comprised of individuals born in the UK. They are more likely than the Supergroup average to have been resident in their terraced, semi-detached, or detached houses for more than 1 year. Levels of multiple car ownership are high. Properties are owned and typically have surplus living space. Associate professionals and administrative occupations are prevalent, and parents are likely to be in middle age or approaching retirement. Educational attainment is above the Supergroup average. Scattered developments and concentrations are found in many small towns.

2a1: Younger Suburban Family Renters

This Subgroup is characterised by a higher proportion of families with dependent children (especially aged 0 to 4) than the Group averages. Young adults predominate, especially over retirees. Many households rent within the private sector, are well qualified to A-Level/Highers or degree level, and work in professional, technical and administrative occupations.

Examples can be found in:

England: Killingworth, Spalding (outskirts)

Wales: Llandudno Junction

Singleton & Longley, 2015; Vickers & Rees, 2007). Although the remit and content of the most recent Census is fundamentally unchanged from its recent forbears, changes in the portfolio of statistics released at small area level influence the cluster solutions. In drawing out these comparisons we refer to '2021 OAC' when referring to observed small area data for 2021 and '2021/2 OAC' when referring to the taxonomy derived using additional 2011 Census data and boundaries. Table 4 presents the 2011 OAC decomposed into 2021/2 OAC Supergroups. The principal amendments to the classification are:

1. Stronger Supergroup concern with labour force participation in an ageing society—the Retired Professionals are prevalent in neighbourhoods populated in 2011 by Rural Residents and Suburbanites.
2. Greater uniformity of geodemographic characteristics observed across some neighbourhoods characterised in 2011 as Urbanite, Suburbanite or Rural. These neighbourhoods are now mainly assigned to Suburbanites and Peri-Urbanites.
3. 2021 Multicultural and Educated Urbanites neighbourhoods were largely classified as Cosmopolitan or Ethnicity Central in 2011. This outcome suggests new foregrounding of educational and occupational outcomes over ethnicity in the core areas of most metropolitan areas in England and Wales.
4. 2021 Low-Skilled Migrant and Student Communities provide a mix of individuals with experiences of education that tend towards polar extremes. These neighbourhoods were principally Multicultural Metropolitan with some Ethnicity Central in 2011, but not drawn from the 2011 Supergroups that were the most economically challenged.
5. The highly dispersed 2021 Supergroup Ethnically Diverse Suburban Professionals, indicates the spread of ethnic minorities across the UK geodemographic structure (see also Catney et al., 2023).
6. The modal census variable values of 2021 Census characteristics embodied in Baseline UK are observed across the four most economically disadvantaged 2011 Supergroups. By contrast, the disadvantage experienced in 2021 Semi- and Un-Skilled Workforce and Legacy Communities maps much more directly into 2011 Hard-Pressed Living and Constrained City Dwellers, respectively.

We have created an interactive website to enable users to examine the new 2021/2 OAC and to compare it with the 2011 classification. This is available at mapmaker.cdrc.ac.uk/#/output-area-classification-2021. The map allows users to pan across all of England and Wales and to search for the classification of specific postcodes. Figure 5 presents four very different geodemographic extracts from this web resource, showing (a) the city conurbation of Manchester, (b) the market town of King's Lynn, (c) the rural settlement system between Wendover and Princes Risborough, and (d) part of the West Midlands conurbation around Northfield. The apparent differences between 2011 and 2021 manifest both changes in the spatial organisation of society and changes in the melange of census variables that we choose to represent them. Some changes raised by these extracts that are also apparent from the website are:

TABLE 4 Confusion matrix showing assignment of 2011 OAC Supergroups to 2021/2 OAC Supergroups.

