




# Can we empirically derive a geographic definition of 'coastal' for use in cancer data reporting? An ecological modelling study using England's national cancer registry

Claire Welsh <sup>1</sup>, Catherine Welham,<sup>1</sup> Jack Anderson,<sup>1</sup> Mark Alan Green <sup>2</sup>, Clarice Quinn,<sup>1</sup> Jennifer Lai,<sup>1</sup> Sally Vernon,<sup>1</sup> Lizz Paley <sup>1</sup>

**To cite:** Welsh C, Welham C, Anderson J, *et al*. Can we empirically derive a geographic definition of 'coastal' for use in cancer data reporting? An ecological modelling study using England's national cancer registry. *BMJ Public Health* 2024;**2**:e001067. doi:10.1136/bmjph-2024-001067

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjph-2024-001067>).

SV and LP are joint senior authors.

Received 18 February 2024  
Accepted 29 October 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. Published by BMJ.

<sup>1</sup>NDRS, NHS England, Leeds, UK

<sup>2</sup>Geography & Planning, University of Liverpool, Liverpool, UK

**Correspondence to**  
Dr Claire Welsh;  
[claire.welsh8@nhs.net](mailto:claire.welsh8@nhs.net)

## ABSTRACT

**Background** Reducing avoidable systematic differences in population health requires first understanding which populations are currently disadvantaged. Although the health of coastal communities in England has been of concern for some years, an operationalised definition of 'coastal' is lacking. This study aims to use national cancer statistics to define and validate a small area-level definition of 'coastal' that could be used to better report cancer-related health inequalities in England.

**Methods** Information on the geography and demography of English populations at the Lower Super Output Area (LSOA) level were used to define a suite of candidate coastal variables that considered foreshore proximity, resident population location, rurality and deprivation. Adjusted linear models of LSOA-level statistics of cancer incidence, prevalence and mortality in England (2016 to 2020) were used to identify candidate coastal variable(s) that explained the greatest proportion of variation in cancer outcomes after adjustment.

**Results** The candidate 'G\_25\_5' (LSOA's designated as 'coastal' if 25% or more of postcodes were within 5 km of the coastline) was selected as the candidate that explained the most residual variation in cancer incidence and prevalence after adjustment. This variable would assign 7377 2011 LSOAs as coastal, whose populations summed to 12.3 million people (22% of England's population, in 2016). This candidate variable was not significantly associated with cancer mortality.

**Conclusions** The coastal variable that we identify can explain some of the 'coastal excess' in poor cancer outcomes. We propose that this variable is now embedded into health inequalities reporting and adopted as the working definition of 'coastal' implicated in NHS England's 'Core20PLUS5' approach for use in cancer data reporting.

## INTRODUCTION

Inequalities in health are unfair, avoidable, systematic differences in population-level health that can lead to large disparities in outcomes like life expectancy.<sup>1</sup> Much work

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Despite the health of coastal communities in the UK having come under scrutiny and been found lacking in recent years, research into precisely who is disadvantaged, in what ways and through which causal mechanisms has been lacking. The main cause of this deficit in our knowledge, and thus the ability to affect positive change, has hitherto been hampered by the lack of an accepted definition of 'coastal' for the purposes of cancer research.

## WHAT THIS STUDY ADDS

⇒ We present the first empirically derived, validated definition of what constitutes a 'coastal community' at the small area level, in a manner suitable for immediate adoption across healthcare and research settings where cancer data are reported.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Provision and adoption of this definition within the cancer domain will allow comparative healthcare research to begin, which will accelerate our understanding of coastal health inequality for the betterment of the outcomes in those affected communities.

in recent times has sought to identify, quantify and explain these differences found across many dimensions of geography and patient demography, with the aim of identifying levers to reduce their magnitude. NHS England identified several populations at risk of such inequality using their 'Core20PLUS5'<sup>2</sup> approach, to help target the most at-risk groups. One such group was 'coastal communities with pockets of deprivation hidden among relative affluence'. This population was identified as one facing unacceptable health inequalities (compared with 'non-coastal' areas) by reports such as the

Chief Medical Officer's Annual Report 2021—Health in Coastal Communities.<sup>3</sup> In this report, it was noted that there was no nationally agreed definition nor consensus on what constitutes a coastal community, and definitions that include areas as large as local authorities, or even larger areas, lead to the deprived communities of interest being masked by more affluent inland areas.

