Does geodemographic segmentation explain differences in route of cancer diagnosis above and beyond person-level sociodemographic variables?

C.J. Bright[1], C. Gildea[1], J. Lai[1], L. Elliss-Brookes[1], G. Lyratzopoulos [1,2]

[1]National Cancer Registration and Analysis Service (NCRAS), Public Health England, Wellington House, London SE1 8UG, UK
[2]Epidemiology of Cancer Healthcare and Outcomes (ECHO) Group, Department of Behavioural Science and Health, Institute of Epidemiology & Health Care (IEHC), University College London, 1-19 Torrington Place, London WC1E 7HB, UK

Address correspondence to Chloe J Bright, E-mail: chloe.bright@phe.gov.uk

ABSTRACT

Background Emergency diagnosis of cancer is associated with poorer short-term survival and may reflect delayed help-seeking. Optimal targeting of interventions to raise awareness of cancer symptoms is therefore needed.

Methods We examined the risk of emergency presentation of lung and colorectal cancer (diagnosed in 2016 in England). By cancer site, we used logistic regression (outcome emergency/non-emergency presentation) adjusting for patient-level variables (age, sex, deprivation and ethnicity) with/without adjustment for geodemographic segmentation (Mosaic) group.

Results Analysis included 36 194 and 32 984 patients with lung and colorectal cancer. Greater levels of deprivation were strongly associated with greater odds of emergency presentation, even after adjustment for Mosaic group, which nonetheless attenuated associations (odds ratio [OR] most/least deprived group = 1.67 adjusted [model excluding Mosaic], 1.28 adjusted [model including Mosaic], $P < 0.001$ for both, for colorectal; respective OR values of 1.42 and 1.18 for lung, $P < 0.001$ for both). Similar findings were observed for increasing age. There was large variation in risk of emergency presentation between Mosaic groups (crude OR for highest/lowest risk group = 2.30, adjusted OR = 1.89, for colorectal; respective values of 1.59 and1.66 for lung).

Conclusion Variation in risk of emergency presentation in cancer patients can be explained by geodemography, additional to deprivation group and age. The findings support proof of concept for public health interventions targeting all the examined attributes, including geodemography.

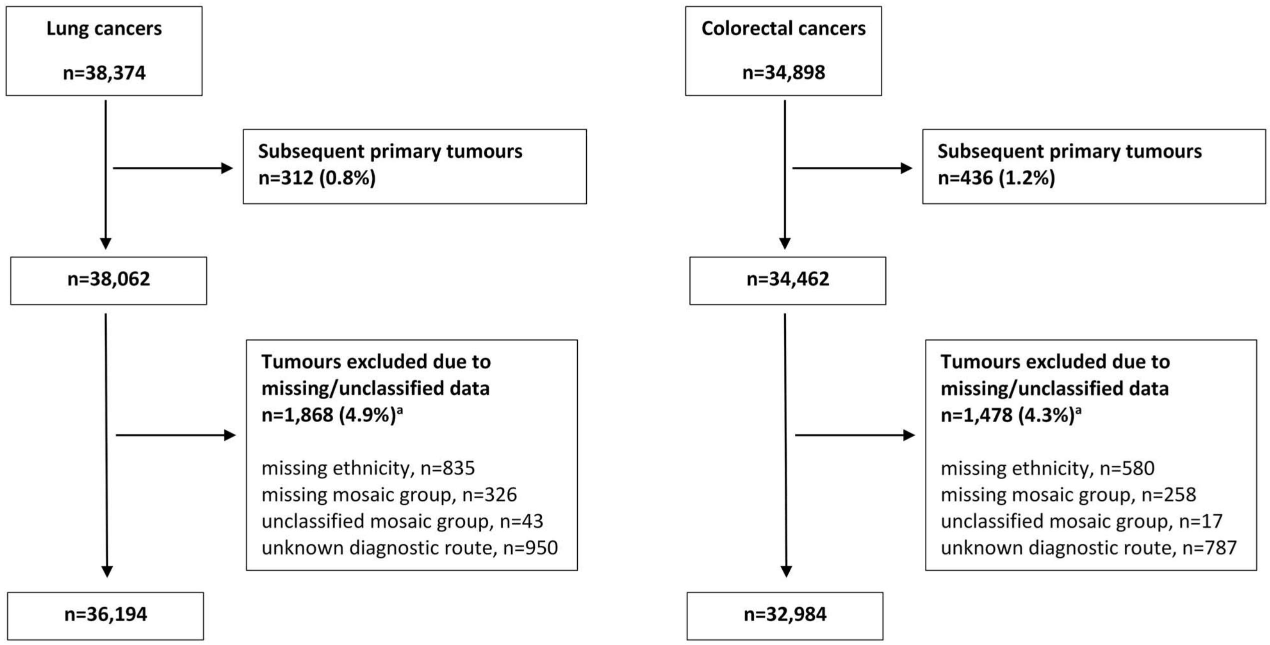
Keywords Cancer, geodemographic segmentation, route of diagnosis, Mosaic

