Using clustering techniques to identify localities with multiple health and social needs

Mark A. Bellis [a,*], Ian Jarman [b], Jenny Downing [a], Clare Perkins [a], Caryl Beynon [a],
Karen Hughes [a], Paulo Lisboa [b]

[a] Centre for Public Health, Liverpool John Moores University, Henry Cotton Campus, 15-21 Webster Street, Liverpool L3 2ET, UK
[b] Computing and Mathematical Sciences, Liverpool John Moores University, James Parsons Building, Bryom Street, Liverpool L3 3AF, UK

ARTICLE INFO

ABSTRACT

Development of health promoting policies requires an understanding not just of the interplay between different measures of health but also their relationship with broader education, criminal justice and other social issues. Methods to better utilise multi-sectoral data to inform policy are needed. Applying clustering techniques to 30 health and social metrics we identify 5 distinct local authority types, with poor outcomes for the majority of metrics concentrated in the same cluster. Clusters were distinguished especially by levels of: child poverty; breastfeeding initiation; children's tooth decay; teenage pregnancy; healthy eating; mental illness; tuberculosis and smoking deaths. Membership of the worst cluster (C5) was focused in Northern England which contains 15.7% of authorities analysed ($n = 324$), but 63.0% of those in C5. The concentration of challenges in certain areas creates disproportionate pressures that may exceed the cumulative effects of individual challenges. Such distinct health clusters also raise issues of transferability of effective policies between areas with different cluster membership.

1. Background

The presence of health inequalities between and within countries is not only a product of differential poverty and inequitable access to healthcare services, but is also associated with structural inequalities that affect the circumstances in which people live and work (Murray et al., 2007; Commission on Social Determinants of Health, 2008; Department of Health, 2009; Marmot, 2005, 2010). In even the most affluent countries, those who are richest experience lower rates of ill health and greater life expectancy than those who are poorest, with a disparate array of factors, for example social exclusion, nutritional adequacy and transport and employment policies, all playing their role (Wilkinson and Marmot, 2003). Consequently, the health status of a population should be of concern to policy makers in a number of sectors, not solely those working within health (Marmot, 2005). In addition to data describing ill health, public health practitioners and policy makers now frequently utilise behavioural, environmental, economic, educational and even criminal justice data (Stachenko, 2008; Bellis et al., 2005; Rutherford et al., 2007) in order to assess the health of the population.

Historically, academic analyses and policy responses have often followed a silo approach examining issues such as obesity, mental health, teenage pregnancy, violence and overall life expectancy separately, yet more recent work is considering the links between different issues (Felitti et al., 1998; Harden et al., 2009; Hillis et al., 2010; Duarte et al., 2010; Weiler et al., 2010; Barton and Pretty, 2010). Thus, it is often the combination of these factors that affects the well-being of different communities, with interactions between different health and social problems impeding the effectiveness of potential interventions (e.g. encouraging walking and other outdoor exercise in areas with high levels of violence; Bellis et al. 2008).

While the need to base policy changes on scientific evidence is unquestionable, increasing the breadth of health and social measures calls for an integrated approach to their analysis and interpretation, adding understanding rather than just additional complexity to local health strategies. Thus, relationships between different metrics should expose links between objectives of different stakeholders and consequently inform joint working. Furthermore, broader public health metrics can identify localities with similar public health profiles that either require comparable external support or may usefully work together and learn from each other's actions. Examples of such processes at the international level include: the use of multi-sectoral data to identify globally the least developed countries (Klugman, 2010); clustering techniques on sub-national regions of the European Union (I2SARE, 2010) to establish areas with similar health status; and clustering of smaller populations within a country

for largely socio-economic purposes (Office for National Statistics, 2001; Vickers and Rees, 2007). More specifically profile matching has also been used to inform comparisons and develop links between areas with similar profiles from a judicial perspective (e.g. Association of Police Authorities, 2006).

In 2005, the Association of Public Health Observatories produced (for the Department of Health, England) a set of local authority public health metrics for England (Association of Public Health Observatories, 2006). Production followed a systematic process to identify what were both important data sets relating to public health and ones that could be provided with reliable quality and regularity at a local administrative level. Available as single profiles for each local authority ($n=324$), the latest iterations (2010) comprise 31 data items; each calculated according to quality assured, published and peer reviewed methodologies. Measures include health risks and protective behaviours, morbidity and mortality as well as data relating to crime, education, economics and the environment (Department of Health and the Association of Public Health Observatories, 2010). Standard clustering techniques provide objective methodologies for combining public health metrics to better identify discrete groups of localities with similar public health profiles and to establish relationships between trends in each metric which can be visualised by cohort (Friedman and Rubin, 1967, Lisboa et al., 2008). Here we use an established clustering methodology (K-means) on the public health metrics calculated for the 2010 local authority profiles in order to explore how a range of factors across the life course affect areas in England differently. We identify how clustering patterns relate to each public health metric and to independent established measures of inequality (i.e. disposable income and deprivation). Finally, we discuss the benefits of this approach in the development of national policy and local multi-agency working.