Having a coastal definition that assigns small areas as populations at risk of health inequalities would allow the NHS, Office for National Statistics (ONS) and other health bodies to publish data comparing those areas to others inland, allowing deeper understanding of the distributions of health outcomes, quantification of the 'coastal excess' in adverse outcomes that is not accounted for by deprivation and population age, and fulfilment of the NHS's Public Sector Equality Duty.<sup>4</sup>

This study aims to develop and evaluate a range of empirically derived coastal definitions at the small area level for cancer data, to identify a candidate variable that captures the maximum possible variance in cancer-related health between coastal and non-coastal areas that is not explained by known demographic and geographic characteristics. The candidate variable chosen will be embedded within the National Disease Registration Service's (NDRS) databases to allow future cancer data reporting to include breakdowns by coastal/non-coastal groupings. Our goal is to develop a resource that is easy to interpret and could be deployed by a range of organisations that may not have the analytical capacity to use more complicated methods for defining coastal areas.

As long ago as 2006, it was noted that coastal populations of the UK were experiencing dramatic demographic ageing.<sup>5</sup> Alongside an ageing population, coastal communities tend to suffer from a range of socioeconomic disadvantages that have been linked to worse psychological and physical health,<sup>6</sup> including reliance on the seasonally variable, service-led employment sector in the wake of reduced maritime industry, a comparatively low-skilled workforce with worse job security and a limited range of employment opportunities, and a higher density of houses of multiple occupation. The stress afforded by these conditions can increase the risk of non-communicable diseases like cancer.<sup>7</sup> These issues are not confined to UK shores, as international work on coastal inequalities has shown.<sup>8 9</sup> This work represents a cancer-focused first step towards reporting of coastal inequality, in the hopes that future collaboration can lead to effective reduction in health inequality across a wider set of health domains.

## MATERIALS AND METHODS

### Geographical data

Shapefiles of the English coastline, ultra-generalised to 500 m, were obtained from ONS, along with shapefiles of the boundaries of all English Lower Super Output Areas (LSOAs, small census geographies containing on average 1500 residents, or 650 households,  $n=32\ 844$ ) as of

December 2011 (the boundaries version released closest to the start of the period of interest). Geographical coordinates for all the centroids of English postcodes were obtained from the ONS. We also accessed two contextual LSOA-level variables to further stratify our definitions of 'coastal': the 2019 Index of Multiple Deprivation (IMD) quintile was selected to measure neighbourhood socioeconomic deprivation.<sup>10</sup> This is a widely used measure and is a key reporting metric given the wide inequalities for cancer outcomes for this indicator.<sup>11 12</sup> 2011 Rural Urban Classification (the dataset closest temporally to the start of the period of interest) was used to compare differences between urban and rural areas.<sup>13</sup>

### Outcomes: cancer data

Cancer data were obtained from the National Cancer Registration Dataset, which is collated, maintained and quality assured by the NDRS, part of NHS England (NHSE).<sup>14 15</sup> The cohort definition used in this study aligns with the cancer definitions used by NDRS for publication of national cancer statistics ([https://www.cancerdata.nhs.uk/incidence\\_and\\_mortality](https://www.cancerdata.nhs.uk/incidence_and_mortality), see online supplemental file 1). For each LSOA, counts of incident and prevalent cancer cases, and deaths from cancer (all malignant ICD-10 'C' codes excluding C44) for the period 2016–2020 were calculated. Patients were assigned to LSOAs based on their postcode at the time of diagnosis for incident and prevalent cases, and postcode at time of death for mortality counts. Directly age-standardised rates of cancer incidence and mortality per 100 000 people were calculated using the method detailed in the online supplemental file 1. Crude prevalence rates per 100 000 persons were calculated for all LSOAs. Cancer prevalence was not age-standardised due to definitional difficulties in determining the 'age at the point of prevalence'.

Cancer cases (incident, prevalent and mortality cases separately) over the period 2016–2020 were each summed per Lower Tier Local Authority (a higher-level administrative geography into which LSOAs are nested, and which is a common level at which cancer data are reported). Should this aggregation lead to many (empirically, >5) Local Authorities with too few cancer cases in their 'coastal' communities to enable safe data reporting (ie, for the avoidance of potential information disclosure), the candidate coastal variable would be dropped. Data for the Isles of Scilly and City of London Local Authorities were excluded due to known epidemiological inconsistencies and small populations ( $n=7$  LSOAs).