Background

Cancer patients diagnosed through emergency presentation have poorer short-term survival and worse patient experience compared with patients diagnosed through other routes.[1–3]Therefore, reducing emergency presentations is one of the main priorities of early diagnosis initiatives and is monitored as a routine cancer outcome indicator.[4, 5] In England, the proportion of patients diagnosed as emergency presentations has decreased from about one in four patients to less than one in five; whilst these reductions were similar across sociodemographic groups over time,[4–6] inequalities prevail.[7] Specifically, older patients and those living in areas of higher deprivation are at substantially greater risk of being diagnosed with cancer as emergencies.[7, 8] About a third of emergency presenters have not previously seen a general practitioner (GP) with relevant symptoms, and public health awareness interventions may help prevent such emergencies by promoting earlier help-seeking.[9, 10] Such 'social marketing for public health' interventions are customarily targeted to sociodemographic groups at greater risk (e.g. older age and lower income groups). In principle, geodemographic segmentation offers additional potential for public health intervention targeting. Geodemographic segmentation classifies people into groups based on age, deprivation, consumer patterns, wealth (e.g. income, home ownership), communication preferences (e.g. technology,

C.J. Bright, Dr. Senior Analyst
C Gildea, Ms. Senior Analyst
J Lai, Dr. Senior Analyst
L Elliss-Brookes, Ms. Head of Analysis
G Lyratzopoulos, Professor

© The Author(s) 2020. Published by Oxford University Press on behalf of Faculty of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com



[a] tumours could have missing/unclassified information for more than one variable, therefore the breakdown may sum to more than the total number of tumours excluded.

Fig. 1 Flow of data to define final cohort.

media) and household make-up and lifestyle (e.g. smoking, physical activity, diet, hobbies). Key applications include credit rating, marketing, public service/health targeting, service area analysis (geographical area as defined by distance from a certain point) and housing market analysis.[11]

Experian's Mosaic classification system is a type of geodemographic segmentation, which was developed as a marketing tool to provide information on consumer communication preferences and help users gain a detailed understanding of target populations. The Mosaic data use information from several sources—Experian data sources, public and trusted third-party sources including research findings and behavioural data.[12] Although primarily used by commercial companies, Mosaic data have also been used by public health researchers to gain a geodemographic profile of populations of interest.[13–17] Compared to commonly reported measures of deprivation, the Mosaic data generate geodemographic descriptions at a much finer granularity—postcode and household.

Some prior literature has identified geodemographic segmentation group disparities in alcohol consumption,[18] smoking prevalence,[16] uptake of chronic disease screening programmes[17, 19] and hospital admissions.[13, 15] Considering cancer studies, variation in colorectal cancer screening uptake[20] and lung cancer incidence[14] has been reported using geodemographic segmentation. Yet, with a single exception,[20] most evidence thus far does not quantify the unique contribution ('added value') of geodemographic profiling, above and beyond that which can be derived by 'standard' sociodemographic stratification (by age, sex, deprivation and ethnicity). We set out to examine this question empirically.

Methods

We analysed data on incident lung and colorectal cancer diagnosed in England in 2016 registered by the National Cancer Registration and Analysis Service of Public Health England.[21] Information was available on diagnostic route, age, sex and ethnicity. Cancers were defined using the International Classification for Disease version 10; Lung C33-C34 and Colorectal C18-C20. These cancers were chosen as they are common and unlike breast and prostate cancer they occur in both genders. Considering diagnostic route, they provide a clear contrast, with a moderately high (colorectal) or high (lung) proportion presenting as an emergency, and a proportion of colorectal cancer patients being diagnosed through screening whereas this does not currently apply to lung cancer patients in England.

Patients' postcodes of residence were grouped according to the 2011 census Lower Super Output Areas (LSOA) and linked to corresponding income domain scores of IMD 2015 (deprivation quintile). Data on Mosaic consumer classification group[22] was obtained by linking at a household (Unique Delivery Point Reference Number [UDPRN]) or postcode level if UDPRN unavailable (16% cases). After excluding the 'unclassified' group (Fig. 1), 15 geodemographic segmentation groups were considered (Box 1).

Statistical analysis

For each cancer, unadjusted logistic regression models were used to explore the independent association of each variable listed above with each of three outcomes: emergency presentation (yes/no), two-week-wait (TWW) referral (yes/no) and screening detected (yes/no, for colorectal cancer only). Because of likely confounding, we subsequently used multivariable logistic regression models adjusted for age, sex, deprivation and ethnicity. To explore whether there is variation in diagnostic route by geodemographic segmentation group, we added Mosaic group to the multivariable model, examining the degree of change in odds ratios [OR] in all other variables in the presence of adjustment for Mosaic group. To test the significance of each variable in the model and obtain a P value, Wald tests were used. Crude proportions were calculated as number of cancers by diagnostic route over the total number of cancers, multiplied by 100. Adjusted proportions were predicted from the multivariable logistic regression models previously specified. Spearman's rank and Kendal–Tau correlations were used to measure the strength and direction of association between crude and adjusted odds for Mosaic group; higher correlation values indicate a similar ordering of Mosaic groups by either their crude or adjusted OR values, whilst low correlations indicate the opposite. We considered stage at diagnosis in supplementary analysis. We focused on emergency presentations, with other routes (TWW referral and screening) provided in supplementary material for comparison.