2. Methods

Full details of the calculation, selection and quality assurance processes used in the production of the health profiles metrics are described in detail elsewhere (Department of Health and the Association of Public Health Observatories, 2010). Of 31 metrics in the health profiles only IMD (Index of Multiple Deprivation) was excluded as it is a broad composite measure containing elements of other metrics in the profiles. All data items in the health profiles were examined for missing data and appropriateness for cluster analysis. From the 30 health profile metrics (see footnote Table 1 for a description of each metric) missing values were present in five variables (smoking in pregnancy [8.6%], breast-feeding initiation [6.8%], statutory homelessness [6.5%], children's tooth decay [4.3%] and new cases of tuberculosis [1.2%]) and missing cells in each metric were imputed to the corresponding metric median.

Data for each metric were re-scaled to the same range (0 and 100) so that no single indicator dominated the clustering process. The standardised metrics were then clustered using K-means clustering. Initially, clustering randomly generates a small number of cluster centres, defined by values (i.e. coordinates) for each of the health profile data variables used. All local authorities are then assigned to their closest cluster centre and the centre point redefined as the mean value for each of the health profiles data variables across the local authorities in that cluster. An iterative process then continues to re-assign all points to the nearest of the updated cluster centres and again recalculates the centre point values. The process is ended by reaching a point where no further changes in assignment to cluster are seen. By also repeating the whole process starting with a range of different numbers of initial cluster centres and values for starting clusters, the most robust

clustering patterns with greatest differentiation between clusters were identified (Bacciu et al., 2009). The statistical analyses used SPSS (release 17.0.0; SPSS, 2008) and MATLAB (version 7.8.0 R2009a; MATLAB, 2009) 64bit for clustering.

Differences in metrics were measured using a test of medians across all clusters. In order to visualise local authority clusters, data are presented as a scatter plot (Fig. 1). Here x and y axes (labelled Principal Separating Axes; PSAs) are combinations of the re-scaled health profile variables used in the clustering process. The PSAs are comparable to the principal components generated in principal component analysis (Jolliffe, 2002) but here, PSAs are selected on the basis of showing the maximum variance between clusters in order to show their separation in a two dimensional plot (Lisboa et al., 2008). In order to examine the stability of local authority cluster membership (sensitivity analysis), clustering was repeated using subsets of the health profile variables and the contribution of each health profile variable to each PSA, and hence to cluster separation, was examined (Lisboa et al., 2008).

IMD was later used in additional analyses after clustering had been undertaken in order to explore its relationship with measures of cluster separation. Gross disposable household income was also used for these purposes. Gross disposable household income is available at NUTS (Nomenclature of Units for Territorial Statistics) geography for England at level 3 (defined as counties or groups of unitary authorities, $n=92$; Office for National Statistics, 2009). Median gross disposable household income per NUTS3 area was used for each local authority nested within a NUTS3 area. IMD and gross disposable household income were chosen to explore clustering as a robust composite and discrete measure of inequality respectively. Data are also presented graphically using local authority and regional boundaries as well as using hexagon mapping where hexagon areas relate to the population of each local authority; Birmingham local authority, for example, is represented by 11 hexagons, while Rutland, a much smaller local authority is represented by half a hexagon (Map 1a, b; Thomas and Dorling, 2004). For each local authority, population sizes and the ethnic make-up of the population were taken from the Office for National Statistics (Office for National Statistics, 2008; Office for National Statistics, 2011).

3. Results

Fig. 1 shows a two dimensional visualisation that identifies five discrete clusters (C1 to C5) by plotting each local authority on the two composite measures of health profile variables calculated to show maximum cluster separation (see Methods). In this scatter plot, all clusters separate along the first separating axis PSA1. Clusters 1 to 5 represent 11.3, 12.3, 11.5, 5.3 and 10.7 million individuals respectively (based on mid-2008 local authority population estimates; Office for National Statistics, 2008).

Correlation of all metrics with PSA1 (from Fig. 1) is presented in Table 1 with deaths from smoking, teenage pregnancy, life expectancy for males and females, early deaths from heart disease and stroke, incapacity benefits for mental illness, children in poverty, early deaths from cancer and adults who smoke all strongly correlated with the axis ($R > 0.8$ or $R < -0.8$). Each local authority's value for PSA1 was also correlated with measures of inequality; IMD ($R=-0.85$, $P < 0.001$) and gross disposable household income ($R=0.65$; $P < 0.001$).[1] PSA2 primarily distinguishes C4 with 95.5% (21/22) of the local authorities within C4 being based in London (Map 1a).

---

[1] Five local authorities were excluded from the analysis as local authority to NUTS3 matching resulted in them being extreme outliers in the correlation.

Table 1

Cluster medians per metric ordered best to worst and correlation coefficients with Principal Separating Axis 1.

