Environment and Planning A 2010, volume 42, pages 2560 – 2580

doi:10.1068/a42394

The geodemographics of educational progression and their implications for widening participation in higher education

Alex D Singleton
Department of Civic Design, University of Liverpool, 74 Bedford Street South, Liverpool L69 7ZQ, England; e-mail: alex.singleton@liverpool.ac.uk
Received 16 September 2009; in revised form 26 April 2010

Abstract. This paper addresses our ability to analyse progression rates into UK Higher Education (HE) using a range of data available at the individual and neighbourhood levels. The then Department for Children, Schools and Families has recently released data which make it possible to profile national patterns of student educational progression from post-compulsory schooling through to university. However, the linked records lack detailed socioeconomic information, and thus a geodemographic classification is used to analyse the flows of students from different sociospatial backgrounds into the HE system. Rates of progression are shown to vary greatly between these groups, and a disaggregation of HE participants by courses of study demonstrates that the abilities of institutions to attract students from different backgrounds will be constrained by the mix of their course offerings.

1 Stratified access to higher education

Extending access to courses of higher education (HE) remains high on the UK government agenda in its efforts to reduce poverty and maintain competitiveness in an internationalised knowledge-driven economy. Since the 1960s, when only around 20 000 new home students accepted places at HE institutions each year, there has been a huge growth in the absolute numbers of students studying HE. Although an increase in overall HE participation is widely viewed as favourable, academics and policy makers have expressed concerns over the equity of these opportunities across a range of different societal groups both in aggregate (Comptroller and Auditor General, 2008) and by HE institutional type (Chowdry et al, 2008). However, it should be noted that recent evidence (in the case of young participants, at least) suggests that these differentials may have fallen since the mid-2000s (Corver, 2009). Previous research has shown that a series of attributes influence the probability of nonparticipation in HE including levels of affluence (Blanden and Machin, 2004; Machin and Vignoles, 2004), ethnic background (Archer and Hutchings, 2000; Connor et al, 2004), school type attended (Gorrard, 2000), gender (Reay et al, 2005), and class/socioeconomic status (Archer et al, 2003; Ball et al, 2002; Gayle et al, 2002). Unsurprisingly, participation differences have also been shown to manifest spatially (Singleton, 2010; Singleton and Longley, 2009), and indeed, "fewer than one in five young people from the most disadvantaged areas enter higher education" (Corver, 2009, page 1).

Primarily in response to general concerns over differentiation in HE participation rates between socioeconomic strata, central policy measures have been introduced to support and encourage HE institutions to widen access. For example, a 2005 Higher Education Funding Council for England (HEFCE) study created a Participation Of Local Areas (POLAR) classification (Corver, 2005) which is now integrated into both the funding mechanism for HE institutions and their performance metrics. The most recent iteration of this classification[1] presents areal estimates of HE-participation rates among the young based upon those people aged 18 between the years 2000 and 2004, and who then entered HE while still aged 18 or 19 between the academic years

[1] http://www.hefce.ac.uk/widen/polar/polar2/

2000/01 and 2005/06. This classification is disseminated at the 2001 Census Area Statistics Ward level, and the participation rates are divided into quintiles labelled as <16%, 16%–24%, 24%–32%, 32%–43% and >43%. The classification is linked to the HE funding model by providing a monetary premium for those students recruited from areas categorised in the lowest two participation quintiles. These monies are aimed to compensate for the extra efforts required by institutions to recruit and retain students who are domiciled from within these areas. Additionally, POLAR is used as a component of official HE performance indicator calculations(2) which measure the success or failure of an institution in recruiting students from low participation areas.

Concerns over widening participation were enhanced (Callender and Jackson, 2005) with the passing of the 2004 Higher Education Act which introduced legislation enabling an increased but variable fee cap. As part of these changes, HE institutions who wished to charge these additional fees were required to adhere to an 'Access Agreement', with the Office for Fair Access outlining how the extra fee money received would be used to extend access. This reporting requirement was further expanded in 2009 when HEFCE specified that institutions must also file a 'Widening Participation Strategic Assessment', which included a transparent statement of widening participation and access policies, activities, and outcomes. Example initiatives designed by institutions to widen access have included the delivery of outreach events, and provision of bursaries or scholarships. In 2007/08 £192 million of additional government funding was provided to encourage participation amongst lower income and other underrepresented groups (OFFA, 2009).

Given the establishment of equality benchmarks and funding streams linked to widening participation activities, this creates an operational setting in which HE institutions are under increasing pressure to account for those students receiving and accepting offers, therefore driving a demand for better data and analysis mechanisms. As such, the overarching aim of this paper is to address our ability to understand aggregate patterns of access to HE—specifically, by linking summary measures of local neighbourhood characteristics with individual-level educational data which have only recently become available to academic researchers. This study builds cumulatively on previous work utilising these data (see Chowdry et al, 2008) that examined individual characteristics affecting progression rates, albeit with limited consideration of neighbourhood through attributing individuals with deprivation indices and areal proxies for parental education.

2 Linking participation in higher education to neighbourhood characteristics

The aggregate socioeconomic characteristics of neighbourhoods are measured in this study by geodemographic classification which over the years has been used prevalently in commercial sector applications to predict the consumption of a range of private goods and services. However, from a historical perspective, it is the case that the use of geodemographic classifications originated in the UK as a tool for targeting urban deprivation funding (Batey and Brown, 1995; Harris et al, 2005), and the analysis developed here can be seen as part of a resurgence of interest in public sector applications (Longley, 2005)—including examples from targeting urban policy initiatives (Batey and Brown, 2007), policing (Breetzke and Horn, 2009), health (Farr and Evans, 2005) and education (Batey et al, 1999; Butler et al, 2007; Harris et al, 2007; Tonks and Farr, 1995; Webber and Butler, 2007). Geodemographics provides one of a number of potential methods with which to understand differences in educational behaviours between socioeconomic groups; however, there are a number of plausible

(2) See http://www.hesa.ac.uk/index.php/content/category/2/32/141/

reasons why this is the most suitable framework for the type of analyses presented here. Individual-level post-compulsory education (over the age of 16) schools data provide no direct indicator of socioeconomic stratification and, as such, analysis of stratification in their participation rates requires estimation of such indicators from locational or other attributes. Alternative imputation strategies could also include the reuse of indicators linked at the individual level from previous Key Stages, where surrogate measures of socioeconomic status are collected; however, these will suffer temporal decay in relevance and thus may no longer be applicable at a later Key Stage. One of the most common stratification measures used in compulsory education (up to the age of 16) data analysis is eligibility for free school meals (FSM). Families can be means tested to examine their eligibility for FSM, and usually FSM are allocated when they are recipient of different types of social support or have very low earnings. Although used widely in education analysis, this indicator is not collected nor applicable in post-compulsory education; it has also has been argued as deficient in that it reduces disadvantage to a binary outcome (Hobbs and Vignoles, 2007), and introduces biases generated by parents who deem the meals insufficiently nutritious or otherwise inappropriate (Storey and Chamberlin, 2001), or who wish to avoid perceived social stigma associated with accepting them (Styles, 2008). Additionally, alternative measures based on occupation commonly used in social research such as the National Statistics Socio-Economic Classification (NS-SEC), cannot be appended given that parental occupation is not collected in any comprehensive national coverage schools dataset. However, in HE this information is routinely collected as part of the admissions process when applicants complete their Universities and Colleges Admissions Service (UCAS) application form. Although NS-SEC is widely used in HE data analysis, there are a number of significant issues with this measure which are rarely reported. Firstly, the majority of HE data collected originate in the UCAS system, for which 25.5% of applicants accepting degree courses in 2008 had NS-SEC coded as "unknown". Analysis of these missing values suggests that they are not randomly distributed: there is both a school type and an ethnic dimension to their noncompletion (Singleton, 2010). A second issue with the occupational basis to NS-SEC assignments is that they are self-declared and, as such, there is risk of coding errors, particularly if applicants perceive that vague or inappropriate description of occupation might lever relative selection advantage. Within this context, geodemographics can be argued to provide valuable surrogate information in the absence of detailed or reliable individual level socioeconomic indicators.

In common with most other measures of socioeconomic stratification, geodemographic classification is not immune to criticism. Their creation typically uses cluster analysis which aggregates areas sharing similar social, demographic, and physical characteristics into ostensibly homogeneous groupings. These groupings provide the basis of the geodemographic typology, onto which further data are appended and used to describe the salient characteristics of the cluster. As such, geodemographic clusters broadly represent the average characteristics of an area in which people live, and such 'averages' are widely deemed vulnerable to ecological fallacy (confounding the characteristics *of* areas with the individuals who reside *in* them) (Harris et al, 2007), although de Smith et al (2009, page 97) observe that there are few, if any, documented empirical instances of this effect in the literature. Areal aggregations do, nonetheless, subsume heterogeneity in the geographic units used to construct the classification, and the cluster members may themselves be characterised by wide variation (Voas and Williamson, 2001). However, issues of ecological fallacy are of course not limited to geodemographics and arise more generally in most analysis of socioeconomic phenomena where disclosure control is an important issue (Longley et al, 2005).