All analysis was conducted in Stata version 15.0.

Box 1

Experian's Mosaic public section consumer classification system uses a range of information, beyond that available to quantify deprivation, to classify the population into 15 different groups.[12] The Mosaic groups and their key characteristics are listed below:

Country Living: well-off homeowners in rural locations enjoying the benefits of rural life.

Prestige Positions: established families in large detached homes living upmarket lifestyles.

City Prosperity: high-status city dwellers living in central locations and pursuing careers with high rewards.

Domestic Success: thriving families who are busy bringing up children and following careers.

Suburban Stability: mature suburban owners living settled lives in mid-range housing.

Senior Security: elderly people with assets who are enjoying a comfortable retirement.

Rural Reality: householders living in inexpensive homes in village communities.

Aspiring Homemakers: younger households settling down in housing priced within their means.

Urban Cohesion: residents of settled urban communities with a strong sense of identity.

Rental Hubs: educated young people privately renting in urban neighbourhoods.

Modest Traditions: mature homeowners of value homes enjoying stable lifestyles.

Transient Renters: single people who pay modest rents for low cost homes.

Family Basics: families with limited resources who have to budget to make ends meet.

Vintage Value: elderly people reliant on support to meet financial or practical needs.

Municipal Challenge: urban renters of social housing facing an array of challenges.

Results

After excluding patients with missing data (4.6%), a total of 36 194 and 32 984 patients with lung and colorectal cancer, respectively, were included in the analysis (Fig. 1). Among lung/colorectal cancer patients respectively, 54%/56% were male, 93%/91% white, 48%/34% in the two most deprived quintiles, 69%/51% diagnosed at stage III or IV and with median age at cancer diagnosis of 73/72 years. Table 1 provides the demographic characteristics of patients within each Mosaic group. Among lung/colorectal cancer patients, 33%/23% and 29%/35% were diagnosed through an emergency or TWW referral, respectively. Among colorectal cancer patients eligible for screening (i.e. aged 60–74), 25% presented via a screening route.

There was variation in the unadjusted odds of emergency presentation by deprivation group for both lung and colorectal cancers (both $P < 0.001$), with 1.30-fold and 1.60-fold increased odds of emergency presentation comparing the most and least deprived groups, respectively (Tables 2 and 3). Adjustment for all variables (excluding Mosaic group), increased the size of variation slightly, with odds ratio for emergency presentation of to 1.42 and 1.67 for the most compared with the least deprived group in patients with lung and colorectal cancer, respectively. Adding Mosaic group to the model attenuated the size of variation by deprivation group substantially, to odds ratios of 1.18 and 1.28 for the most compared with the least deprived group in patients with

Table 1 Characteristics of lung and colorectal cancer patients by mosaic group

	Lung cancer						*Colorectal cancer*					
Mosaic group	*N*	*%Total*	*%Male*	*Median age*[a]	*%White*	*%Deprivation Q4/5*	*N*	*%Total*	*%Male*	*Median age*[a]	*%White*	*%Deprivation Q4/5*
Total	36194	100	53.5	73	93.2	48.0	32984	100	56.4	72	91.0	34.0
Country Living	2078	5.7	56.4	74	95.4	4.8	3126	9.5	58.5	73	94.3	3.4
Prestige Positions	1946	5.4	57.0	72	92.5	4.3	3392	10.3	58.2	70	91.8	3.9
City Prosperity	885	2.4	56.0	71	78.8	35.3	931	2.8	58.3	69	75.6	28.6
Domestic Success	944	2.6	50.6	71	89.4	9.7	1290	3.9	54.4	63	88.0	10.0
Suburban Stability	2375	6.6	55.3	67	94.8	19.1	2765	8.4	58.2	64	93.4	17.3
Senior Security	6527	18.0	53.7	77	96.4	19.9	6989	21.2	54.2	78	95.6	17.0
Rural Reality	1912	5.3	56.3	72	95.6	18.9	1790	5.4	55.8	72	94.2	16.4
Aspiring Homemakers	1044	2.9	52.1	71.5	92.5	35.9	975	3.0	52.2	64	92.4	30.2
Urban Cohesion	1467	4.1	57.7	72	76.1	63.4	1500	4.5	56.4	71	68.7	59.4
Rental Hubs	818	2.3	55.6	67.5	85.1	55.5	795	2.4	57.4	63	79.7	46.0
Modest Traditions	3219	8.9	54.1	69	96.0	73.4	2130	6.5	62.0	67	93.8	69.4
Transient Renters	1174	3.2	56.2	66.5	92.5	77.8	725	2.2	58.6	65	87.0	73.4
Family Basics	1605	4.4	51.1	65	93.0	85.2	961	2.9	53.4	58	88.2	84.3
Vintage Value	7141	19.7	49.0	76	96.0	75.2	4161	12.6	53.3	78	94.7	69.2
Municipal Challenge	3059	8.5	53.6	67	90.0	95.3	1454	4.4	59.0	66	81.6	95.3
P value[b]			<0.001	<0.001	<0.001	<0.001			<0.001	<0.001	<0.001	<0.001