Indicator name	R (PSA1)	←Best outcome		Median (+/−IQR)		Worst outcome→
1 Deaths from smoking	−0.925	151.9 (140.5–163.8)	182.9 (169.5–193.1)	219.1 (185.6–257.2)	227.0 (216.1–238.3)	281.3 (260.4–301.0)
2 Teenage pregnancy (under 18)	−0.921	23.8 (21.0–28.0)	35.1(31.3–38.5)	46.3 (41.5–50.4)	48.0 (40.3–59.2)	53.9 (49.0–58.4)
3 Life expectancy—Male	0.908	79.9 (79.3–80.4)	78.8 (78.4–79.3)	77.5 (75.9–78.7)	77.2 (76.8–77.7)	75.7 (75.3–76.2)
4 Early deaths: heart disease and stroke	−0.889	55.9 (51.3–59.8)	65.4 (59.8–70.4)	79.7 (75.3–85.9)	90.1 (82.0–100.1)	97.0 (89.7–105.9)
5 Incapacity benefits for mental illness	−0.876	15.2 (13.4–17.6)	21.9 (17.9–25.6)	28.9 (25.1–35.7)	29.5 (24.1–33.0)	40.7 (34.6–47.7)
6 Children in poverty	−0.866	10.8 (8.9–12.6)	15.7 (13.3–18.1)	21.9 (19.2–24.5)	29.6 (24.7–33.3)	38.7 (31.3–44.9)
7 Early deaths: cancer	−0.857	98.5 (94.6–103.4)	108.1 (104.1–112.9)	113.1 (106.1–127.4)	119.7 (115.7–123.4)	134.5 (127.4–143.8)
8 Adults who smoke	−0.845	16.7 (15.3–18.4)	20.5 (18.8–22.1)	21.8 (18.6–27.0)	24.4 (22.4–26.5)	28.0 (26.0–29.9)
9 Life expectancy—Female	0.838	83.5 (83.1–83.9)	82.5 (82.0–83.1)	82.3 (81.2–83.1)	81.5 (81.0–82.0)	80.1 (79.6–80.6)
10 Hospital stays for alcohol-related harm	−0.791	1167.2 (1004.8–1334.2)	1339.9 (1218.0–1523.3)	1596.5 (1466.2–1852.9)	1766.1 (1444.4–1897.0)	2113.7 (1904.3–2388.3)
11 Violent crime	−0.687	9.2 (7.6–12.2)	12.8 (10.6–15.2)	17.1 (14.9–21.7)	18.1 (14.8–23.0)	27.1 (23.9–30.7)
12 GCSE achievement	0.671	58.3 (54.9–62.6)	52.0 (46.8–54.6)	50.3 (47.7–55.3)	46.4 (43.0–51.7)	44.9 (41.0–47.9)
13 Healthy eating adults	0.670	36.9 (33.6–39.3)	32.0 (29.7–33.6)	29.6 (27.9–31.6)	25.4 (23.7–27.8)	22.5 (21.1–24.5)
14 Children's tooth decay (at age 5)	−0.659	0.7 (0.5–0.9)	0.9 (0.7–1.1)	1.1 (0.8–1.4)	1.2 (1.0–1.8)	1.7 (1.4–1.9)
15 Infant deaths	−0.623	3.3 (2.4–4.2)	4.4 (3.5–4.9)	5.0 (4.2–5.7)	5.1 (4.0–5.8)	5.8 (5.0–6.4)
16 Obese children	−0.617	7.6 (6.8–8.5)	9.1 (8.3–9.8)	9.5 (8.8–10.6)	10.2 (9.2–11.5)	12.1 (11.3–13.3)
17 People diagnosed with diabetes	−0.608	3.7 (3.5–4.0)	4.2 (4.0–4.6)	4.4 (4.1–4.7)	4.5 (3.9–5.1)	4.7 (4.3–5.0)
18 Physically active adults	0.564	13.0 (11.7–14.2)	11.2 (10.3–12.4)	10.5 (9.5–11.8)	10.3 (9.4–11.0)	9.2 (7.4–11.1)
19 Obese adults	−0.523	21.2 (18.4–23.4)	22.5 (20.5–23.5)	24.8 (23.2–26.1)	25.4 (23.6–27.3)	26.8 (25.1–28.4)
20 Breastfeeding initiation	0.475	88.0 (83.8–90.6)	76.1 (73.3–79.9)	73.4 (71.7–76.5)	71.9 (66.3–74.3)	59.7 (54.1–67.3)
21 Incidence of malignant melanoma	0.416	5.5 (4.4–7.4)	10.3 (8.6–11.8)	12.4 (10.7–14.2)	13.6 (10.8–15.7)	14.8 (11.7–18.4)
22 New cases of tuberculosis	−0.414	3.5 (2.5–6.0)	4.8 (2.7–8.3)	6.5 (4.1–12.1)	9.8 (4.0–23.9)	50.8 (38.3–62.2)
23 Smoking in pregnancy	−0.408	6.9 (5.1–8.9)	13.3 (8.8–15.1)	15.0 (13.1–17.5)	15.4 (14.0–18.3)	20.8 (18.3–24.5)
24 Road injuries and deaths	0.391	41.8 (36.6–51.7)	43.7 (35.2–56.6)	47.6 (42.5–58.3)	52.7 (40.9–60.9)	60.3 (48.1–75.3)
25 Statutory homelessness	−0.376	1.3 (0.6–1.7)	1.7 (1.1–2.2)	2.0 (1.4–3.0)	2.0 (1.0–3.8)	3.9 (2.7–4.7)
26 Binge drinking adults	−0.327	14.1 (10.5–19.6)	18.3 (16.5–20.8)	18.7 (16.0–21.5)	20.9 (18.8–23.0)	24.4 (20.4–27.3)
27 Hip fracture in over 65 s	−0.264	438.5 (409.0–476.1)	449.8 (415.0–487.7)	462.9 (427.0–514.4)	487.5 (460.2–526.1)	523.8 (489.4–549.4)
28 Carbon emissions	0.246	5.4 (4.7–6.3)	6.1 (5.6–7.5)	6.6 (6.1–7.4)	6.9 (6.0–8.1)	7.3 (6.2–8.3)
29 Physically active children	0.157	51.30 (46.2–56.3)	51.25 (46.6–56.9)	48.9 (43.9–55.4)	48.5 (43.7–55.8)	48.4 (45.0–52.8)
30 Excess winter deaths	0.147	13.7 (11.9–17.6)	14.2 (11.6–16.7)	15.5 (12.4–18.3)	16.3 (12.7–18.9)	16.3 (13.2–19.3)