	2011 OAC							
	Rural residents	**Cosmopolitans**	**Ethnicity central**	**Multicultural metropolitans**	**Urbanites**	**Suburbanites**	**Constrained City dwellers**	**Hard-pressed living**
2021/2 OAC								
Retired Professionals	28.62%	0.01%	0%	0%	1.4%	69.38%	0.02%	0.57%
Suburbanites and Peri-Urbanities	30.19%	0.08%	0.01%	0.02%	36.13%	19.82%	0.34%	13.42%
Multicultural and Educated Urbanites	0.02%	51.61%	41.73%	2.92%	1.04%	0.05%	2.6%	0.04%
Low-Skilled Migrant and Student Communities	0.04%	1.71%	16.77%	74.5%	1.38%	0.11%	2.93%	2.57%
Ethnically Diverse Suburban Professionals	3%	1.44%	0.02%	12.29%	40.93%	39.25%	0.23%	2.84%
Baseline UK	0.73%	6.13%	1.8%	15.46%	32.14%	0.61%	20.43%	22.69%
Semi- and Un-Skilled Workforce	1.88%	0.01%	0%	0.06%	3.68%	0.56%	9.24%	84.57%
Legacy Communities	0.39%	4.46%	0.97%	0.08%	5.68%	0.14%	86.95%	1.33%

1. Sharper definition of urban centres. The 2021 Multicultural and Educated Urbanites are more focused on city and town centres, and more towns exhibit distinctive central cores identified by this Supergroup. These areas of urban living were split between Cosmopolitans and Ethnicity Central in 2011, as illustrated for Manchester and its surrounding towns in Figure 5a. Manchester itself, Salford and Bolton had identifiable single Supergroup urban cores in 2011, but additional cores are apparent in 2021 in the surrounding, smaller urban areas of Stockport, Ashton-under-Lyne, Bury and Oldham. In part, this may manifest increasing studentification of central, ethnically diverse urban areas.
2. Greater differentiation within urban areas. In 2021, many towns include the complete, or near complete, set of eight supergroups—no longer are some Supergroups only generally seen outside urban areas (e.g., the 2011 Rural Residents Supergroup) or within larger urban areas (e.g., the 2011 Ethnicity Central Supergroup). The replacement of a distinctively rural 2011 Supergroup reflects a greater 2021 focus upon occupation rather than industry, while reduced association of ethnic minorities with the centres of major conurbations manifests diffusion of ethnic minorities and recent European migrants into small towns across Britain. This is illustrated by King's Lynn in Figure 5b.
3. Greater differentiation within suburban and outlying urban areas. The 2011 Rural Residents and Suburbanite Supergroups make up almost all of the 2021 Retired Professionals neighbourhoods and, with the addition of 2011 Urbanites, now make up most of the 2021 Suburbanites and Peri-Urbanites Supergroup. This enables greater differentiation of social structure both in small affluent towns and in adjacent countryside. Figure 5c illustrates this with the small towns of Princes Risborough and Wendover, and the small villages and hamlets that separate them. Both the towns and the villages/hamlets show a similar geodemography split across Supergroups.
4. Increasingly variegated urban structure. The nature of residential structure dictates that adjacent Output Areas very frequently share the same Supergroup assignments, but the geographic scale at which this holds appears to be changing. Some areas show greater variation in 2021 across quite small areas, compared with 2011. For example, Northfield in Birmingham (Figure 5d) exhibits an increasingly variegated pattern, suggesting a diminished role of general urban change dynamics relative to local, place-specific considerations.

Table 5 presents a technical comparison of the 2001, 2011 and 2021/2 OACs. The sense is that parsimonious classification has been achieved: the smallest Supergroup is about the same size as that in 2001 and the size range is almost identical, despite there being one fewer Supergroup in 2001, while the 2011 OAC is more unevenly distributed; the 2021/2 Group and Subgroup structure is of identical or lesser complexity; the range in Group size is reduced; and the Subgroup taxonomy is similar in range but lower in complexity than the 2011 OAC. The improvement in parsimony over the 2011 OAC has accrued through more restrictive focus upon variable domains and the use of improved methods (specifically clustergrams) to partition clusters.