[a]Median age at cancer diagnosis.

[b]*P* value calculated using Pearson's chi-squared test, except for 'median age' where *P* value is calculated using Kruskal–Wallis equality-of-populations rank test.

lung and colorectal cancer, respectively. There was a 3%-point and 4%-point difference between the most and least deprived in terms of adjusted proportions in emergency presentations for lung and colorectal cancer patients (lung: 30% versus 34%; colorectal: 21% versus 25%).

The variation in odds of emergency presentation of lung cancer across age groups increased from 4.13 (OR between highest/lowest group) to 4.30 after adjustment for all variables (excluding Mosaic group) and 4.86 (including adjustment for Mosaic group). After adjustment (including for Mosaic group), 25% of people diagnosed aged 18–39 years presented as an emergency compared to 61% aged 90+ Supplementary Table S2. Similarly, for colorectal cancer, we observed large variation in odds of emergency presentation across age groups (5.04 OR between the age group with the highest/lowest odds ratios), but this remained similar after adjustment for all variables (both with/without Mosaic group). After adjustment (including for Mosaic group), 17% of people diagnosed aged 60–69 years presented as an emergency compared to 50% aged 90+ Supplementary Table S3.

Considering geodemographic segmentation group, there was variation in the unadjusted odds of emergency presentation across the 15 Mosaic groups for both lung and colorectal cancer (Tables 2 and 3). For lung, there was 1.59-fold variation in odds of emergency presentation between the highest and the lowest risk Mosaic group for emergency presentation, increasing to 1.66-fold after adjustment for all variables. There was a 10%-point difference between the highest and lowest risk Mosaic group in terms of adjusted proportions in emergency presentations (39% of Rental Hubs versus 29% of Senior Security)—this represents a large variation given the absolute overall (all-patient) percentage was 33%. There was a high degree of concordance in the ranks of the 15 Mosaic groups when considering crude and adjusted odds ratio values, meaning that groups with higher crude odds tended to also have a higher adjusted odds, and vice versa (Spearman's $\rho = 0.92$, $P < 0.001$ Kendall's Tau $\tau a = 0.75$, $P < 0.001$). For colorectal cancer, variation between highest/lowest risk group for emergency presentation was reduced (OR = 2.30 to OR = 1.89) after adjustment for all variables (including

Table 2 Crude and adjusted odds of emergency presentation for lung cancer patients[a]