Neighbourhoods are approximated in the majority of geodemographic classifications by the use of administrative geographies which in the UK typically comprise either 2001 Census Output Areas (40 to ~125 households) or Unit Postcodes [analogous to zip (+4) codes: ~15 households] because of the resolution of the available input data used to create the classifications. Use of such elemental units is, of course, fundamentally a pragmatic decision, as the areal units used to assign the typology are administrative and may not correspond to actual neighbourhoods informed by community structures or interactions. Aggregations of such units also may not approximate 'neighbourhood effects' (Johnston, 1979), which are manifestations of the cumulative results of individual attitudes—exhibited, for example, in voting behaviour—which may be distinct from those patterns expected based upon the sociodemographic characteristics that the individuals hold alone. However, given that a student's schooling and socialisation is more likely than not to occur within an area local to their home, the mix of students within these geographies could be argued as an important contributing factor which influences the attitudinal forming frameworks of the individual and the resulting recorded aggregate behaviours which are measured. It is clear that there is no panacea for profiling socioeconomic differentiation in educational outcomes, and given that direct indicators are argued here as limited at the level of the individual, geodemographic classifications are likely to present a plausible surrogate measure.

3 Integrating educational careers across English school and university data

The English education sector comprises multiple bodies with responsibilities for data collection at different stages in a student's educational career (see figure 1). The majority of these data are collected by the Department for Children, Schools and Families (DCSF); however, in post-compulsory education the Learning and Skills

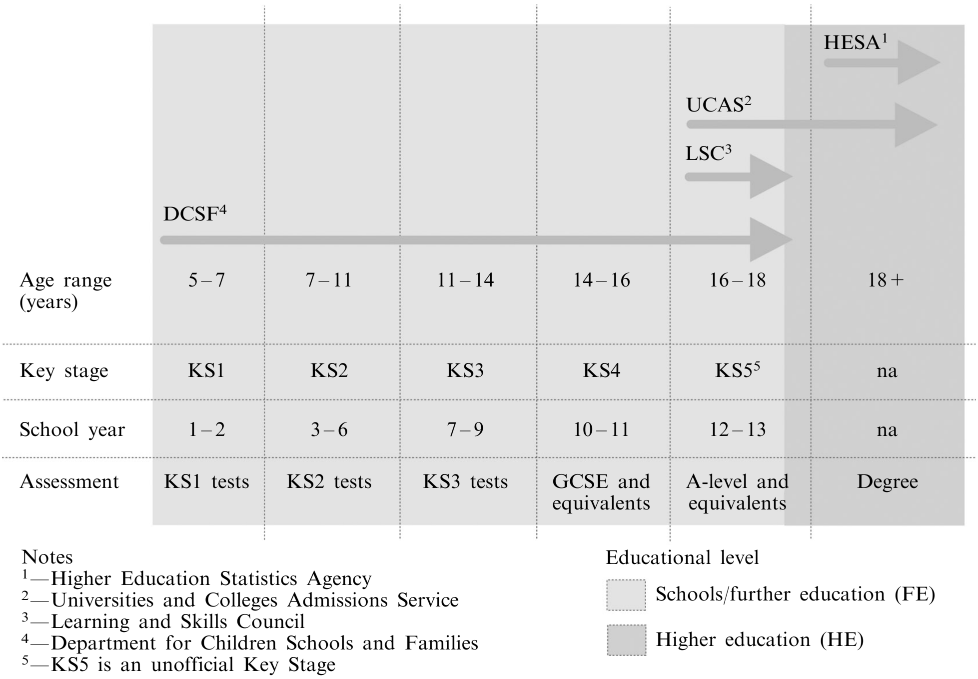

Figure 1. The structure of the English state-funded education sector and those bodies responsible for data collection.

Council (LSC) collect additional data from further education colleges. When students make applications to universities, UCAS manage the offer-making and acceptance data. At the end of each admission round a subset of these data is transferred to universities for incorporation into their own local data-management systems. Finally, each university has a statutory requirement to file a dataset with the Higher Education Statistics Agency (HESA) each year, detailing those students registered at the university.

In order to build a representative picture of the whole education system, data are required for the successive stages of each uniquely identifiable individual's educational career over time, making it possible to relate participation in post-compulsory education and subsequent attainment to characteristics of the individual's education pathway (figure 1). The DCSF has begun to create such a resource through the National Pupil Database (NPD). At an individual level, these data link attainment and choices for students who previously studied in English state-funded schools during Key Stages 1 (KS1) through 5 (KS5). More recently, these data have also been linked to HESA data, thus providing an indication of which students went on to study in HE.

However, there remain challenges and caveats when this information is used in an integrated sense. First, compulsory education may be completed in the independent sector, which includes schools typically supported financially through payment of annual attendance fees. Although the DCSF collects attainment information for students within these institutions, they do not collect demographic data. In the state school sector, demographic data are collected separately under statute through the Pupil Level Annual Schools Census (PLASC), although there are no compliance requirements for the independent sector. Second, at the end of compulsory education, at the age of 16, students can continue their education in either KS5 delivered through a school sixth form, or in a college of further education as part of the learning and skills sector, managed by the LSC. However, the LSC and KS5 data are not mutually exclusive and a student can study in both sectors at the same time. Hence, records from an individual, college, or school can appear in both datasets. Furthermore, because the learning and skills sector supplies a much wider range of courses, the data are structured in a way that is less amenable for analysis. For example, there is only a single variable indicating that a qualification may be at a level that could qualify a learner for potential entry to HE. However, this simply indicates that a qualification is above level three (equivalent to an A-Level) in the National Qualifications Framework,[3] and not necessarily something that an HE institution would consider a suitable entry requirement. In addition, there are a number of specific issues when HE and schools data are linked. For example, UCAS admissions data are not currently integrated into the NPD, and so the initial applications a student makes cannot yet be tracked at an individual level. Given these various challenges it is reasonable to surmise that the assembly and interpretation of educational statistics is complex, and cannot currently give a full and comprehensive picture of educational careers. However, with appropriate caveats, there is still a wealth of information that can be derived about the flows of students from the school system into HE.

4 Measuring rates of participation in higher education

This study uses DCSF and HESA data which were initially sourced through the PLASC/NPD User Group[4] at the Centre for Market and Public Organisation in the University of Bristol. The NPD data are organised around a unique identifier for each student within the system and enable the tracking of individual performance and participation behaviours (eg course choices) over time (see figure 2).

(3) http://www.qcda.gov.uk/5967.aspx

(4) http://www.bris.ac.uk/Depts/CMPO/PLUG/

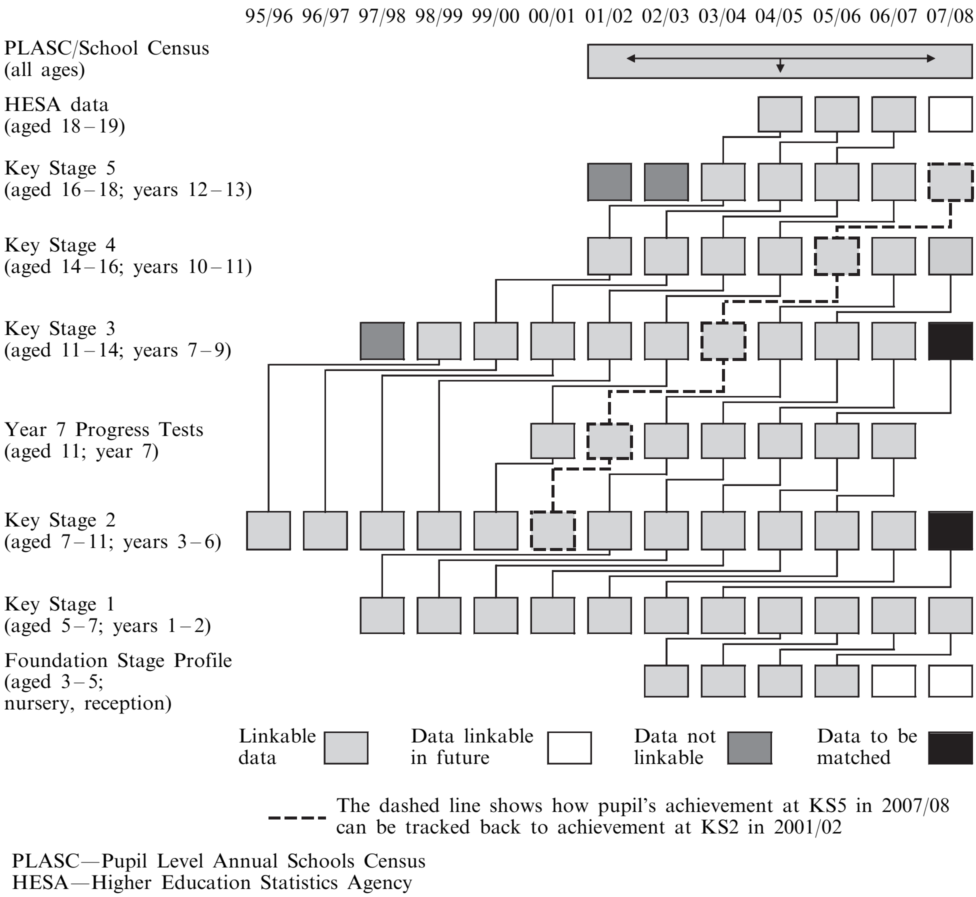

Figure 2. A schematic representation of educational careers (source: PLASC/PLUG User Group).