A secondary requirement of this work was that the classification be created with a schema and nomenclature that could accommodate small area Census data for Northern Ireland (for 2021) and Scotland (for 2022) when these become available. Although it is recognised that such classification will encompass subpopulations at different times, our expectation is that a consistent UK-wide 2021/2 classification can be framed into the nomenclature set out here and that it will be useful for planning and resource allocation in the post-pandemic world. Undoubtedly, the inclusion of new data and use of changed small area boundaries may greatly affect small area assignments. It may also be that the cluster pen portraits are deemed to require renaming in light of the allocations of 2021 Northern Irish and 2022 Scottish data. We anticipate that the cluster centroids identified in this research will be usable to assign Northern Irish and Scottish data when these are available. However, we also advocate, as a minimum, testing this assumption by reclustering the entire UK small area data to validate that this is tenable, and to validate that UK-wide reclustering does not fundamentally change the cluster allocations already published for England and Wales. Should our expectations not be fulfilled, we would advocate issuing an updated 2021/2 OAC for England and Wales along with supporting metadata.

5 | CONCLUSION

From a statistical perspective, the clusters represent the latent structure of the domains of the classification as represented by the input variables extracted from the Census. The remit of UK censuses is defined by UK central government statute, and our general purpose classification is thus framed by the data arising from the 2021/2 censuses. Our variable selection is informed by correlation analysis, but ultimately is subjective in that other variable specifications, cluster solutions or clustering methods might identify other latent structures. As with any geodemographic classification, empirical