Cluster 1 ▮ Cluster 2 ▮ Cluster 3 ▮ Cluster 4 ▮ Cluster 5 ▮

1 Deaths attributable to smoking, directly age standardised rate, 35 years +, 2006–08. **2** Under-18 conception rate per 1000 females aged 15–17 (crude rate) 2006–08 provisional. **3** Life expectancy at birth, years, all ages, 2006–08, males. **4** Mortality from all circulatory diseases, directly age-standardised rate, per 100,000 population under 75, 2006–08. **5** Claimants of incapacity benefit with mental or behavioural disorders, crude rate per 1000 working age population 2008. **6** Percentage of children living in families receiving means-tested benefits, 2007 (part of Indices of Deprivation 2007). **7** Mortality from all cancers, directly age-standardised rate per 100,000 population under 75, 2006–08. **8** Percentage of adults who smoke, modelled estimate using Health Survey for England, 2006–2008. **9** Life expectancy at birth, years, all ages, 2006–08, females. **10** Hospital admissions for alcohol-related harm 2008–09, directly age and sex standardised rate per 100,000 population. **11** Recorded violence against the person offences, crude rate per 1000 population 2008–09. **12** Pupils achieving 5 or more GCSEs, including maths and English, at grades A*–C, percentage at end of Key Stage 4 in Local Education Authority schools 2008–09. **13** Percentage of healthy eating adults, modelled estimate using Health Survey for England 2006–08. **14** Mean number of teeth per child sampled (at age 5) which were either actively decayed, missing or filled 2007–08. **15** Infant deaths, crude rate per 1,000 live births 2006–08. **16** Percentage of obese school children in reception year 2008–09. **17** Percentage of people on GP registers with a recorded diagnosis of diabetes 2008–09. **18** Percentage of physically active adults aged 16 and over 2008–09. **19** Percentage of obese adults, modelled estimate using Health Survey for England 2007–08. **20** Percentage of mothers initiating breast feeding where status is known 2008–09. **21** Directly age standardised rate of malignant melanoma incidence per 100,000 population over age 75 2004–06. **22** Crude rate of tuberculosis (new cases) per 100,000 population 2006–08. **23** Percentage of mothers smoking in pregnancy where status is known 2008–09. **24** People killed or seriously injured on the roads, rate per 100,000 population 2006–08. **25** Statutory homeless households, crude rate per 1000 households 2008–09. **26** Percentage of binge drinking adults, modelled estimate using Health Survey for England 2007–08. **27** Directly age-standardised rate of hip fractures per 100,000 for emergency admission 2008–09. **28** Total end user CO_2 emissions per capita (tonnes CO_2 per resident) 2007. **29** Percentage of year 1–13pupils who spend at least 3 hours per week on high quality PE and school sport 2008–09. **30** Ratio of excess winter deaths (observed winter deaths minus expected deaths based on non-winter deaths) to average non-winter deaths 01.08.05–31.07.08. R(PSA1)=Correlation coefficient (Spearman's) for each metric with Principal Separating Axis 1.

Examination of both PSA1 and PSA2 against health profile variables identifies that the main contributions to the separating axes (i.e. the health profile variables with the largest loading in absolute value in either PSA1 or PSA2) were children in poverty; breastfeeding initiation; children's tooth decay; teenage pregnancy; healthy eating adults; mental illness; new cases of tuberculosis and deaths from smoking. Repeating the clustering process using only this subset of indicators results in just 12.35% of local authorities changing their cluster membership. Exclusion of increasing proportions of this subset of variables incrementally alters the clustering solution.

Table 1 shows the median and inter-quartile ranges by metrics calculated for each cluster and orders them from best to worst well-being outcome (by median). The cell colours in Table 1 represent the cluster allocation; for metrics one to 12, the best outcome is observed within cluster one, while for the thirteenth metric (healthy eating adults), the best outcome falls within cluster four, and so on. For metrics, differences across all cluster medians per variable are significant ($P < 0.001$), with the exception of physically active children ($P=0.072$). For the vast majority of indicators, clusters align with C1 (most affluent) having the best outcomes and C5 (most deprived) the worst. Thus, C5 has the worst median value for

20 of the public health metrics and C1 the best for 19 (Table 1). C4 (mainly London) is anomalous as, despite it having relatively high to moderate deprivation, it has the best median values for nine public health metrics (e.g. breast feeding initiation, obese adults and excess winter deaths) and the poorest for five (e.g. violent crime, children in poverty). C4 is also distinguished by a very different ethnic composition (C1 to C5 percentage white; 92.6%, 93.0, 91.3%, 64.5% and 86.6% respectively, mid-year 2007). The North West and North East account for 63.0% of all the local authorities in the worst cluster C5, but only 15.7% of all local authorities in England included in the profiles, while the East of England and South East account for 35.2% of all local authorities but 58.2% of all those in the best public health cluster (C1, Map 1a, b).