The analysis presented in this section illustrates some of the challenges faced when calculating progression rates at the geodemographic cluster level and makes comparison with other commonly used HE-participation indices. These comparisons illustrate how the methodology used in calculating official national participation estimates is not directly applicable in a geodemographic context and, additionally, how the NPD-derived estimates are differentiated from other measures of young participation in HE.

The official measure of aggregate HE participation in the UK is calculated as the initial participation rate (HEIPR) statistic (DCSF, 2007). HEIPR was introduced as a method of monitoring progress towards the government's participation-rate target of 50% of the relevant age cohort, which relates to the summation of first-time HE participants aged between 18 and 30 years against denominators for the same age bands derived from up-to-date national population estimates. The precise methods of calculating the numerators and denominators of the HEIPR are thoroughly documented elsewhere (see DCSF, 2003). The HEIPR statistic is presented in aggregate for the total population and it is suggested that rates may have reached 43% during the 2007/08 academic year (DIUS, 2009). This measure is supplemented with the Fulltime Young Participation by Socio-Economic Class (FYPSEC) (Kelly and Cook, 2007) measure which estimates the 'participation gap' between the rates of students entering HE from lowest three NS-SEC groups relative to the highest four; and which in 2007/08 stood at 20.2% (BIS, 2009). However, it should be noted that this measure utilises UCAS data and, as such, suffers those limitations discussed earlier related to missing NS-SEC information for a large proportion of the accepted cohort. The imputation methods utilised in FYPSEC to allocate admitted students who do not have NS-SEC codes involves linkage of the individual to a POLAR quintile related to the ward in which their home address is located. The national distribution of NS-SEC by POLAR is then used to apportion the missing codes and this is also likely to introduce uncertainty into this measure.

It could be argued that a plausible method of calculating differential participation between geodemographic clusters could be simply to disaggregate the HEIPR by each classification group. However, the utility of the HEIPR within this context is eroded: firstly, through analytical uncertainties related to the estimation of appropriate denominator base populations given the transient or migrant characteristics of the age ranges under study (eg the effects of term-time student accommodation); and, secondly, challenges related to disaggregating population base estimate by geodemographic clusters. This second issue arises because of mismatch between the spatial resolution used to append the geodemographic classification (eg Unit Postcode) and the areal unit used to disseminate national base population estimates (eg Lower Super Output Area: LSOA).(5) Thus, although numerator counts of students attending HE for each of a set of geodemographic clusters are reasonably straightforward to extract from HE data, it is much more difficult to create an appropriate estimate of the base population from which these students were drawn. One method commonly used in commercial applications of geodemographics is to manipulate Office for National Statistics (ONS) LSOA mid-year population estimates into the aggregating geographic units of the geodemographic classification—a task which is further complicated by the fact that the LSOA estimates are released only for the year age bands 0–15, 16–29, 30–44, 45–64 (males), 45–59 (females), 65+ (males), and 60+ (females). Although in the commercial sector a pragmatic manipulation of population estimate data may be routine; public sector applications may be concerned with apportionment of potential life chances, and it may be deemed that these methods induce an unacceptable level of uncertainty into profiling practices.

With the advent of the NPD, a more robust solution can be presented which provides significant improvement over previous geodemographic studies, which have relied either on out-of-date population counts derived from Census small area statistics (Batey et al, 1999), or the use of approximated age cohorts based on commercial 'black box' population estimates (Tonks and Farr, 1995). As outlined earlier, the NPD provides a link between school pupils and their subsequent appearance in the HE data. Thus, for a single national state school year cohort the population and location both of the numerator (those students who go to university) and of the denominator (those students who are eligible to go to university) are both known. These data can be aggregated into any organising geography (eg schools, areas, geodemographic clusters) and progression rates calculated. In these calculations, a single school year requires matching against multiple years of HE data to capture those students who do not attend HE directly after finishing their schooling (eg because of gap years). However, in reality, there could be other students who also attended HE from a national school year cohort much later in their lives, and hence would not be captured by this analysis. To illustrate this new method, those students at the end of KS5 in 2004 were matched to the 2004/05, 2005/06 and 2006/07 HE data at the individual level. Table 1 shows the percentages of all (both private school and state school) students appearing in the KS5 data who later attended HE relative to the base population of all students at the end of KS5. It should be noted

(5) Areal population estimates are available from the ONS: http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14357

Table 1. Student flows from the KS5 2004 data into higher education.

	2004	2005	2006	Overall
% of KS5	51.3	20.9	4.5	72.3

that the 'overall' figure is smaller than the sum of the 2004 – 06 percentages as the students who reapplied to new courses of HE during this period have been de-duplicated. Additionally, given the limitations of the NPD, these figures do not include those students studying HE entry qualifications within in the learning and skills sector.

The overall progression rate using this measure was 72.3%, which is high when compared with the latest HEIPR of around 43%. However, it is important to differentiate this initial NPD-derived rate from the HEIPR in two ways. Firstly, the HEIPR considers participation across the population age range 18 – 30 years, rather than a single school-leaving cohort predominantly aged 18 – 19 years. Secondly, and more significantly, the consideration of students progressing from KS5 into HE does not account for those students who left school at the age of 16; that is, those who would appear within the denominator of the HEIPR. A more appropriate rate may therefore be obtained by matching the same student cohort appearing two years previously at the end of KS4 in 2002. When these students are matched to the same HE data, a new rate of 32.0% of the cohort appears to have entered into HE during the recording period (see table 2). The decision to attend HE occurs much earlier than the end of KS5, and indeed the motivation for the majority of students seeking post-compulsory education is to gain qualifications which enable them to attend HE later, as illustrated previously by the high rates of progression in table 1.

Table 2. Student flows from KS4 2002 data into higher education.

	2004	2005	2006	Overall
% of KS4	20.2	10.4	3.5	32.0

Caution must be taken when comparing the NPD derived figures with official estimates of the HEIPR, especially given the differences in age ranges considered by the two measures. These NPD-derived figures are predominantly representative of the young participation that accounts for the majority of HE places (see figure 3) and, as such, could usefully be compared with the HEFCE POLAR classification.

The KS4 progression rates were summarised with the same participation rate break points as POLAR. A cross-tabulation was made between the POLAR classification and these NPD-derived progression rate bands, the results of which are presented in table 3. Although there is reasonable correspondence between the two measures ($p = <0.0001$; $\chi^2 = 7921.49$; df $= 16$) there is some dissimilarity, which may reflect differences between the underlying construction methodologies of the two measures. For example: POLAR rates include students from the independent sector; some wards could not be assigned a score where the frequency of state school students was very low (eg City of London); and the NPD rates correspond to a single school year and, as such, may suffer from temporal instability, whereas POLAR utilises multiple years of admissions data. The differences between the classifications mainly emerge in the POLAR band '24% – 32%' which matches with only 37.1% of the corresponding band in the NPD-derived classification; however, also 40.6% in the band '32% – 43%'. The differences between the categorical assignments of areas into participation quintiles were mapped and compared and there was no apparent geographical bias in these results.

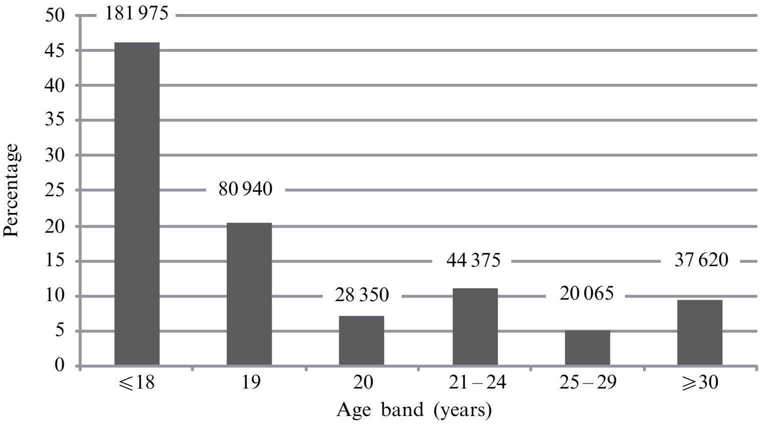

Figure 3. All first-year full-time undergraduates in 2005/06 higher education (source: Higher Education Statistics Agency).