	Lung cancer n = 36 194								
	Crude			*Adjusted for sex, age, ethnicity, deprivation quintile—not including mosaic group*			*Adjusted for sex, age, ethnicity, deprivation quintile and mosaic group*		
	%	OR (95% CI)	*P* value	%	OR (95% CI)	*P* value	%	OR (95% CI)	*P* value
Deprivation quintile									
1—Least deprived	29.4%	Ref		28.5%	Ref		30.4%	Ref	
2	30.9%	1.08 (0.99–1.17)	0.072	30.4%	1.10 (1.01–1.19)	0.021	31.8%	1.07 (0.99–1.16)	0.106
3	32.6%	1.16 (1.08–1.26)	<0.001	32.5%	1.22 (1.13–1.32)	<0.001	33.0%	1.13 (1.04–1.23)	0.004
4	33.6%	1.21 (1.13–1.31)	<0.001	33.9%	1.30 (1.20–1.40)	<0.001	33.3%	1.15 (1.05–1.26)	0.002
5—Most deprived	35.0%	1.30 (1.20–1.39)	<0.001	35.8%	1.42 (1.31–1.53)	<0.001	33.8%	1.18 (1.07–1.30)	0.001
Overall *P* value[b]			<0.001			<0.001			0.010
Mosaic group									
Country Living	28.6%	0.71 (0.63–0.78)	<0.001				29.2%	0.82 (0.73–0.92)	0.001
Prestige Positions	28.1%	0.69 (0.62–0.77)	<0.001				29.5%	0.83 (0.73–0.94)	0.003
City Prosperity	32.0%	0.83 (0.71–0.96)	0.014				32.6%	0.96 (0.83–1.13)	0.654
Domestic Success	37.5%	1.06 (0.92–1.22)	0.435				38.1%	1.23 (1.06–1.43)	0.007
Suburban Stability	28.4%	0.70 (0.63–0.77)	<0.001				31.4%	0.91 (0.82–1.02)	0.104
Senior Security	30.3%	0.77 (0.71–0.82)	<0.001				28.5%	0.79 (0.73–0.86)	<0.001
Rural Reality	31.2%	0.80 (0.72–0.89)	<0.001				32.4%	0.96 (0.85–1.07)	0.447
Aspiring Homemakers	38.4%	1.10 (0.96–1.26)	0.166				38.3%	1.24 (1.08–1.43)	0.002
Urban Cohesion	32.2%	0.84 (0.74–0.95)	0.004				32.4%	0.95 (0.84–1.08)	0.456
Rental Hubs	38.4%	1.10 (0.95–1.27)	0.218				39.4%	1.31 (1.12–1.53)	0.001
Modest Traditions	29.9%	0.75 (0.69–0.82)	<0.001				32.2%	0.94 (0.86–1.03)	0.217
Transient Renters	37.2%	1.05 (0.92–1.19)	0.499				38.7%	1.27 (1.11–1.44)	<0.001
Family Basics	34.3%	0.92 (0.82–1.03)	0.158				36.0%	1.13 (1.00–1.27)	0.048
Vintage Value	36.2%	Ref					33.4%	Ref	
Municipal Challenge	35.3%	0.96 (0.88–1.05)	0.406				37.1%	1.18 (1.07–1.30)	0.001
Overall *P* value[b]			<0.001						<0.001
OR between highest/lowest group		1.59						1.66	
Spearman's rank test					$\rho = 0.918$, $P < 0.001$				
Kendall–Tau test					τa =0.752, $P < 0.001$				

[a]For sex, age, ethnicity, deprivation quintile and mosaic group we used male gender, age 50–59 years, white, least deprived and vintage value (largest group), respectively, as the reference categories in the models.
[b]*P* value calculated using Wald test.

for deprivation group). There was a 11%-point difference between the highest and lowest risk Mosaic group in terms of adjusted proportions in emergency presentations (31% of Transient Renters versus 20% of Prestige Positions)—this represents a large variation given the absolute overall (all-patient) percentage was 23%. There was high concordance in the ranks of crude/adjusted odds across the 15 Mosaic groups (Spearman's $\rho = 0.95$, $P < 0.001$, Kendall's Tau τa = 0.82, $P < 0.001$).

To summarize, people belonging to Prestige Positions group ('people living in detached houses living upmarket lifestyles') and Senior Security group ('elderly people living a comfortable retirement') were least likely to present as an emergency. In contrast, people belonging to Rental Hubs group ('young educated private renters') and Transient Renters group ('single people with modest rents') were most likely to present as an emergency.

Sensitivity analysis: the findings presented remained practically unchanged after adjustment for stage at diagnosis (results not shown).

Supplementary analysis: additional analyses for other diagnostic routes (TWW, and screening for colorectal cancer)

Table 3 Crude and adjusted odds of emergency presentation for colorectal cancer patients[a]