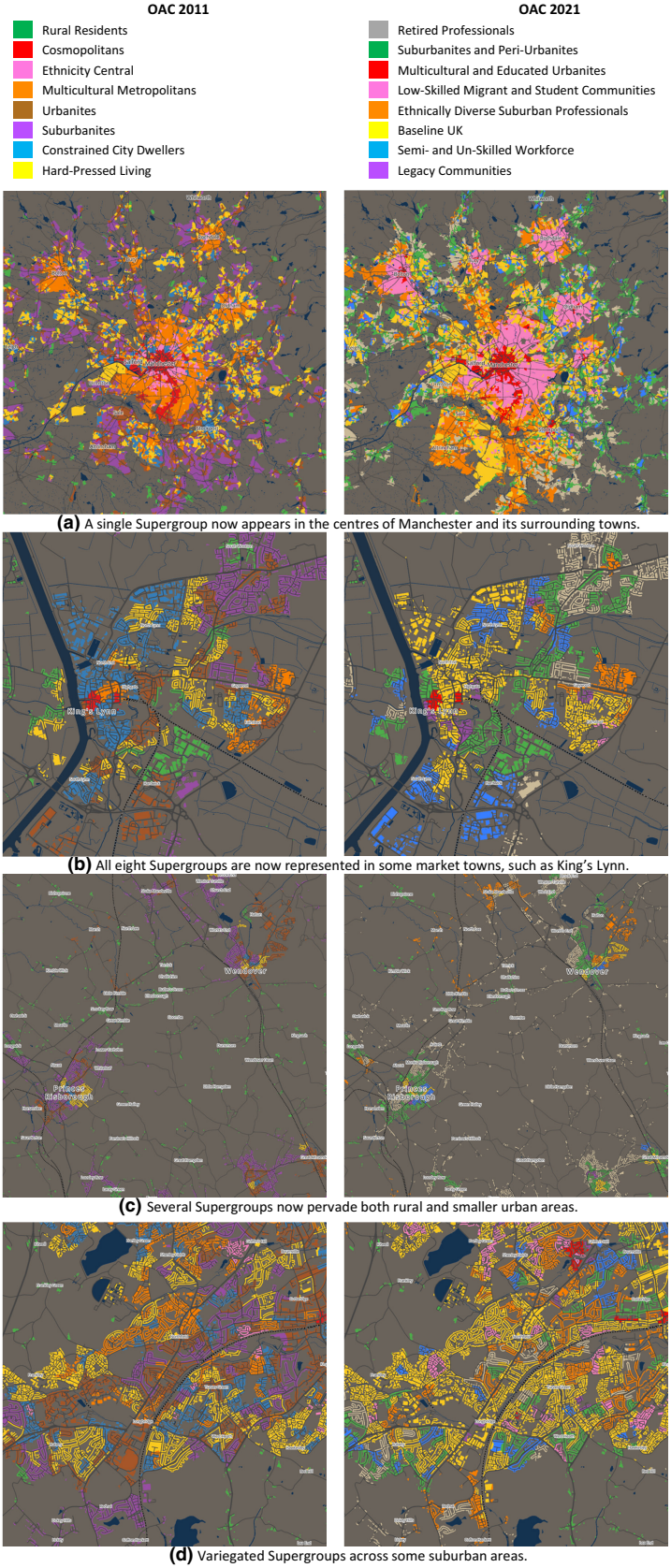

FIGURE 5 Comparisons of OAC 2011 (left) and OAC 21 (right) for: (a) Manchester and surrounding towns; (b) King's Lynn; (c) Wendover and Princes Risborough; and (d) Northfield in Birmingham. The horizontal extent of the maps are 6 km, 40 km, 15 km and 7 km, respectively. The colours used should not be compared across years. Building block/built-up area indications are approximate and include all building types (*Source*: mapmaker.cdrc.ac.uk/#/output-area-classification-2021).

TABLE 5 Comparison of the 2001 OAC, the 2011 OAC and the 2021/2 OAC.

	2001 OAC	2011 OAC	2021/2 OAC
Data source	2001 UK Census	2011 UK Census	2011 and 2021 UK Census
Number of variables	41	60	
Data manipulation techniques	Percentages log10 Range	Percentages Inverse Hyperbolic Sine Range	
Clustering technique	*K*-means with squared Euclidian distance		
Cluster numbers	7 Supergroups 21 Groups 52 Subgroups	8 Supergroups 26 Groups 76 Subgroups	8 Supergroups 21 Groups 52 Subgroups
Range of cluster assignment	7.5%–21.2% 2.3%–8.2% 0.8%–3.3%	5.1%–20.2% 0.7%–11.6% 0.1%–4.2%	7.7%–20.3% 2.1%–8.3% 0.4%–5.4%

evaluation and analytical experience guide choices and conventions in developing what is deemed to be the best solution. The resulting neighbourhood typology is in many respects data driven but also informed by the remit of censuses and experience in interpreting and describing a set of clusters, cognisant of geographical context.

Apparent changes over the period 2011–21 are objective only insofar as the underlying OACs use many of the same variables and that the membership and dispositions of the research teams that created them are likely to have led to choice of similar numbers of clusters. But no geodemographic classification is truly objective. The sizes and compositions of the Supergroups, Groups and Subgroups also dictate that the results are not directly comparable unless data from multiple time periods are pooled into a single classification (see, for example, Singleton et al., 2016). Comparison does nevertheless capture average population characteristics and neighbourhood attributes, and so comparison does chart how the characteristics of a neighbourhood have changed through demographic ageing, residential mobility and migration.

This research has demonstrated the feasibility of accommodating changing geographies, asynchronous data collation and phased data release in devising a UK-wide frame for geodemographic classification. The resulting classification enables timely release of results for England and Wales, and a clear path to inclusion of Northern Irish and Scottish data when these are both available. The remit of geodemographic classifications is typically to provide general purpose shorthand summaries of neighbourhood conditions and, framed by the remit of the 2021 and 2022 censuses, the 2021 England and Wales OAC is no exception. In substantive terms the results manifest secular trends in British society, such as the blurring of urban and rural divides, the diffusion of ethnic minorities across the country and across geodemographic categories, and increasing differentiation of urban and suburban social structure. In important respects, the 2021 OAC also presents a unique 'pandemic geodemographic' in which changes in work patterns are embedded within the observed social structures. Some of these, such as the prevalence and nature of homeworking, may prove more enduring than others, such as patterns of participation in, and attendance of, higher education. The 2021 England and Wales OAC remains a general-purpose classification, albeit with some scaling back of remit because of the unique circumstances prevailing during, and immediately following, the COVID-19 pandemic. The Open methodology has also been utilised by the Greater London Authority to create a geodemographic classification specific to neighbourhood conditions prevailing in London.