Fig. 1. Two dimensional visualisation of health profile clusters against Principal Separating Axis 1 and 2.
All local authorities for which health profiles were developed ($n=324$) were included in the clustering analysis. Health profiles were not developed for City of London and Isles of Scilly.

4. Discussion

Using a K-means methodology we have identified five distinct public health clusters at a local administrative level (here, local authorities). Despite the wide range of metrics used the vast majority correlate strongly with a single separating axis (PSA1, Fig. 1) representing strong interrelations between metrics. Identifying strong relationships between different public health measures is not new (Bellis et al., 2011; Schuit et al., 2002; Poulton et al., 2002). However, that such a wide range of measures can be refined into a set of clustered public health geographies strongly supports the need for multi-disciplinary approaches to health and well-being; especially in those areas facing the greatest public health challenges (e.g. C5). Moreover, that aggregated public health metrics correlate strongly with composite deprivation (IMD) and even a single measure of affluence (gross disposable household income) reinforces ongoing national and international attention on inequalities and the need to address underlying determinants of poor health (Commission on Social Determinants of Health, 2008; Department of Health, 2010a; Marmot, 2010).

Differences presented here between C4 and other clusters support previous research which used clustering techniques to classify areas of the UK according to multiple measures of the physical environment. These also identified a unique cluster which incorporated most of London (Richardson et al., 2010; Shortt et al., 2011). Together these findings demonstrate: the unique combination of public health and environmental challenges faced by administrations in a multi-cultural megacity (comprised of more than 10 million people); the need for bespoke solutions in such settings; and the potential dangers of extrapolating solutions developed in such cities to other settings. Indeed, the need to identify and address London administrations as demographically and economically different entities has been raised outside of public health (Vickers and Rees, 2007). Equally, however, the most deprived communities (North West and North East) are also faced with unique combinations of poor public health with over 60% of C5 administrations falling within their boundaries. Results identify that individuals in such administrations face a lifelong public health deficit which starts
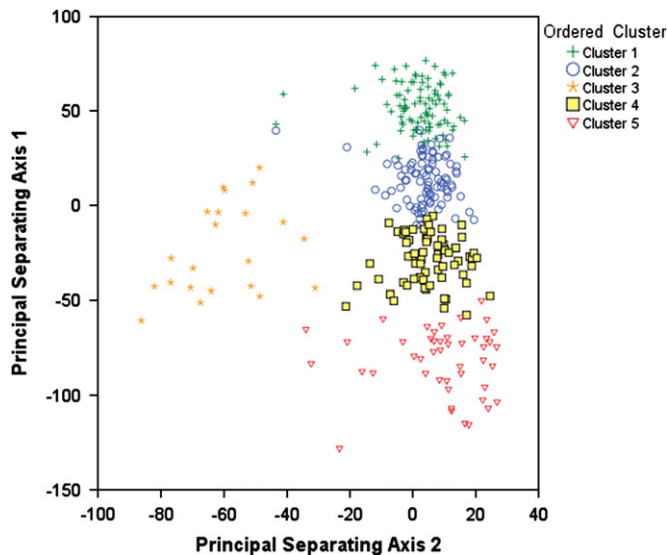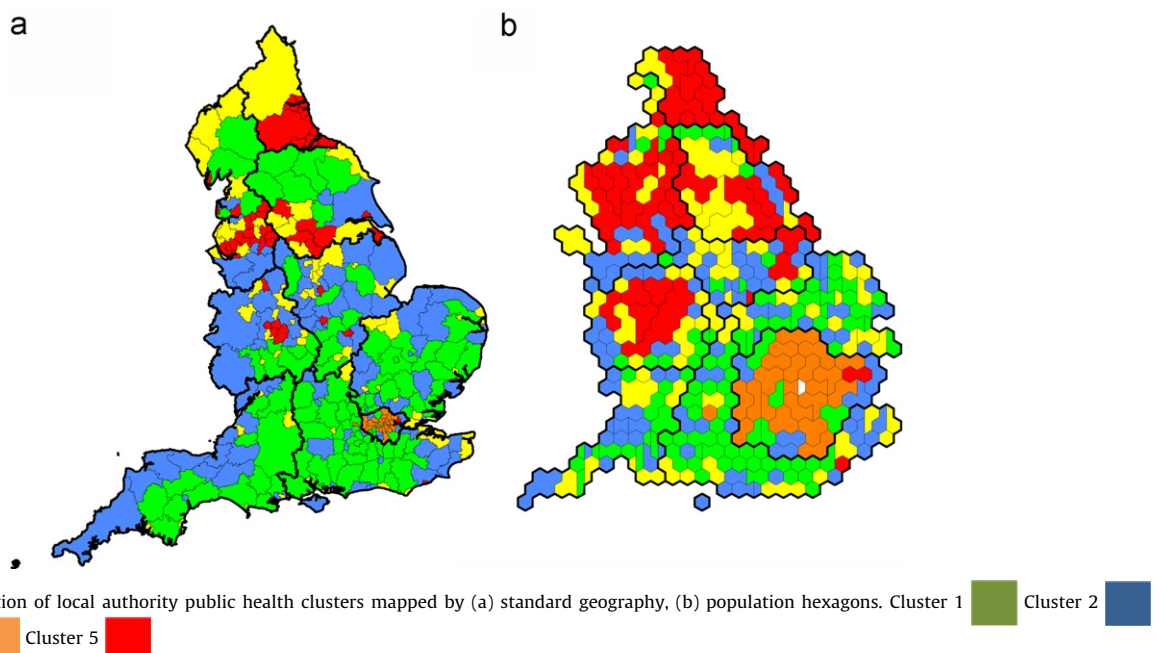