### Candidate variable derivation

#### Geography-only candidates ('G candidates')

Three postcode-level binary variables were computed by determining whether each postcode was within 1 km, 3 km or 5 km of the English coastline. A qualitative examination of coastal towns (eg, Crosby, Blackpool, Skegness, Hoylake, Bridlington) suggested that a distance of 5 km would include most of these towns (including

larger settlements like Blackpool). We used a varying distance threshold since no agreed definition exists and to account for areas that may be viewed as ‘coastal’ despite not sharing a contiguous border with the coast (eg, seaside towns that extend inland).

To determine whether each LSOA was ‘coastal’, four separate definitions were developed. First, an LSOA was considered ‘coastal’ if it intersected with any part of the English coastline (candidate variable ‘G1’). Next, we attempted to identify LSOAs that had at least one coastal ‘neighbourhood’ (analogous to an Output Area as defined by the ONS). Since the ONS reports that 79.6% of Output Areas (‘neighbourhoods’) contain 110–139 households,<sup>16</sup> and each postcode represents around 15 households on average, an LSOA was considered ‘coastal’ if 10 or more postcodes were within  $x$  kilometres of the coast. Values for  $x$  were 1, 3 or 5 km, thus producing three more candidate variables labelled ‘G\_10\_x’. Lastly, LSOAs were defined as coastal if a proportion of their postcodes were located within  $x$  kilometres of the coastline. The proportions tested were, 50% and 25% (‘G\_50\_x’ and ‘G\_25\_x’ candidates, respectively). Again the ‘distance’ portion (‘x’) of the definition was set at 1, 3 or 5 km for both proportional definitions.

#### Combination candidates (‘GR candidates’, ‘GD candidates’ and ‘GDR candidates’)

To ascertain whether the cancer-related health inequality gradients between coastal and non-coastal LSOAs were better defined when deprivation or rurality were part of the variable definition, the above ‘geography only’ candidates (‘G candidates’) were combined with other demographic information.

First, we restricted the definition of coastal to only include LSOAs that met the geographical criteria, and which were also in IMD quintiles 1 or 2 (ie, the two most deprived quintiles nationally). These candidate variables were analogously labelled ‘GD1’, ‘GD\_10\_x’, ‘GD\_25\_x’ and so on (‘GD candidates’).

Second, an LSOA would be labelled ‘coastal’ if it met the geographical criteria, and had a rural classification code beginning D, E or F (rural codes).<sup>17</sup> These candidate variables were labelled ‘GR1’, ‘GR\_10\_x’ and so on (‘GR candidates’).

Lastly, a fourth set of candidate variables were created that combined the geography, deprivation and rurality indicators into a more restricted definition. These would classify an LSOA as coastal if it met the geographical criteria, was in IMD quintile 1 or 2, and had a rurality code of D/E/F. These candidate variables were labelled ‘GDR1’, ‘GDR\_10\_x’ and so on (‘GDR candidates’).

Counts and descriptions of all the candidate variables tested can be found in S2.1 in online supplemental file 1.

#### Covariates

This study aimed to identify the ‘best’ candidate variable to explain variation in cancer outcomes, rather than to determine the best predictive model for each outcome.

Thus, no formal model-building procedure was undertaken, and we aimed for a parsimonious model to minimise overfitting. Covariates were chosen based on existing knowledge of cancer risk factors, and readily available public data. Mid-2016 population estimates by single year of age and gender were obtained from ONS (those closest in date to the start of the period used). From this, we calculated mean age of the population of each LSOA. The age of a community is an important confounder, since older populations are more likely to live in coastal areas<sup>18</sup> and cancer risk increases with age,<sup>19</sup> with worse outcomes for those diagnosed at older ages.<sup>20</sup> The proportion of female residents per LSOA was also derived, as cancer rates differ significantly by gender.<sup>21</sup> The proportion of the population who were of the white ethnic group was calculated from 2016 figures downloaded from Nomis.<sup>22</sup>