In addition to the research required to classify Northern Irish and Scottish data when these are both available, the cluster centroids might also be used subsequently to update the UK-wide classification with data of a higher temporal granularity. In this sense, frequently updated observed or modelled small area data sources (e.g., Lansley et al., 2019; Longley et al., 2011) might be used to shift cluster assignments between cluster centroids in light of changes in neighbourhood conditions—measured, for example, on an annual rather than decennial basis (see Lloyd, 2021; Patias & Rowe, 2022). We believe that such attempts are particularly important considering the perceived 'twilight of the Census' and the rising interest in the utilisation of administrative data in the provision of census-like data (Coleman, 2013).

ACKNOWLEDGEMENTS

We acknowledge funding for this paper from the Office for National Statistics, and the Economic and Social Research Council via the Consumer Data Research Centre grant, reference ES/L011840/1.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in NOMIS at https://www.nomisweb.co.uk/sources/census_2021, National Records of Scotland at https://www.nrscotland.gov.uk/statistics-and-data, and NISRA at nisra.gov.uk/statistics. The reference variable names as listed in Table 1, and the and terms and conditions are given at https://www.nomisweb.co.uk/home/terms.asp.

ORCID

Paul A. Longley https://orcid.org/0000-0002-4727-6384

REFERENCES

Anderson, T.K. (2010) Using geodemographics to measure and explain social and environment differences in road traffic accident risk. *Environment and Planning A: Economy and Space*, 42, 2186–2200. Available from: https://doi.org/10.1068/a43157

Ashby, D.I. & Longley, P.A. (2005) Geocomputation, geodemographics and resource allocation for local policing. *Transactions in GIS*, 9, 53–72. Available from: https://doi.org/10.1111/j.1467-9671.2005.00205.x

Bracken, I. & Martin, D.J. (1989) The generation of spatial population distributions from census centroid data. *Environment and Planning A*, 21(4), 537–543.

Catney, G., Lloyd, C.D., Ellis, M., Wright, R., Finney, N., Jivraj, S. et al. (2023) Ethnic diversification and neighbourhood mixing: A rapid response analysis of the 2021 Census of England and Wales. *The Geographical Journal*, 189, 63–77. Available from: https://doi.org/10.1111/geoj.12507

Charlton, M., Openshaw, S. & Wymer, C. (1985) Some new classifications of census enumeration districts in Britain: A poor man's ACORN. *Journal of Economic and Social Measurement*, 13, 69–98.

Coleman, D. (2013) The twilight of the census. *Population and Development Review*, 38, 334–351.

Dewdney, J.C. (1981) *The British census, concepts and techniques in modern geography*. Norwich, UK: Geo Abstracts.

Fleischmann, M. (2023) Clustergram: Visualization and diagnostics for cluster analysis. *Journal of Open Source Software*, 8(89), 5240. Available from: https://doi.org/10.21105/joss.05240

Gale, C.G., Singleton, A.D., Bates, A.G. & Longley, P.A. (2016) Creating the 2011 area classification for output areas (2011 OAC). *Journal of Spatial Information Science*, 12, 1–27. Available from: https://doi.org/10.5311/JOSIS.2016.12.232

Gulma, U.L. (2022) A new geodemographic classification of the influence of neighbourhood characteristics on crime: The case of Leeds, UK. *Computers, Environment and Urban Systems*, 92, 101748. Available from: https://doi.org/10.1016/j.compenvurbsys.2021.101748

Harris, R., Sleight, P. & Webber, R. (2005) *Geodemographics, GIS and neighbourhood targeting, mastering GIS*. West Sussex, England, Hoboken, N.J: Wiley.

Lansley, G., Li, W. & Longley, P.A. (2019) Creating a linked consumer register for granular demographic analysis. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 182, 1587–1605. Available from: https://doi.org/10.1111/rssa.12476

Liu, Y., Singleton, A. & Arribas-Bel, D. (2019) A principal component analysis (PCA)-based framework for automated variable selection in geodemographic classification. *Geo-spatial Information Science*, 22, 251–264. Available from: https://doi.org/10.1080/10095020.2019.1621549

Lloyd, C.D. (2010) *Local models for spatial analysis*. Boca Raton, MA: CRD Press.