Map 1. Geographic distribution of local authority public health clusters mapped by (a) standard geography, (b) population hexagons. Cluster 1 ▮ Cluster 2 ▮ Cluster 3 ▮ Cluster 4 ▮ Cluster 5 ▮

Bold black lines represent regional boundaries and thin lines local authorities. Population hexagons resize and represent local authorities in hexagons according to the populations that each represent (Thomas and Dorling, 2004).

from smoking in pregnancy, low breast feeding initiation, and early tooth decay through poor school performance and teenage pregnancy to, as adults, obesity, smoking, alcohol misuse, poorer mental and physical health and ultimately early death. Consequently these results support the need for a life course approach to improving public health (Lynch and Davey Smith, 2005; Braveman and Barclay, 2009; Department of Health, 2010a). A child's options in later life appear largely constrained by factors relating to where they grew up (e.g. schools, job opportunities, housing opportunities; Dorling et al., 2001). Indeed increasing evidence suggests that the origins of much adult ill health are routed in exposure to poor social and environmental conditions during childhood with, for instance, early exposure to abuse and neglect predisposing individuals to behaving anti-socially and higher risks of substance abuse, depression and even cancer (Felitti et al., 1998; Anda et al., 2006).

The individual relationships of poverty with each of the health profile metrics are well studied but lifelong reinforcement and complex interaction between each measure is poorly understood. However, public health measures to increase physical activity in adults or children are impeded by local violence restricting people to their houses (Bellis et al., 2008), access to health information are hindered by high levels of illiteracy (Sanders et al., 2009) and efforts to improve parenting are affected by poor levels of parental health (Waylen and Stewart-Brown, 2010). Even recovery from diseases such as cancer are inevitably complicated by the general health of the patient and the community and family support on which they can draw (Falagas et al., 2007; Kroenke et al., 2006; Woods et al., 2006). As a result, the combined effects of multiple challenges to public health are potentially greater than those related to each individually, as people's and communities' abilities to cope with poor well-being are overwhelmed. Consequently, those areas with C5 profiles not only require more study of complex interactions between threats to well-being but may also require support that grows geometrically rather than in direct proportion to each of their public health challenges.

A variety of methodologies exist to define characteristics at an individual level, traditionally developed for social marketing or health purposes (e.g. P2 People and Places, Mosaic, Acorn; Dedman et al., 2006). However, such geodemographic tools typically segment individuals into groups with different characteristics rather than profile and cluster localities at geographies responsible for public health policy. Nationally, clustering techniques offer a mechanism for objectively identifying areas facing the greatest public health challenges and provide an opportunity for similar local authorities to exchange information on what works and what does not. In England, the Police have also adopted methods of comparisons between similar geographies in order to make them both more meaningful and realistic when setting local objectives (Association of Police Authorities, 2006). Moreover, these methods can help avoid the need of a locality being measured by a single or small set of variables that may describe its overall status in either an overly positive or negative fashion.

For England, combining health data (e.g. on smoking, alcohol intake and obesity) with data more traditionally the focus of local authorities (e.g. violent crime and carbon emissions) in a way which is more easily interpreted is timely. After nearly 40 years of local public health responsibilities sitting with health bodies (since being transferred from local authorities in 1974; Draper, Glenhorn and Best, 1976), the government is set to transfer responsibility for public health back to local authorities (Department of Health, 2010b). A better understanding of how different public health metrics correlate and cluster in specific localities should help both local priority setting and national policy development. Such intelligence could be used to refine resource allocations to and by local

administrations, taking into account not just the individual health and social issues identified by each health profile variable but also additional requirements relating to their interactions.

The public health metrics used here were independently identified and quality assured and their methods are peer reviewed for their inclusion in national Health Profiles (Department of Health and the Association of Public Health Observatories, 2010). However, as ecological measures they inevitably hide inequalities at smaller geographies within each authority and the validity of each measure in any area depends on the accuracy of data collection and appropriate collection and use of population denominators. The data sets used here also included different years but always represented the most current available at analysis. Other metrics are obtainable and other methodological approaches could have been applied even to the data sets used. However, the strong correlation between metrics suggests that other methods and the addition or substitution of other metrics would still support the concentration of multiple poor health outcomes into a specific cluster of local authorities. Any long term adoption of clustering methodologies to inform public health measures would have to explore the frequency at which clusters should be re-examined in order to map changes in the status of different administrative areas.