	Colorectal cancer n = 32 984								
	Crude			*Adjusted for sex, age, ethnicity, deprivation quintile—not including mosaic group*			*Adjusted for sex, age, ethnicity, deprivation quintile and mosaic group*		
	%	OR (95% CI)	*P* value	%	OR (95% CI)	*P* value	%	OR (95% CI)	*P* value
Deprivation quintile									
1—Least deprived	19.7%	Ref		19.6%	Ref		21.2%	Ref	
2	20.7%	1.07 (0.98–1.15)	0.124	20.8%	1.08 (1.00–1.18)	0.051	21.9%	1.04 (0.96–1.14)	0.313
3	22.6%	1.19 (1.10–1.29)	<0.001	22.4%	1.19 (1.10–1.30)	<0.001	22.5%	1.09 (0.99–1.19)	0.068
4	25.1%	1.37 (1.26–1.48)	<0.001	25.0%	1.39 (1.27–1.51)	<0.001	23.8%	1.17 (1.06–1.29)	0.001
5—Most deprived	28.2%	1.60 (1.47–1.74)	<0.001	28.5%	1.67 (1.54–1.83)	<0.001	25.3%	1.28 (1.14–1.43)	<0.001
Overall *P* value[b]			<0.001			<0.001			<0.001
Mosaic group									
Country Living	18.7%	0.59 (0.53–0.66)	<0.001				20.2%	0.77 (0.68–0.87)	<0.001
Prestige Positions	17.3%	0.54 (0.48–0.60)	<0.001				19.6%	0.74 (0.65–0.84)	<0.001
City Prosperity	22.6%	0.75 (0.63–0.88)	<0.001				23.0%	0.92 (0.77–1.10)	0.343
Domestic Success	24.7%	0.84 (0.73–0.97)	0.019				25.0%	1.03 (0.88–1.20)	0.737
Suburban Stability	18.8%	0.59 (0.53–0.67)	<0.001				22.1%	0.87 (0.76–0.99)	0.032
Senior Security	21.1%	0.69 (0.63–0.75)	<0.001				20.1%	0.76 (0.69–0.85)	<0.001
Rural Reality	21.6%	0.71 (0.62–0.81)	<0.001				22.6%	0.89 (0.78–1.03)	0.120
Aspiring Homemakers	27.1%	0.95 (0.81–1.11)	0.543				25.2%	1.04 (0.88–1.23)	0.665
Urban Cohesion	23.5%	0.79 (0.69–0.90)	0.001				22.6%	0.89 (0.77–1.04)	0.135
Rental Hubs	30.2%	1.11 (0.94–1.31)	0.220				29.1%	1.28 (1.07–1.52)	0.007
Modest Traditions	21.4%	0.70 (0.62–0.79)	<0.001				23.4%	0.93 (0.82–1.06)	0.307
Transient Renters	32.6%	1.24 (1.05–1.47)	0.013				31.0%	1.40 (1.18–1.67)	<0.001
Family Basics	32.6%	1.24 (1.07–1.44)	0.005				29.8%	1.32 (1.13–1.56)	0.001
Vintage Value	28.0%	Ref					24.5%	Ref	
Municipal Challenge	29.0%	1.05 (0.92–1.19)	0.508				28.7%	1.25 (1.09–1.44)	0.002
Overall *P* value[b]			<0.001						<0.001
OR between highest/lowest group		2.30						1.89	
Spearman's Rank test					ρ = 0.949, $P < 0.001$				
Kendall–Tau test					τa =0.819, $P < 0.001$				

[a]For sex, age, ethnicity, deprivation quintile and mosaic group we used male gender, age 50–59 years, white, least deprived and vintage value (largest group), respectively, as the reference categories in the models.
[b]*P* value calculated using Wald test.

are included in Appendix/Online Tables. There was little variation between unadjusted and adjusted odds of all diagnostic routes by ethnicity and sex Supplementary Tables S2–S6. There was 2-fold variation in the odds of screening presentation for colorectal cancer between the highest and lowest risk Mosaic groups after adjustment for age, sex, ethnicity and deprivation Supplementary Table S6. There was a 15%-point difference between the highest and lowest risk Mosaic group in terms of adjusted proportions in screening presentation for colorectal cancer (30% Prestige Positions versus 15% of Family Basics).

Discussion

Main finding of this study

We found that large differences in the risk of emergency presentation in cancer patients can be explained by geodemography, additional to age and conventional measures of deprivation. This is also true of screening presentation for colorectal cancer. Thus, the use of geodemography could help to target public health interventions for earlier detection (screening programmes) and earlier diagnosis (awareness campaigns, which aim to reduce emergency presentations).

What is already known on this topic

Previous studies have shown that there are inequalities in diagnostic route for cancer, in particular persons of older age at diagnosis and higher deprivation are at greater risk of presenting via an emergency.[1, 7–9] Our study is consistent with these findings, though it adds one more dimension to appreciation of variation in risk of diagnosis of cancer through an emergency presentation, regarding the influence of geosegmentation group. Not all emergency presentations can be considered preventable as some chiefly reflect tumour biology in patients with minimal or no prodromal symptoms; in such cases, the patient and health system factors have little influence on the risk of emergency presentation.[2, 23] However, the fact that emergency presentations vary greatly by sociodemographic factors (including very large differences in proportions between most and least deprived patients), and their rapid reduction in recent years cannot be principally explained by sociodemographic differences in tumour biology and strongly indicate a sizeable component that reflects patient behaviour and/or health system factors.[2, 7, 23] It is therefore important that efforts continue to decrease both the absolute number of emergency presentations and related inequalities, until the minimum proportion dictated by tumour biology is reached.

Only a small body of evidence exists regarding the use of geodemographic segmentation to help understand inequalities in health outcomes.[13–20] In the context of health studies on cancer, previous studies have examined variation by geodemographic group in lung cancer incidence[14] and colorectal cancer screening uptake.[20] The latter study adjusted for both deprivation (IMD) and geodemographic segmentation and found evidence for independent variation in uptake of screening by both IMD and ‘P^2 segmentation’ group. In contrast, a different study reported no variation in cancer characteristics or outcomes (stage, performance status, therapy received, survival) by geodemographic segmentation.[24]

What this study adds

Identifying subsets of the population most at risk of emergency presentation can help to guide interventions to reduce inequalities in relation to the route to cancer diagnosis. In turn, doing so can help to achieve reductions in persistent inequalities in cancer survival.[25] Our study adds substantially to the present evidence base, regarding the added value of geodemographic segmentation above and beyond what is already known using ‘standard’ sociodemographic profiling of diagnostic route.