Table 3. Numbers (with percentages shown in parentheses) of census area statistics wards by the POLAR (participation of local areas) 2 classification and NPD (National Pupil Database)-banded estimates.

NPD	POLAR2				
	<16%	16% – 24%	24% – 32%	32% – 43%	>43%
<16%	621 (51.0)	126 (9.3)	43 (2.7)	20 (1.1)	7 (0.4)
16 – 24%	490 (40.2)	525 (38.7)	225 (14.1)	82 (4.4)	34 (1.8)
24% – 32%	95 (7.8)	559 (41.2)	592 (37.1)	305 (16.2)	91 (4.7)
32% – 43%	10 (0.8)	141 (10.4)	648 (40.6)	886 (47.2)	375 (19.6)
>43%	1 (0.1)	6 (0.4)	88 (5.5)	585 (31.1)	1404 (73.2)
Missing	1 (0.1)	1 (0.1)	0 (0)	1 (0.1)	7 (0.4)
Total	1218 (100)	1358 (100)	1596 (100)	1879 (100)	1918 (100)

5 Calculating rates of progression by geodemographic clusters

In order to examine HE differentials between geodemographic clusters, in the following analysis the overall rate of HE progression from the 2002 KS4 data into geodemographic clusters is disaggregated. The specific geodemographic typology used for this analysis is A Classification Of Residential Neighbourhoods (ACORN) supplied under an academic licence by CACI.(6) This classification is created at the Unit Postcode level using input data both from the 2001 Census of the Population and various CACI consumer databases. These input data were standardised and clustered using computer algorithms which seek to order areas into a series of groups that are considered most homogeneous across their input variables. A detailed methodology can be found in Harris et al (2005). The ACORN classification sorts every postcode in the UK into fifty-six clusters called 'types', which are subsequently hierarchically nested by further clustering algorithms into two aggregate levels of seventeen 'groups' and five 'categories'. Further data are appended to these clusters and are used by the classification-building team to name and describe their characteristics. In order to reduce duplication, the ACORN names are not presented here in situ, but instead form part of the results presented later in this section. Detailed descriptive information on the characteristics of the clusters can be found on the CACI website.(7)

(6) http://www.caci.co.uk/

(7) http://www.caci.co.uk/acorn/

By use of the matching methodology outlined in the previous section, disaggregated progression rates were calculated for each ACORN Type. In addition, upper and lower 95% confidence intervals were calculated using Byar's approximation (Rothman, 2002). Although there are numerous methods for estimating confidence intervals for data with a Poisson distribution (Swift, 2009), Byar's approximation is commonly used in epidemiological analysis and is equally applicable for the count of observed HE participants distributed across ACORN types. Byar's approximation is illustrated by equation (1) and equation (2).

$$P_{\text{lower}} = C\left(1 - \frac{1}{9C} - \frac{z}{3C^{1/2}}\right)^3, \quad (1)$$

$$P_{\text{upper}} = C\left[1 - \frac{1}{9(C+1)} - \frac{z}{3(C+1)^{1/2}}\right]^3, \quad (2)$$

where P is the participation count, C is the count of the observed values, and z is the standard deviation measure (1.96) relating to the 95% value of a standard normal distribution curve.

The progression rates are shown in table 4 alongside category, group, and type labels, and additionally are visualised with the confidence intervals for each ACORN type in figure 4. The HE progression rates are shown with a line between the mid-grey and dark grey bars that illustrate the upper and lower bounds of the confidence intervals. When interpreting the short descriptions of the ACORN types, it is important to remember that these relate to the general characteristics of the area, and not to the characteristics of the students themselves. For example, although an area could be

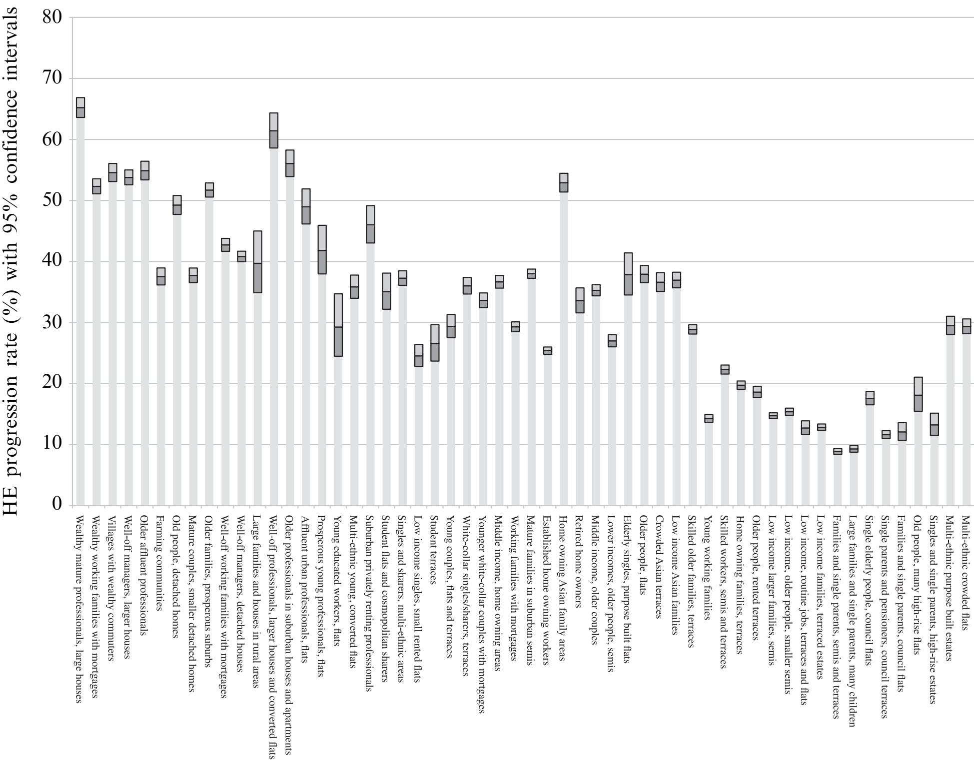

Figure 4. KS4 to higher education (HE) progression rates by ACORN (A Classification of Residential Neighbourhoods) types.

Table 4. KS4 to HE participation rates by ACORN (A Classification of Residential Neighbourhoods) types.

ACORN category	ACORN group	ACORN type	Participation rate (%)	Participation rank	Income index score
1: Wealthy Achievers	1A: Wealthy Executives	1A1: Wealthy mature professionals, large houses	64	1	151
		1A2: Wealthy working families with mortgages	51	8	153
		1A3: Villages with wealthy commuters	53	5	130
		1A4: Well-off managers, larger houses	53	6	133
	1B: Affluent Greys	1B5: Older affluent professionals	53	4	115
		1B6: Farming communities	36	19	100
		1B7: Old people, detached homes	48	10	106
		1B8: Mature couples, smaller detached homes	37	17	94
	1C: Flourishing Families	1C9L Older families, prosperous suburbs	51	9	127
		1C10: Well-off working families with mortgages	42	13	135
		1C11: Well-off managers, detached houses	40	14	112
		1C12: Large families and houses in rural areas	35	24	102
2: Urban Prosperity	2A: Prosperous Professionals	2A1: Well-off professionals, larger houses and converted flats	59	2	143
		2A2: Older professionals in suburban houses and apartments	54	3	126
	2B: Educated Urbanites	2B3: Affluent urban professionals, flats	46	11	143
		2B4: Prosperous young professionals, flats	38	15	140
		2B5: Young educated workers, flats	24	39	99
		2B6: Multi-ethnic young, converted flats	34	28	118
		2B7: Suburban privately renting professionals	43	12	118
	2C: Aspiring Singles	2C8: Student flats and cosmopolitan sharers	32	30	97
		2C9: Singles and sharers, multi-ethnic areas	36	20	102
		2C10: Low income singles, small rented flats	23	41	84
		2C11: Student terraces	24	40	100
3: Comfortably Off	3A: Starting Out	3A1: Young couples, flats and terraces	28	36	111
		3A2: White-collar singles/sharers, terraces	35	25	110
	3B: Secure Families	3B3: Younger white-collar couples with mortgages	32	29	123
		3B4: Middle income, home owning areas	36	22	108
		3B5: Working families with mortgages	29	32	113
		3B6: Mature families in suburban semis	37	16	108

Table 4 (continued).