5. Conclusions

The clustering analysis used here identified how a broad range of different freely available measures relating to public health are concentrated into specific localities. The effects of such concentrations are not well studied but it is likely that interactions between such challenges to health and well-being will contribute to disproportionate pressures on communities, services and other resources. Such interactions are not typically reflected in policies determined by individual government departments but require a cross-departmental consideration of the general environment in which health, social, educational, economic or judicial policies are expected to be implemented. To address the clusters of public ill health (Table 1), local public health leaders (e.g. Directors of Public Health) require influence across a range of public service responsibilities; both to implement health service-based interventions and to align other systems, including schools, transport, physical activity and social care, with public health priorities. The national counterparts of such local leaders require the same influence across other government departments. Critically, such developments should recognise that interventions successful in clusters with relatively few public health problems (e.g. C1 or C2) may not successfully translate into, for instance, C5 type areas or may require considerably more resource in order to achieve comparable results. However, clustering approaches may provide public health with a mechanism to identify areas between which policy and interventions may successfully be replicated. With an increasing array of intelligence pertinent to health and well-being, such mechanisms help move from considering disparate metrics to public health syndromes which describe the collections of challenges faced by different communities and help engage a wider range of stakeholders in finding solutions.

References

Anda, R.F., Felitti, V.J., Bremner, D., Walker, J.D., Whitfield, C.H., Perry, B.D., Dube, S.H.R., Giles, W.H., 2006. The enduring effects of abuse and related adverse experiences in childhood. European Archives of Psychiatry and Clinical Neuroscience 256, 174–186.
Association of Police Authorities, 2006. iQuanta: a police authority guidesecond ed. Association of Police Authorities, London.
Association of Public Health Observatories, 2006. Health Profiles Project: Project brief. Association of Public Health Observatories, York.

Bacciu, D., Jarman, I.H., Etchells, T.A., Lisboa, P.J.G., 2009. Patient stratification with competing risks by multivariate Fisher distance. International Joint Conference on Neural Networks, 213–220.

Barton, J., Pretty, J., 2010. What is the best dose of nature and green exercise for improving mental health? A multi-study analysis. Environmental Science and Technology 44, 3947–3955.

Bellis, M.A., Hughes, K., Tocque, K., Hennell, T., Humphrey, G., Wyke, S., 2005. Assessing and communicating the health and judicial impact of alcohol use. Public Health 119, 253–261.

Bellis, M.A., Hughes, K., Anderson, Z., Tocque, K., Hughes, S., 2008. Contribution of violence to health inequalities in England: demographics and trends in emergency hospital admissions for assault. Journal of Epidemiology and Community Health 62, 1064–1071.

Bellis, M.A., Hughes, K., Wood, S., Wyke, S., Perkins, C., 2011. National five-year examination of inequalities and trends in emergency hospital admission for violence across England. Injury Prevention 10.1136/ip.2010.030486.

Braveman, P., Barclay, C., 2009. Health disparities beginning in childhood: a life-course perspective. Pediatrics 124, S163–S175.

Commission on Social Determinants of Health, 2008. Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health. World Health Organization, Geneva.

Dedman, D., Jones, A., Tocque, K., Bellis, M.A., 2006. Population targeting: Tools for social marketing. Synthesis Report, Issue 6. Liverpool: North West Public Health Observatory.

Department of Health, 2009. Tackling health inequalities: 10 Years On—A review of developments in tackling health inequalities in England over the last 10 years. Department of Health, London.

Department of Health, 2010a. Healthy lives, healthy people: Transparency in outcomes. Department of Health, London.

Department of Health, 2010b. Equity and excellence: Liberating the NHS. The Stationary Office, London.

Department of Health and the Association of Public Health Observatories, 2010. The indicator guide Health Profiles 2010. Department of Health, London.

Dorling, D., Smith, G., Noble, M., Wright, G., Burrows, R., Bradshaw, J., Joshi, H., Pattie, C., Mitchell, R., Green, A.E., McCulloch, A., 2001. How much does place matter? Environment and Planning A 33, 1335–1369.

Draper, P., Glenhorn, G., Best, G., 1976. The organization of health care: A critical review of the 1974 reorganization of the National Health Service. in: Tuckett, D. (Ed.), An Introduction to Medical Sociology Travistock Publications Ltd., London, pp. 254–289.

Duarte, C.S., Sourander, A., Nikolakaros, G., Pihlajamaki, H., Helenius, H., Piha, J., Kumpulainen, K., Moilanen, I., Tamminen, T., Almqvist, F., Must, A., 2010. Child mental health problems and obesity in early adulthood. The Journal of Pediatrics 156, 93–97.

Falagas, M.E., Zarkadoulia, E.A., Ioannidou, E.N., Peppas, G., Christodoulou, C., Rafailidis, P.R., 2007. The effect of psychosocial factors on breast cancer outcome: a systematic review. Breast Cancer Research 9, R44.

Felitti, V.J., Anda, R.F., Nordenberg, D., Williamson, D.F., Spitz, A.M., Edwards, V., Koss, M.P., Marks, J.S., 1998. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. American Journal of Preventive Medicine 14, 245–258.

Friedman, H.P., Rubin, J., 1967. On some invariant criteria for grouping data. Journal of the American Statistical Association 62, 1159–1178.