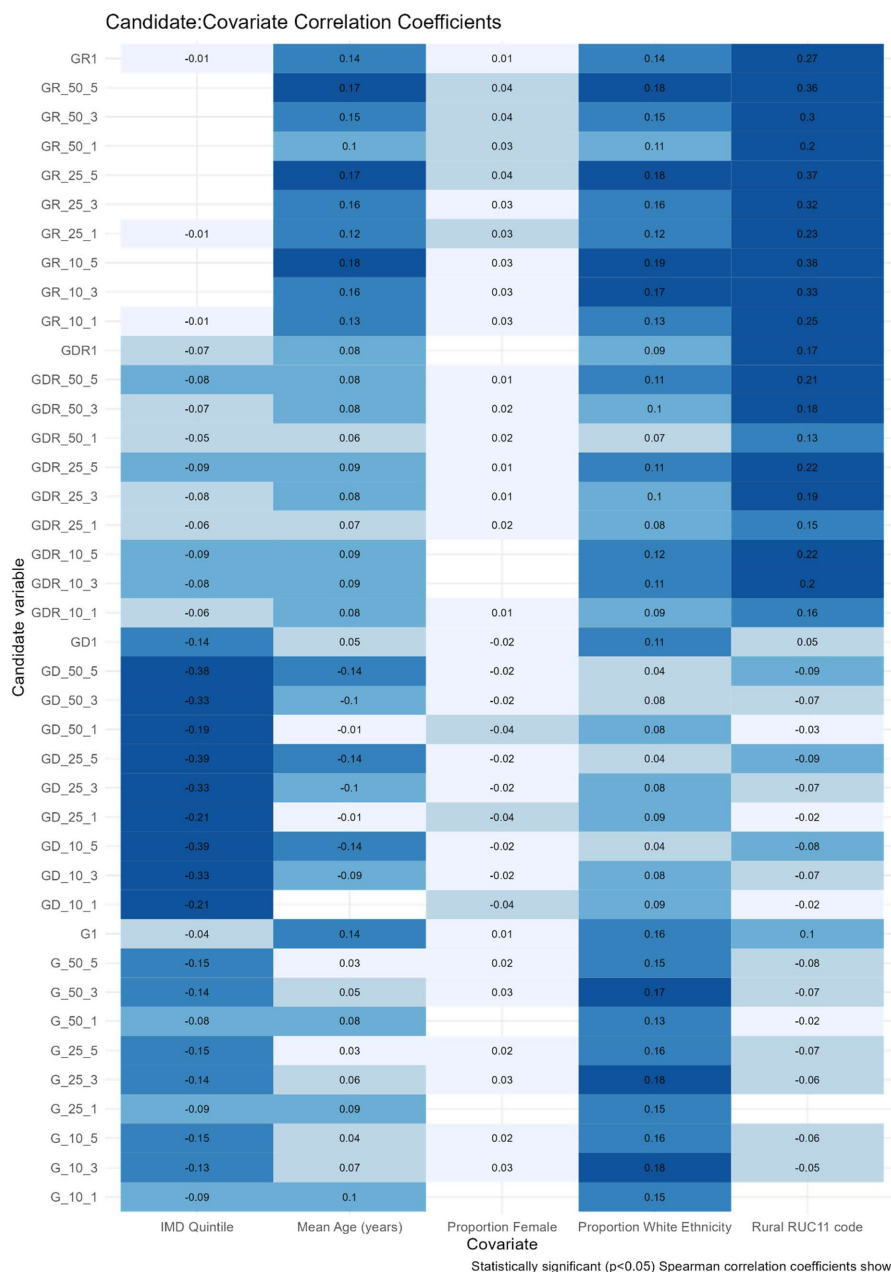
#### Statistical analysis

Descriptive analyses were undertaken to explore the variation of our outcomes and candidate variables. We visualised candidate variables using geographic mapping. Spearman correlation coefficient for the association of each candidate with each covariate was computed to identify potential multicollinearity issues, with an a priori cut-off of  $r = \pm 0.4$  (figure 1).

A set of linear regression models was created. Each cancer outcome (age-standardised incidence rate, crude prevalence and age-standardised mortality rate) was regressed against each candidate variable, additionally adjusted for mean population age, proportion of female residents, proportion of white ethnicity residents, IMD quintile and rural classification. Where a candidate variable incorporated either rural classification or IMD quintile (‘GD’, ‘GR’ and ‘GDR’ candidates), those covariates were excluded from the adjustment set to avoid over fitting.

Residual plots for each model were visually inspected for outliers and heteroskedasticity, leading to the omission of two outlier LSOAs for which data quality issues may have caused misleading figures. After excluding the Isles of Scilly and City of London ( $n=7$  LSOAs), the final dataset used in all models thus included complete data for 32 835 LSOAs.