Geodemographic segmentation can offer additional insights beyond the consideration of deprivation measures. This may arise from two principle means. Firstly, the former is based at household (or postcode) level, whereas the latter is based on LSOA (400–1200 households)[26]; this reduces misclassification error due to ecological fallacy.[27] Secondly, geodemographic segmentation classification includes information on a much wider range of variables beyond those which are also included in conventional deprivation groups i.e. area statistics on markers of socioeconomic deprivation.

Our study is consistent with previous studies which have shown a relationship between diagnostic route with deprivation, age and sex.[7, 8] We found age at diagnosis was the variable associated with the greatest size of variation in diagnostic route, with older age most at risk of emergency presentation, and least likely to present via TWW referral (lung) or for colorectal cancer only, via screening. There was also large variation in diagnostic route by deprivation group, with the most deprived groups being at greater risk of emergency presentation and least likely to present via TWW referral or for colorectal cancer via screening. Unique to our study, we have also shown that geodemographic segmentation adds information to identify which subsets of the population are most at risk of emergency presentation. We observed large variation in the risk of emergency presentation by Mosaic group. These findings provide proof of concept for public health interventions targeting all of the examined attributes, including geodemographic segmentation group.

Nonetheless, whilst the information gleaned from this study could be used to further stratify the population and target interventions to reduce inequalities in routes to diagnosis, such interventions will need to be rigorously evaluated for their efficacy and cost-effectiveness. Further work would be required to investigate which specific factors, of those used for the Mosaic classification, contribute additional information in explaining variation in diagnostic route.

Limitations of this study

This population-based study design utilizing electronically linked data, enabled analysis of data on diagnostic route, person-level sociodemographic factors and geodemographic segmentation among a large and representative cancer patient population in England. In the adjusted analysis we were able to account for the sociodemographic composition of Mosaic groups by adjusting for age and deprivation among other variables.

There are several limitations. Firstly, we used deprivation and geodemographic segmentation measures that are based on English data, therefore the results of this study may not be applicable to other countries with different geo- and sociodemographics context.

Secondly, we restrict ourselves to one important aspect of cancer epidemiology (diagnostic route) given its prognostic significance for survival. However, future studies may wish to investigate the relationship between geodemographic segmentation and other important cancer measures such as treatment and survival. Similarly, although we have concentrated on lung and colorectal cancers future research may wish to explore other cancers. We must also consider the possibility (albeit small given the range of factors already considered) that an unobserved factor could better explain the variation in diagnostic route of cancer.

Whilst the ability to link Mosaic groups at a household level is a strength of the study, for 16% of patients, UDPRN was not available and therefore these patients were linked at a postcode level (<0.1% unable to link overall). This is likely to have attenuated the size of true underlying associations, i.e. the findings can be conservative with regard to the influence of geodemographic segmentation group.

The size of some Mosaic groups with the highest/lowest odds of emergency presentation is small. However, this is often the case for stratified approaches targeting only small segments of the general population.

Given the size of differences in the estimated OR for age group and deprivation in models either including or excluding the geosegmentation group variable, and the p values obtained from likelihood ratio tests comparing such model pairs ($P < 0.001$ for all), we are confident that the amount of information added by including the geosegmentation variable is substantial.

Conclusions

Differences in risk of emergency presentation in cancer patients are associated with geodemographic group, above and beyond variation by deprivation and age. The findings provide support that information provided by geodemographic segmentation approach, in addition to the more frequently used conventional sociodemographic variables could help to target public health interventions to reduce inequalities in routes to diagnosis for cancer.

Supplementary data

Supplementary data are available at the *Journal of Public Health* online.

Conflict of interest

None declared.

Acknowledgements

This work uses data that have been provided by patients and collected by the NHS as part of their care and support. The data is collated, maintained and quality assured by the National Cancer Registration and Analysis Service, which is part of Public Health England (PHE). The authors acknowledge Experian for the provision of the Mosaic data that has been used in this work. G.L. is supported by a Cancer Research UK Advanced Clinician Scientist Fellowship (grant number C18081/A18180); additionally, he is an associate director of the multi-institutional CanTest Collaborative, which is funded by Cancer Research UK (grant number C8640/A23385).

References

1 McPhail S, Elliss-Brookes L, Shelton J *et al.* Emergency presentation of cancer and short-term mortality. *Br J Cancer* 2013;**109**: 2027–34.

2 Zhou Y, Abel GA, Hamilton W *et al.* Diagnosis of cancer as an emergency: a critical review of current evidence. *Nat Rev Clin Oncol* 2017;**14**:45–56.

3 Salika T, Abel GA, Mendonca SC *et al.* Associations between diagnostic pathways and care experience in colorectal cancer: evidence from patient-reported data. *Frontline Gastroenterol* 2018;**9**: 241–8.