Lloyd, C.D. (2021) Neighbourhood change, deprivation, and unemployment in Belfast. *The Geographical Journal*, 188, 190–208. Available from: https://doi.org/10.1111/geoj.12424

Longley, P.A., Cheshire, J.A. & Mateos, P. (2011) Creating a regional geography of Britain through the spatial analysis of surnames. *Geoforum*, 42, 506–516. Available from: https://doi.org/10.1016/j.geoforum.2011.02.001

Longley, P.A., Webber, R. & Li, C. (2008) The UK geography of the E-society: A national classification. *Environment and Planning A*, 40, 362–382. Available from: https://doi.org/10.1068/a3912

Office for National Statistics. (2015) *Pen portraits for the 2011 area classification for output areas*. Available from: https://data.cdrc.ac.uk/output-area-classification-2011/resource/pen-portraits-2011-output-area-classification/resource/pen-portraits-2011-output-area-classification

Office for National Statistics. (2023) *Comparing census 2021 and labour force survey estimates of the labour market, England and Wales: 13 March 2023*. Available from: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/comparingcensus2021andlabourforcesurveyestimatesofthelabourmarketenglandandwales/13march2023

Patias, N. & Rowe, F. (2022) Trajectories of neighbourhood inequality in Britain: Unpacking inter-regional socioeconomic imbalances, 1971−2011. *The Geographical Journal*, 188, 150–165. Available from: https://doi.org/10.1111/geoj.12420

Petersen, J., Gibin, M., Longley, P., Mateos, P., Atkinson, P. & Ashby, D. (2011) Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *Journal of Geographical Systems*, 13, 173–192. Available from: https://doi.org/10.1007/s10109-010-0113-9

Singleton, A., Arribas-Bel, D., Murray, J. & Fleischmann, M. (2022) Estimating generalized measures of local neighbourhood context from multispectral satellite images using a convolutional neural network. *Computers, Environment and Urban Systems*, 95, 101802. Available from: https://doi.org/10.1016/j.compenvurbsys.2022.101802

Singleton, A., Pavlis, M. & Longley, P.A. (2016) The stability of geodemographic cluster assignments over an inter-censal period. *Journal of Geographical Systems*, 18, 97–123. Available from: https://doi.org/10.1007/s10109-016-0226-x

Singleton, A.D. & Longley, P. (2015) The internal structure of greater London: A comparison of national and regional geodemographic models: The internal structure of greater London. *GEO: Geography and Environment*, 2, 69–87. Available from: https://doi.org/10.1002/geo2.7

Singleton, A.D. & Longley, P.A. (2009) Creating open source geodemographics—Refining a national classification of census output areas for applications in higher education. *Papers in Regional Science*, 88, 643–666. Available from: https://doi.org/10.1111/j.1435-5957.2008.00197.x

Stillwell, J. (Ed.). (2017) *The Routledge handbook of census resources, methods and applications: Unlocking the UK 2011 census*, 1st edition. New York, NY: Routledge, Abingdon, Oxon. Available from: https://doi.org/10.4324/9781315564777

The Economist. (2021) *Britain's census form reveals the obsessions of different eras*. Available from: https://www.economist.com/britain/2021/03/20/britains-census-form-reveals-the-obsessions-of-different-eras [Accessed 10th October 2023].

Vickers, D. & Rees, P. (2007) Creating the UK National statistics 2001 Output Area Classification. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 170, 379–403. Available from: https://doi.org/10.1111/j.1467-985X.2007.00466.x

Webber, R.J. (1977) *The national classification of residential neighbourhoods: An introduction to the classification of wards and parishes, PRAG technical papers*. Planning Research Applications Group, Centre for Environmental Studies.

Yang, Y., Dolega, L. & Darlington-Pollock, F. (2023) Ageing in place classification: Creating a geodemographic classification for the ageing population in England. *Applied Spatial Analysis and Policy*, 16, 583–623. Available from: https://doi.org/10.1007/s12061-022-09490-y

How to cite this article: Wyszomierski, J., Longley, P.A., Singleton, A.D., Gale, C. & O'Brien, O. (2024) A neighbourhood Output Area Classification from the 2021 and 2022 UK censuses. *The Geographical Journal*, 190, e12550. Available from: https://doi.org/10.1111/geoj.12550