ACORN category	ACORN group	ACORN type	Participation rate (%)	Participation rank	Income index score
3: Comfortably Off	3B: Secure Families	3B7: Established home owning workers	25	38	93
		3B8: Home owning Asian family areas	51	7	107
	3C: Settled Suburbia	3C9: Retired home owners	32	31	77
		3C10: Middle income, older couples	34	27	95
		3C11: Lower incomes, older people, semis	26	37	84
	3D: Prudent Pensioners	3D12: Elderly singles, purpose built flats	35	26	84
		3D13: Older people, flats	37	18	96
4: Moderate Means	4A: Asian Communities	4A1: Crowded Asian terraces	35	23	78
		4A2: Low income Asian families	36	21	80
	4B: Post-Industrial Families	4B3: Skilled older families, terraces	28	34	106
		4B4: Young working families	14	49	88
	4C: Blue-Collar Roots	4C5: Skilled workers, semis and terraces	22	42	91
		4C6: Home owning families, terraces	19	43	83
		4C7: Older people, rented terraces	18	44	79
5: Hard-Pressed	5A: Struggling Families	5A1: Low income larger families, semis	14	48	81
		5A2: Low income, older people, smaller semis	15	47	73
		5A3: Low income, routine jobs, terraces and flats	12	51	79
		5A4: Low income families, terraced estates	12	50	73
		5A5: Families and single parents, semis and terraces	8	56	67
		5A6: Large families and single parents, many children	9	55	74
	5B: Burdened Singles	5B7: Single elderly people, council flats	17	45	66
		5B8: Single parents and pensioners, council terraces	11	53	65
		5B9: Families and single parents, council flats	11	54	69
	5C: High-Rise Hardship	5C10: Old people, many high-rise flats	15	46	58
		5C11: Singles and single parents, high-rise estates	12	52	64
	5D: Inner City Adversity	5D12: Multi-ethnic purpose built estates	28	35	86
		5D13: Multi-ethnic crowded flats	28	33	86

characterised as 'Elderly singles, purpose built flats', this does not mean that the student or parents possess these characteristics—simply that these are the dominant characteristics of the area in which they live. Additionally, the rates presented for the ACORN type are for progression into HE by those pupils recorded within the DCFS data rather than a participation rate for the area against some base population. This affects interpretation; for example, the ACORN type 'Student terraces' might be hypothesised as having a high participation rate, given the dominance of student accommodation within these areas; however, the recorded progression rate is actually low (24%). Student areas will usually have low rents and often accommodation that is of poor quality, and as such those families living within these areas (not attending HE) may be of lower income/ more deprived, thus decreasing the probability that they will progress into HE.

Overall there are very large differentials between type-level rates of progression, ranging from 64% in areas classified as 'Wealthy mature professionals, large houses' through to only 8% in the type 'Families and single parents, semis and terraces'. These differentials are staggering and illustrate huge variance in the apportionment of potential life chances between different aggregate geodemographic neighbourhoods. As one might expect, these inequality profiles share reasonable correspondence to measures of affluence—which is illustrated as a plot of participation against a commercially derived (supplied with ACORN) average household income estimate for each type (see figure 5). However, there are other interesting patterns that contradict this broad relationship and which appear to correspond to the ethnic composition of the geodemographic clusters. For example, from within the deprived category 'Hard-Pressed' progression into HE from the types 'Multi-ethnic purpose built estates' and 'Multi-ethnic crowded flats' are both 28%, which is 10% greater than all other types within this same category, albeit still low. Another example would be the type 'Home owning Asian Family Areas' where the rate of progression is 52%, again around 10% higher than other types in the category 'Comfortably Off'.

Although these type-level patterns appear reasonably logical and reinforce findings from previous studies (see section 1), within categories there is marked diversity

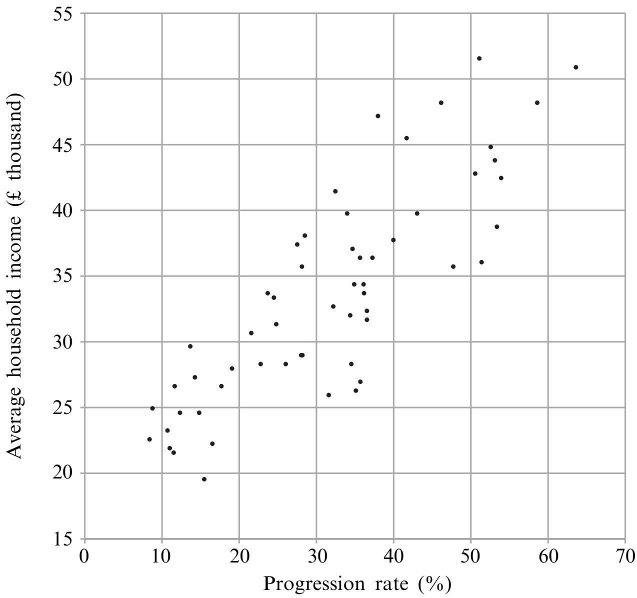

Figure 5. KS4 to HE progression rates and average household income estimates of ACORN (A Classification of Residential Neighbourhoods) types.

in progression rates between Types. The three lowest participation types within the Category 'Urban Prosperity' are in fact lower than the highest two types within the category 'Hard-Pressed'. For example, within the most affluent category 'Wealthy Achievers', type-level participation rates range from to 64% ('Wealthy mature professionals, large houses') to 35% ('Large families and houses in rural areas'). Similarly, within the category 'Urban Prosperity', participation rates range from 59% ('Well-off professionals, larger houses and converted flats') to 23% ('Low income singles, small rented flats'). Although the geodemographic classification does not explain the occurrence of these different rates, it does provide a method of hypothesising about the potential processes underlying the observations. For example, 'Low income singles, small rented flats' are areas typically inhabited by single people in their twenties and, as such, people living within these areas are unlikely to have children of school age. Thus, it could be hypothesised that if students were found living with a parent in these areas, they may be doing so under adverse circumstances (eg perhaps after a divorce) given the typical unsuitability of the accommodation found in these areas for raising a family. However, in order to draw firm conclusions from these types of hypothesis, more robust qualitative or quantitative testing would be required. Given these issues, and under the assumption that the type-level classification is robust, the organisation of these clusters into aggregate groups or categories using application specific data might be more appropriate. This would be especially useful for those categories and groups containing types with low HE progression characteristics which are described in language which might not normally be interpreted as indicative of low HE participation areas (eg 'Wealthy', 'Prosperity', 'Aspiring', 'Educated', 'Affluent'). These labels could potentially lead to erroneous targeting, especially given a prevailing tendency of new users to interpret the most aggregate levels of a classification superficially or uncritically.

6 The geodemographics of course selection and implications for widening participation

Through integration of the NPD and HE data at an individual level, in the preceding section I highlighted how a geodemographic classification could be utilised as a method of identifying and characterising those types of areas that may be subject to low rates of progression into HE. However, this analysis can be extended to examine subjects of study typically chosen by entrants to HE within different geodemographic groups. Operationally, a better understanding of these differences could help provide targeted information that would be useful in efforts to increase participation in specific subjects. From a widening-participation perspective this analysis also challenges any assumption that all subject areas are equally attractive or attainable by different geodemographic strata.

In these calculations, the HE data alone were used to create subject profiles. Although it is possible to use the link methodology presented in the previous section at a subject level, these analyses would simply reflect the aggregate participation profile rather than demonstrating any differentials between subjects of study. Thus, in this analysis the numerators are the sum of those students entering HE to study a specific course over a denominator of all those students entering HE. The HE population considered was first year, first degree, UK-domiciled full-time students. Confidence intervals were again calculated using the method presented earlier in equations (1) and (2). Subjects were ordered using the Joint Academic Coding System (JACS)[8] which aggregates all courses of HE into a series of subject lines. There are 1281 subject lines which include descriptions such as "A1—Pre-clinical Medicine" or

[8] The full JACS schema is available from the HESA website: http://www.hesa.ac.uk/dox/jacs/JACS_complete.pdf

Table 5. JACS (Joint Academic Coding System) subject groups.

Group code	Subject description
A	Medicine and Dentistry
B	Subjects allied to Medicine
C	Biological Sciences
D	Veterinary Sciences, Agriculture and related subjects
F	Physical Sciences
G	Mathematical and Computer Sciences
H	Engineering
J	Technologies
K	Architecture, Building and Planning
L	Social studies
M	Law
N	Business and Administrative studies
P	Mass Communications and Documentation
Q	Linguistics, Classics and related subjects
R	European Languages, Literature and related subjects
T	Eastern, Asiatic, African, American and Australasian Languages, Literature and related subjects
V	Historical and Philosophical studies
W	Creative Arts and Design
X	Education

"G4—Computer Science". These subject lines nest within more aggregate subject groups, which are shown in table 5.

The geodemographic rates of participation are shown in table 6 for all subjects, and are visualised for a selected number of the larger and better defined JACS groups in figure 6. Course groups are organised in this figure by small multiple radial diagrams, in which the grey area represents the participation rates by specific ACORN groups, and the two lighter grey radial lines the upper and lower 95% confidence intervals—again calculated using Byars's approximation. The scale used on each chart has been variably assigned to account for differences in the overall size of the JACS groups presented. These analyses therefore illustrate the relative subject preference by students who are resident within different geodemographic clusters.