4 National Cancer Registration and Analysis Service. Routes to diagnosis 2006–2016 year breakdown. Public Health England; 2018. available from: http://www.ncin.org.uk/publications/routes_to_diagnosis

5 National Cancer Registration and Analysis Service. Cancer Outcomes: Emergency Presentations. Public Health England; 2020. available from: http://www.ncin.org.uk/cancer_type_and_topic_specific_work/topic_specific_work/cancer_outcome_metrics

6 Herbert A, Abel GA, Winters S *et al.* Cancer diagnoses after emergency GP referral or A&E attendance in England: determinants and time trends in routes to diagnosis data, 2006-2015. *Br J Gen Pract* 2019;**69**:e724–e30.

7 Herbert A, Abel GA, Winters S *et al.* Are inequalities in cancer diagnosis through emergency presentation narrowing, widening or remaining unchanged? Longitudinal analysis of English population-based data 2006–2013. *J Epidemiol Community Health* 2019;**73**:3.

8 Elliss-Brookes L, McPhail S, Ives A *et al.* Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *Br J Cancer* 2012;**107**:1220–6.

9 Abel GA, Mendonca SC, McPhail S *et al.* Emergency diagnosis of cancer and previous general practice consultations: insights from linked patient survey data. *Br J Gen Pract* 2017;**67**:e377–e87.

10 Murchie P, Smith SM, Yule MS *et al.* Does emergency presentation of cancer represent poor performance in primary care? Insights from a novel analysis of linked primary and secondary care data. *Br J Cancer* 2017;**116**:1148–58.

11 Webber RJ, Butler T, Phillips T. Adoption of geodemographic and ethno-cultural taxonomies for analysing big data. *Big Data Soc* 2015;**2**:205395171558391.

12 Experian Information Solutions. Mosaic - consumer classification for consistent cross-channel marketing. Available from: https://www.experian.co.uk/business/marketing/segmentation-targeting/mosaic/.

13 Doos L, Uttley J, Onyia I *et al.* Mosaic segmentation, COPD and CHF multimorbidity and hospital admission costs: a clinical linkage study. *J Public Health* 2014;**36**:317–24.

14 Iyen-Omofoman B, Hubbard RB, Smith CJ *et al.* The distribution of lung cancer across sectors of society in the United Kingdom: a study using national primary care data. *BMC Public Health* 2011;**11**:857.

15 Kadam UT, Lawson CA, Moody DK *et al.* Consumer segmentation and time interval between types of hospital admission: a clinical linkage database study. *J Public Health* 2018;**40**:154–62.

16 Sharma A, Lewis S, Szatkowski L. Insights into social disparities in smoking prevalence using mosaic, a novel measure of socioeconomic status: an analysis using a large primary care dataset. *BMC Public Health* 2010;**10**:755.

17 Waqar S, Bullen G, Chant S *et al.* Cost implications, deprivation and geodemographic segmentation analysis of non-attenders (DNA) in an established diabetic retinopathy screening programme. *Diabetes Metab Syndr* 2012;**6**:199–202.

18 Powell J, Tapp A, Sparks E. Social marketing in action—geodemographics, alcoholic liver disease and heavy episodic drinking in great Britain. *Int J Nonprofit Voluntary Sect Mark* 2007;**12**:177–87.

19 Sheringham J, Sowden S, Stafford M *et al.* Monitoring inequalities in the National Chlamydia Screening Programme in England: added value of ACORN, a commercial geodemographic classification tool. *Sex Health* 2009;**6**:57–62.

20 Nnoaham KE, Frater A, Roderick P *et al.* Do geodemographic typologies explain variations in uptake in colorectal cancer screening? An assessment using routine screening data in the south of England. *J Public Health* 2010;**32**:572–81.

21 Henson KE, Elliss-Brookes L, Coupland VH *et al.* Data resource profile: National Cancer Registration Dataset in England. *Int J Epidemiol* 2019;**49**:16.

22 Experian. Mosaic consumer classification for consistent cross-channel marketing. 2019. Available from: https://www.experian.co.uk/business/marketing/segmentation-targeting/mosaic/.

23 Lyratzopoulos G, Saunders CL, Abel GA. Are emergency diagnoses of cancer avoidable? A proposed taxonomy to motivate study design and support service improvement. *Future Oncol* 2014;**10**:1329–33.

24 Cheyne L, Taylor A, Milton R *et al.* Social deprivation does not affect lung cancer stage at presentation or disease outcome. *Lung Cancer* 2013;**81**:247–51.

25 Exarchakou A, Rachet B, Belot A *et al.* Impact of national cancer policies on cancer survival trends and socioeconomic inequalities in England, 1996-2013: population based study. *BMJ* 2018;**360**:k764.

26 Office for National Statistics. Census. *Geography*. 2019. Available from:. https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography.

27 Firebaugh G. Ecological Fallacy, Statistics of. In: Wright JD (ed). *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edn. Oxford: Elsevier, 2015, 865–7.