Harden, A., Brunton, G., Fletcher, A., Oakley, A., 2009. Teenage pregnancy and social disadvantage: systematic review integrating controlled trials and qualitative studies. British Medical Journal 339, b4254.

Hillis, S.D., Anda, R.F., Dube, S.R., Felitti, V.J., Marchbanks, P.A., Macaluso, M., Marks, J.S., 2010. The protective effect of family strengths in childhood against adolescent pregnancy and its long-term psychosocial consequences. The Permanente Journal 14, 18–27.

I2SARE, 2010. Development of a regional typology—aim, methods and results. Available at : ⟨http://www.i2sare.eu/Default.aspx⟩ Accessed on 1st June 2011.

Jolliffe, I.T., 2002. Principal Component Analysis second ed. Springer, New York.

Klugman, J., 2010. Human development report 2010. The real wealth of nations: pathways to human development. United Nations Development Programme, New York.

Kroenke, C.H., Kubzansky, L.D., Schernhammer, E.S., Holmes, M.D., Kawachi, I., 2006. Social networks, social support and survival after breast cancer diagnosis. Journal of Clinical Oncology 24, 1105–1111.

Lisboa, P., Ellis, I., Green, A., Ambriogi, F., Dias, M., 2008. Cluster-based visualisation with scatter matrices. Pattern Recognition 29, 1814–1823.

Lynch, J., Davey Smith, G., 2005. A life course approach to chronic disease. Annual Review of Public Health 26, 1–35.

MATLAB version 7.8.0. R2009a, 2009. Natick, Massachusetts: The MathWorks Inc.

Marmot, M., 2005. Social determinants of health inequalities. The Lancet 365, 1099–1104.

Marmot, M., 2010. Fair society, healthy lives. Strategic review of health inequalities in England post-2010. The Marmot review final report. University College London, London.

Murray, C.J.L., Lopez, A.D., Black, R., Mathers, C.D., Shibuya, K., Ezzati, M., Salomon, J.A., Michaud, C.M., Walker, N., Vos, T., 2007. Global burden of disease 2005: call for collaborators. The Lancet 370, 109–110.

Office for National Staticis, 2001. National statistics 2001 area classification. Available at: ⟨www.statistics.gov.uk/about/methodology_by_theme/area_classification/default.asp⟩ Accessed on 9th June 2011.

Office for National Statistics, 2008. Mid-year population estimates. Availabe at: ⟨http://www.statistics.gov.uk/statbase/product.asp?vlnk=15106⟩ Accessed on 14th February 2011.

Office for National Statistics, 2009. Regional household income. Available at: ⟨http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14651⟩ Accessed on 14th February 2011.

Office for National Statistics, 2011. Population estimates for ethnic group (experimental), 2011. Available at: ⟨http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14238&More=Y⟩ Accessed on 14th February 2011.

Poulton, R., Caspi, A., Milne, B.J., Thomson, W.M., Taylor, A., Sears, M.R., Moffitt, T.E., 2002. Association between children's experience of socioeconomic disadvantage and adult health: a life-course study. The Lancet 360, 1640–1645.

Richardson, E.A., Mitchell, R., Shortt, N.K., Pearce, J., Dawson, T.P., 2010. Developing summary measures of health-related multiple environmental deprivation for epidemiological research. Enviornment and Planning A 42, 1650–1658.

Rutherford, A., Zwi, A.B., Grove, N.J., Butchart, A., 2007. Violence: a priority for public health? (part 2). Journal of Epidemiology and Community Health 61, 764–770.

Sanders, L.M., Federico, S., Klass, P., Abrams, M.A., Dreyer, B., 2009. Literacy and child health: a systematic review. Archives of Pediatrics and Adolescent Medicine 163, 131–140.

Schuit, A.J., van Loon, A.J.M., Tijhuis, M., Ocké, M.C., 2002. Clustering of lifestyle risk factors in a general adult population. Preventive Medicine 35, 219–224.

Shortt, N.K., Richardson, E.A., Mitchell, R., Pearce, J., 2011. Re-engaging with the physical environment: a health related environmental classification of the UK. Area 43, 76–87.

SPSS for windows, release 17.0.0., 2008. Chicago: SPSS Inc.

Stachenko, S., 2008. Challenges and opportunites for surveillance data to inform public health policy on chronic non-communicable diseases: Canadian perspectives. Public Health 122, 1038–1041.

Thomas, B., Dorling, D., 2004. Advances in the human cartography of the UK. Cartographic Journal 41, 109–115.

Vickers, D., Rees, P., 2007. Creating the UK national statistics 2001 output area classification. Journal of the Royal Statistical Society, Series A 170, 379–402.

Waylen, A., Stewart-Brown, S., 2010. Factors influencing parenting in early childhood: a prospective longitudinal study focusing on change. Child: Care, Health and Development 36, 198–207.

Weiler, R., Stamatakis, E., Blair, S., 2010. Should health policy focus on physical activity rather than obesity? Yes.. British Medical Journal 340, c2603.

Wilkinson, R., Marmot, M., 2003. The solid facts Second Edition World Health Organization, Geneva.

Woods, L.M., Rachet, B., Coleman, M.P., 2006. Origins of socio-economic inequalities in cancer survival: a review. Annals of Oncology 17, 5–19.