The participation rate differentials between both the subject and the ACORN groups are very apparent. Figure 6 shows how JACS groups vary in distinctiveness of their geodemographic profile. For example, subjects such as 'Medicine and Dentistry' can be seen to be a relatively more prevalent option in specific geodemographic clusters, whereas subjects like 'Engineering' or 'Biological Sciences' are more uniformly appealing across clusters. Some specific observations that can be made from these analyses include strong bias (relative to other geodemographic groups) for students from the ACORN group 'Asian Communities' to study specific subjects including 'Law', 'Business and Administrative Studies' and 'Mathematical and Computer Sciences', but not language-based subjects, 'Creative Arts and Design', or 'Historical and Philosophical Studies'. There are also some acceptances to 'Medicine and Dentistry', but very few to 'Veterinary Sciences, Agriculture and related subjects' despite the similar entrance requirements of these courses. Overall, it appears that students resident within this ACORN group appear to be favouring subjects that could be argued as having greater perceived vocational relevance. In the JACS group 'Medicine and Dentistry', there are very few acceptances from the most deprived ACORN groups, presumably related to the high entrance requirements and/or work experience required to study within these areas. There also appear to be relatively fewer students from the ACORN groups that

Table 6. Subject participation rates (numbers with percentages shown in parentheses) frequency by JACS (Joint Academic Coding System), and ACORN (A Classification of Residential Neighbourhoods) groups.

ACORN group	JACS subject group																		
	A	B	C	D	F	G	H	J	K	L	M	N	P	Q	R	T	V	W	X
1A: Wealthy Executives	1364 (4.5)	2026 (6.7)	3567 (11.8)	488 (1.6)	1890 (6.2)	1546 (5.1)	1668 (5.5)	291 (1.0)	1006 (3.3)	2558 (8.4)	1264 (4.2)	3263 (10.8)	853 (2.8)	1276 (4.2)	339 (1.1)	222 (0.7)	1829 (6.0)	3946 (13.0)	834 (2.7)
1B: Affluent Greys	573 (3.3)	1397 (8.2)	1979 (11.6)	454 (2.7)	1122 (6.6)	930 (5.4)	944 (5.5)	184 (1.1)	562 (3.3)	1197 (7.0)	683 (4.0)	1653 (9.7)	467 (2.7)	672 (3.9)	170 (1.0)	94 (0.5)	911 (5.3)	2449 (14.3)	586 (3.4)
1C: Flourishing Families	603 (2.7)	1861 (8.5)	2728 (12.4)	319 (1.5)	1295 (5.9)	1380 (6.3)	1129 (5.1)	241 (1.1)	645 (2.9)	1618 (7.4)	988 (4.5)	2203 (10.0)	697 (3.2)	892 (4.1)	210 (1.0)	113 (0.5)	1123 (5.1)	3019 (13.7)	817 (3.7)
2A: Prosperous Professionals	255 (4.6)	257 (4.7)	554 (10.0)	38 (0.7)	311 (5.6)	282 (5.1)	241 (4.4)	50 (0.9)	150 (2.7)	573 (10.4)	201 (3.6)	470 (8.5)	171 (3.1)	338 (6.1)	91 (1.6)	64 (1.2)	517 (9.4)	842 (15.3)	97 (1.8)
2B: Educated Urbanites	220 (3.5)	422 (6.8)	628 (10.1)	31 (0.5)	215 (3.4)	379 (6.1)	340 (5.4)	49 (0.8)	147 (2.4)	631 (10.1)	270 (4.3)	708 (11.3)	209 (3.3)	316 (5.1)	83 (1.3)	60 (1.0)	477 (7.6)	955 (15.3)	73 (1.2)
2C: Aspiring Singles	132 (2.3)	523 (9.0)	597 (10.3)	19 (0.3)	171 (2.9)	535 (9.2)	305 (5.3)	47 (0.8)	144 (2.5)	504 (8.7)	366 (6.3)	844 (14.5)	204 (3.5)	180 (3.1)	38 (0.7)	17 (0.3)	196 (3.4)	836 (14.4)	114 (2.0)
3A: Starting Out	76 (1.9)	322 (7.9)	441 (10.8)	39 (1.0)	197 (4.8)	244 (6.0)	182 (4.5)	45 (1.1)	106 (2.6)	320 (7.9)	167 (4.1)	386 (9.5)	146 (3.6)	178 (4.4)	39 (1.0)	24 (0.6)	206 (5.1)	811 (19.9)	123 (3.0)
3B: Secure Families	675 (2.2)	3047 (9.7)	3831 (12.2)	333 (1.1)	1593 (5.1)	2348 (7.5)	1508 (4.8)	293 (0.9)	719 (2.3)	2206 (7.1)	1514 (4.8)	3513 (11.2)	1107 (3.5)	1073 (3.4)	239 (0.8)	111 (0.4)	1364 (4.4)	4366 (14.0)	1290 (4.1)
3C: Settled Suburbia	163 (1.9)	834 (9.6)	1139 (13.1)	116 (1.3)	513 (5.9)	594 (6.9)	415 (4.8)	90 (1.0)	188 (2.2)	594 (6.9)	396 (4.6)	779 (9.0)	302 (3.5)	321 (3.7)	67 (0.8)	45 (0.5)	345 (4.0)	1293 (14.9)	435 (5.0)
3D: Prudent Pensioners	95 (2.6)	295 (8.1)	442 (12.1)	32 (0.9)	200 (5.5)	213 (5.8)	176 (4.8)	39 (1.1)	82 (2.2)	297 (8.1)	159 (4.4)	360 (9.9)	126 (3.4)	144 (3.9)	39 (1.1)	17 (0.5)	216 (5.9)	584 (16.0)	123 (3.4)
4A: Asian Communities	113 (2.4)	588 (12.3)	444 (9.3)	9 (0.2)	116 (2.4)	603 (12.6)	272 (5.7)	19 (0.4)	157 (3.3)	339 (7.1)	528 (11.1)	909 (19.0)	88 (1.8)	77 (1.6)	5 (0.1)	1 (0.0)	71 (1.5)	259 (5.4)	136 (2.8)
4B: Post-industrial Families	98 (1.4)	612 (8.6)	921 (12.9)	68 (1.0)	336 (4.7)	589 (8.3)	330 (4.6)	65 (0.9)	135 (1.9)	461 (6.5)	372 (5.2)	756 (10.6)	253 (3.5)	257 (3.6)	40 (0.6)	30 (0.4)	295 (4.1)	1165 (16.3)	317 (4.4)
4C: Blue-collar Roots	90 (1.1)	797 (9.6)	1029 (12.4)	104 (1.3)	418 (5.1)	615 (7.4)	382 (4.6)	91 (1.1)	148 (1.8)	584 (7.1)	435 (5.3)	827 (10.0)	302 (3.6)	305 (3.7)	51 (0.6)	29 (0.4)	357 (4.3)	1255 (15.2)	402 (4.9)
5A: Struggling Families	99 (0.9)	1115 (10.2)	1400 (12.8)	95 (0.9)	531 (4.8)	945 (8.6)	466 (4.3)	97 (0.9)	221 (2.0)	759 (6.9)	623 (5.7)	1190 (10.9)	392 (3.6)	348 (3.2)	57 (0.5)	30 (0.3)	363 (3.3)	1616 (14.7)	531 (4.8)
5B: Burdened Singles	20 (0.8)	243 (9.5)	300 (11.7)	24 (0.9)	120 (4.7)	237 (9.2)	123 (4.8)	28 (1.1)	60 (2.3)	188 (7.3)	140 (5.5)	295 (11.5)	102 (4.0)	83 (3.2)	14 (0.5)	10 (0.4)	69 (2.7)	395 (15.4)	96 (3.7)
5C: High-rise Hardship	5 (1.1)	47 (10.3)	42 (9.2)	2 (0.4)	19 (4.2)	43 (9.4)	23 (5.0)	5 (1.1)	9 (2.0)	39 (8.5)	24 (5.3)	59 (12.9)	20 (4.4)	9 (2.0)	1 (0.2)	1 (0.2)	13 (2.8)	80 (17.5)	13 (2.8)
5D: Inner City Adversity	67 (1.6)	358 (8.7)	498 (12.1)	4 (0.1)	62 (1.5)	443 (10.8)	226 (5.5)	20 (0.5)	98 (2.4)	310 (7.6)	306 (7.5)	770 (18.8)	166 (4.0)	88 (2.1)	13 (0.3)	10 (0.2)	96 (2.3)	472 (11.5)	79 (1.9)

could be considered more affluent (eg 1A – 2B) in 'Mathematical and Computer Sciences' and 'Law', perhaps related to changing perceptions of future employment opportunities. A further interesting pattern that emerges in 'Social Studies', languages, and 'Historical and Philosophical studies' is the dominance of ACORN groups that could be considered as representing professional backgrounds (2A – 2B), yet relatively fewer students from the groups in the most affluent areas (broadly, category 1). The inverse of this relationship appears to hold for 'Education'.

Overall, the differential rates of participation between ACORN groups demonstrated by this analysis are reflective of similar findings from a range of other previous non-geodemographic studies in which entrance profiles of HE degrees in 'mathematics' (Kitchen, 1999), 'medicine' (Seyan et al, 2004) 'physics', and 'chemistry' (Elias et al, 2006) have been examined. These specific studies, and other broader analyses of HE participation rates draw together a number of factors influencing an individual's probability of participating in specific HE subjects—including attainment (Leslie, 2003), ethnicity (Connor et al, 2004), gender (Mastekaasa and Smeby, 2008), and social class (Reay et al, 2005). Thus, it is quite reasonable to expect that subject selection differentials will also manifest between ACORN groups where aggregations of people who exhibit some of these influencing factors also cluster spatially.

The analysis presented in this section has implications for widening participation in HE given that those students attending an institution will be at least partially determined by the portfolio of courses that are on offer. In an operational setting, subject-profiling activities could be used in attempts to widen participation: for example, through the stratified allocation of bursaries to courses with low participation across ACORN types typically underrepresented in HE as a whole; or, alternatively, institutions might utilise subject-based benchmarking for internal widening participation audits across their departments. This information would be useful to determine differences between those departments which are genuinely widening participation versus those which are attracting students as a result of the inherent bias induced by the subjects that they offer.

7 Conclusions

The analyses presented in this paper have demonstrated how individually linked records enabled by the NPD make it possible to track and profile flows of students from school into HE institutions. These linkages were utilised to establish an alternative basis for the calculation of differential progression rates across a geodemographic typology. It is proposed that this method provides a more robust basis for HE institutions to use geodemographic classifications in widening-participation activities, and it is argued as being a significant improvement on previously demonstrated techniques which have relied on commercially modelled measures or outdated census population data. This analysis led logically to geodemographic profiling of an entire HE cohort in order to analyse the degree to which students from different neighbourhood types participated in specific HE subjects of study. These expanded analyses demonstrated stark differences in the revealed preferences for subjects of study across different neighbourhoods. These results reinforce findings from previous studies; however, they also demonstrated the utility of geodemographics to differentiate subject preferences by those students from areas which might simply be regarded as 'affluent' when alternative metrics were used. The utility of the analyses and data presented in this paper should be of interest to HE institutions, both for increasing participation and for widening access. However, the tension between academic excellence and equality of opportunity is something which HE institutions cannot address independently, and it is only through better investment in the school system supporting the most deprived neighbourhoods that we are likely to see an improvement in the rates of HE participation across all neighbourhoods.

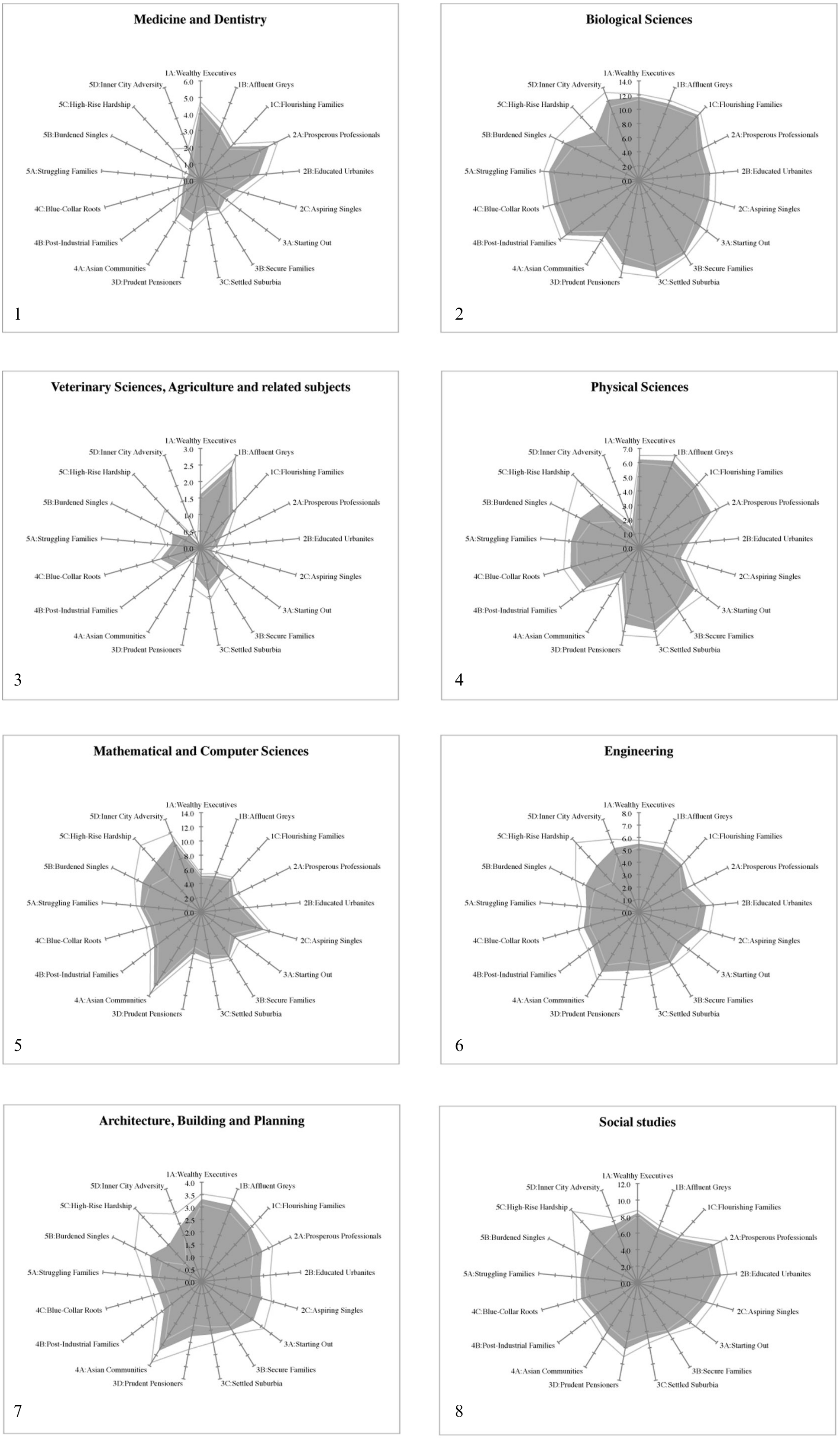

Figure 6. Selected JACS (Joint Academic Coding System) subject participation rates with confidence intervals by ACORN (A Classification of Residential Neighbourhoods) groups.

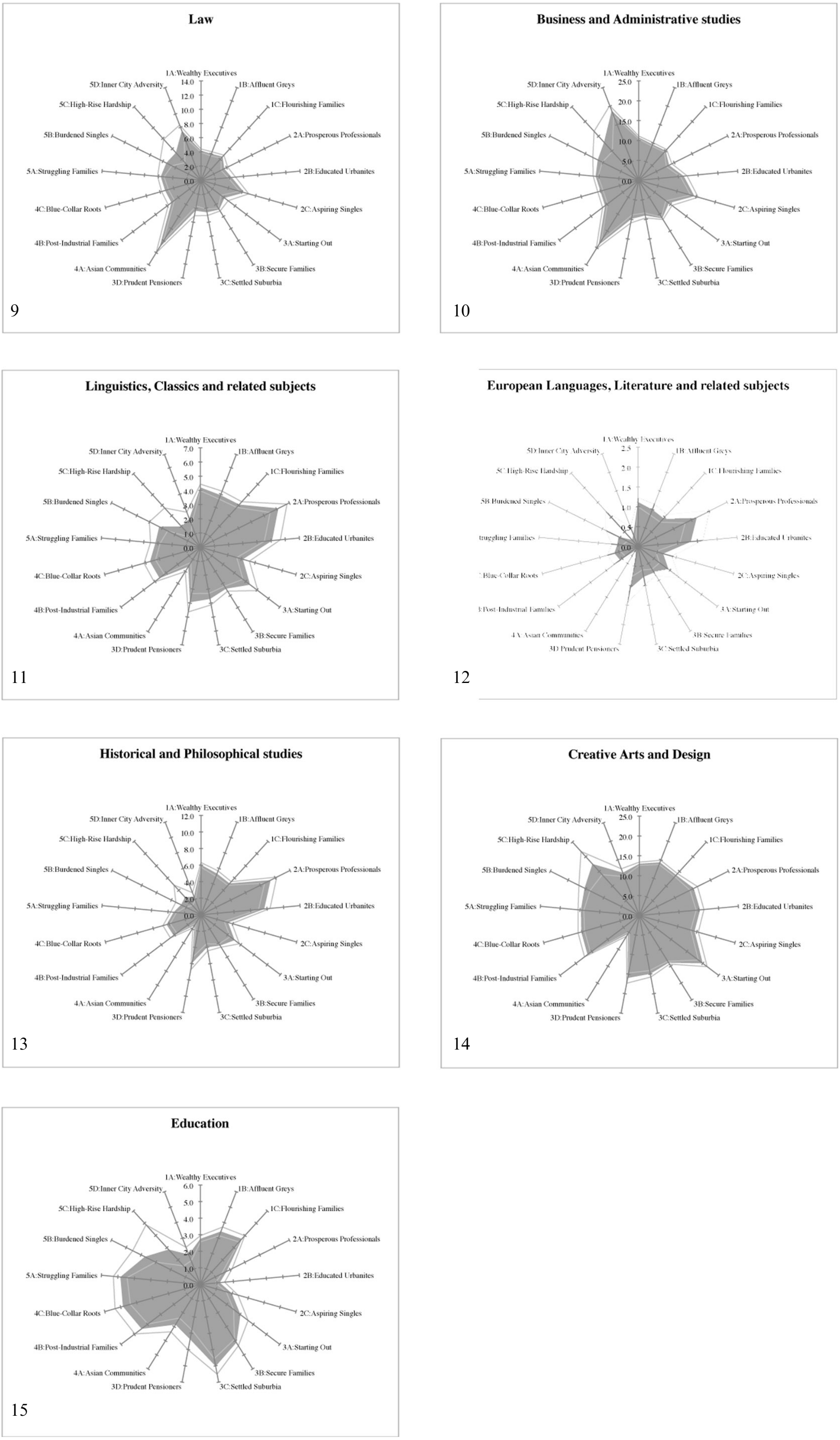

Figure 6 (continued).

References

Archer L, Hutchings M, 2000, "'Bettering yourself'? Discourses of risk, cost and benefit in ethnically diverse, young working-class non-participants' constructions of higher education" *British Journal of Sociology of Education* **21** 555 – 574

Archer L, Hutchings M, Ross A, 2003 *Higher Education and Social Class* (Routledge, London)

Ball S J, Davies J, David M, Reay D, 2002, " 'Classification' and 'judgement': social class and the 'cognitive structures' of choice of higher education" *British Journal of Sociology of Education* **23** 55 – 72

Batey P W J, Brown P J B, 1995, "From human ecology to customer targeting: the evolution of geodemographics", in *GIS for Business and Service Planning* Eds P Longley, G Clarke (GeoInformation International, Cambridge) pp 77 – 103

Batey P, Brown P J B, 2007, "The spatial targeting of urban policy initiatives: a geodemographic assessment tool" *Environment and Planning A* **39** 2774 – 2793

Batey P, Brown P J B, Corver M, 1999, "Participation in higher education: a geodemographic perspective on the potential for further expansion in student numbers" *Journal of Geographical Systems* **1** 277 – 303

BIS, 2009, "Full-time young participation by socio-economic class (FYPSEC) update 2009", Department for Business, Innovation and Skills, London

Blanden J, Machin S, 2004, "Educational inequality and the expansion of UK higher education" *Scottish Journal of Political Economy* **51** 230 – 249

Breetzke G D, Horn A C, 2009, "A geodemographic profiler for high offender propensity areas in the city of Tshwane, South Africa" *Environment and Planning A* **41** 112 – 127

Butler T, Hamnett C, Ramsden M, Webber R, 2007, "The best, the worst and the average: secondary school choice and education performance in East London" *Journal of Education Policy* **22** 7 – 29

Callender C, Jackson J, 2005, "Does the fear of debt deter students from higher education?" *Journal of Social Policy* **34** 509 – 540

Chowdry H, Crawford C, Dearden L, Goodman A, Vignoles A, 2008, "Widening participation in higher education: analysis using linked administrative data", The Institute for Fiscal Studies, London

Comptroller and Auditor General, 2008, "Widening Participation in Higher Education", National Audit Office, London

Connor H, Tyers C, Modood T, Hillage J, 2004, "Why the difference?: a closer look at higher education minority ethnic students and graduates", Department for Education and Skills, London

Corver M, 2005, "Young participation in higher education", Higher Education Funding Council for England, Bristol

Corver M, 2009, "Trends in young participation in higher education: core results for England", Higher Education Funding Council for England, Bristol

DCSF, 2003 *Review of the Initial Entry Rate into Higher Education* Department for Children, Schools and Families, London

DCSF, 2007 *Methodological Revisions to the Higher Education Initial Participation Rate* Department for Children, Schools and Families, London

de Smith M, Goodchild M F, Longley P A, 2009 *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools* (Troubador Publishing, Leicester)

DIUS, 2009 *Participation rates in higher education: academic years 1999/2000 – 2007/2008 (Provisional)* Department for Innovation, Universities and Skills, London

Elias P, Jones P, McWhinnie S, 2006, "Representation of ethnic groups in chemistry and physics: a report prepared for the Royal Society of Chemistry and the Institute of Physics", Royal Society of Chemistry and the Institute of Physics, London

Farr M, Evans A, 2005, "Identifying 'unknown diabetics' using geodemographics and social marketing" *Interactive Marketing* **7** 47 – 58

Gayle V, Berridge D, Davies R, 2002, "Young people's entry into higher education: quantifying influential factors" *Oxford Review of Education* **28** 5 – 20

Gorrard S, 2000 *Education and Social Justice: The Changing Composition of Schools and its Implications* (University of Wales Press, Cardiff)

Harris R, Sleight P, Webber R, 2005 *Geodemographics, GIS and Neighbourhood Targeting* (John Wiley, Chichester, Sussex)

Harris R, Johnston R, Burgess S, 2007, "Neighborhoods, ethnicity and school choice: developing a statistical framework for geodemographic analysis" *Population Research and Policy Review* **26** 553—579

Higher Education Act, 2004 *Public General Acts—Elizabeth II* chapter 8 (The Stationery Office, London)

Hobbs G, Vignoles A, 2007, " Is free school meal status a valid proxy for socio-economic status (in schools research)?", Department for Children, Schools and Families, London

Johnston R, 1979 *Political, Electoral and Spatial Systems* (Oxford University Press, Oxford)

Kelly K, Cook S, 2007, "Socio-economic class: a new widening participation measure in higher education", Department for Children, Schools and Families, London

Kitchen A, 1999, "The changing profile of entrants to mathematics at A Level and to mathematical subjects in higher education" *British Educational Research Journal* **25** 57 – 74

Leslie D, 2003, "Using success to measure quality in British higher education: which subjects attract the best-qualified students?" *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **166** 329 – 347

Longley P A, 2005, "Geographical information systems: a renaissance of geodemographics for public service delivery" *Progress in Human Geography* **29** 57 – 63

Longley P A, Goodchild M F, Maguire D J, Rhind D W, 2005 *Geographic Information Systems and Science* (John Wiley, Chichester, Sussex)

Machin S, Vignoles A, 2004, "Educational inequality: the widening socio-economic gap" *Fiscal Studies* **25** 107 – 128

Mastekaasa A, Smeby J, 2008, " Educational choice and persistence in male and female dominated fields " *Higher Education* **55** 189 – 202

OFFA, 2009, "Access agreement monitoring: outcomes for 2007 – 08", Office for Fair Access, Bristol

Reay D, David M, Ball S J, 2005 *Degrees of Choice: Social Class, Race, Gender and Higher Education* (Trentham Books, Stoke on Trent)

Rothman, K J 2002 *Epidemiology: An Introduction* (Oxford University Press, Oxford)

Seyan K, Greenhalgh T, Dorling D, 2004, "The standardised admission ratio for measuring widening participation in medical schools: analysis of UK medical school admissions by ethnicity, socioeconomic status, and sex " *British Medical Journal* **328** 1545 – 1546

Singleton A D, 2010 *Educational Opportunity: The Geography of Access to Higher Education* (Ashgate, Farnham, Surrey)

Singleton A, Longley P, 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

Storey P, Chamberlin R, 2001, "Improving the take up of free school meals." (Department for Education and Employment, London)

Styles B, 2008, "Moving on from free school meals: national census data can describe the socio-economic background of the intake of each school in England" *Educational Research* **50** 41 – 53

Swift M B, 2009, "Comparison of confidence intervals for a Poisson mean—further considerations" *Communications in Statistics – Theory and Methods* **38** 748 – 759

Tonks D G, Farr M, 1995, "Market segments for higher education using geodemographics" *Marketing Intelligence and Planning* **13** 24 – 33

Voas D, Williamson P, 2001, "The diversity of diversity: a critique of geodemographic classification" *Area* **33** 63 – 76

Webber R, Butler T, 2007, "Classifying pupils by where they live: how well does this predict variations in their GCSE results?" *Urban Studies* **44** 1229 – 1254