For each metric, all candidate variables whose Wald  $p$  value was  $<0.05$  within their fully adjusted model were ranked according to Bayesian Information Criterion (BIC), which was scaled within each cancer metric to have a mean of 0 and a SD of 1 (BIC within each metric was scaled by subtracting the mean and dividing by the SD of the metric). The candidate variable whose model produced the lowest scaled BIC was selected as the ‘best’ model for that outcome/statistic combination. By adjusting for key socioeconomic and demographic characteristics, we aimed to move beyond simply mapping prevalence, thus strengthening our confidence that the chosen variable would be capturing more of the as-yet



**Figure 1** Heatmap of Spearman correlation coefficients for the associations between each candidate coastal variable and each covariate. Non-statistically significant ( $p$  value  $> 0.05$ ) correlations are not shown. Darker blue boxes indicate larger absolute coefficient values. IMD, Index of Multiple Deprivation

ill-defined health inequality dimension that underpins the residual variation in outcomes.

### Sensitivity analyses

This study aimed to identify a descriptive candidate for coastal cancer health inequality mapping, rather than to produce causal estimates, inferences or predictions. We also planned to create a variable that would be easily recreated by stakeholders using accessible information, therefore we employed a simple model that adjusted for the most common confounders in such research: socio-economic deprivation (IMD), rurality index, population age structure, ethnicity and gender. We did, however, consider an additional set of confounders in a sensitivity

analysis to confirm that our initial adjustment set was capturing most of the variation in outcomes. To do this, we additionally adjusted for LSOA-level variables as follows: distance to the nearest hospital (in minutes), the Air quality Domain Score and distance to nearest general practitioner (GP) practice (minutes) (downloaded from the Consumer Data Research Centre Access to Healthy Assets and Hazards (AHAAH) Version 3<sup>23</sup>), and for the index of the Small Area Mental Health Index (SAMHI) downloaded from the Place-Based Longitudinal Data Resource. Modelling and candidate selection proceeded as described above.



We sought to validate our candidate variable as a proxy for variation in coastal health by using a non-cancer outcome, which is known to vary by coastal status. We therefore augmented our analyses by assessing the significance of the coastal variables in fully adjusted models (same covariates as for cancer models) for the prevalence of coronary heart disease (CHD), which is known to be associated with coastal proximity.<sup>7</sup> CHD prevalence as percentage of LSOA population in 2016 was downloaded from the Place-Based Longitudinal Data Resource and was used in equivalent regression models to cancer outcomes.<sup>24</sup>

### Patient and public involvement

Patients and the public were not involved in the design or production of this work.

## RESULTS

After data cleaning, there remained 32 835 LSOAs with complete data available for analysis, covering a population of 55.3 million inhabitants. Mean population size per LSOA was 1683 persons (range 422–11 514 persons).

Spearman correlation coefficients between each candidate variable ( $n=40$ ) and each outcome metric ( $n=3$ ) did not exceed  $r = \pm 0.4$ , suggesting that multicollinearity was limited (figure 1).

All 40 candidate coastal variables were significantly associated with at least one cancer metric (incidence, prevalence or mortality) in fully adjusted models (figure 2). All geography-only, deprivation and rurality candidates ('G', 'GD' and 'GR' candidates) significantly improved models of two of the three cancer statistics, whereas all but one of the 'GDR candidates' were significantly associated with incidence, prevalence and mortality.

The 'best performing' (ie, lowest scaled BIC) candidate variable for capturing coastal-related variation in cancer incidence and prevalence was the geography-only model 'G\_25\_5'. Table 1 describes the distributions of model covariates across those LSOAs deemed 'coastal' by the 'G\_25\_5' model. This candidate variable assigns LSOAs to the 'coastal' identifier if 25% or more of their postcodes have centroids within 5 km of the coastline. When considering cancer mortality however, no candidate variables were a significant addition to the model. In this case, the 'best performing' candidate variable for capturing coastal-related variation in cancer mortality was 'GR\_10\_5', which assigns LSOAs with 10 or more postcodes within 5 km of the coast, provided they are also rural (figure 2). This candidate variable was not found to significantly improve the model of mortality rate based on likelihood ratio tests, despite the increased mortality rate in coastal areas as defined by this variable (table 2). This suggests that the confounder variables included in our models were responsible for most of the variability in cancer mortality, leaving little explanatory power for the 'GR\_10\_5' variable to account for. We therefore conclude that, when considering LSOA-level cancer

mortality, additional adjustment for coastal location is not advisable.

Table 2 presents summary statistics for the regression models presenting the associations between coastal regions (defined here using 'G\_25\_5') and cancer outcome statistics. We show that coastal areas have poorer cancer outcomes compared with non-coastal areas (after adjusting for known covariates). In coastal areas, age-standardised incidence was 9.1 per 100 000 persons higher (95% CIs=5.6 to 12.6), crude prevalence was 6.4 per 100 000 people higher (95% CIs=4.4 to 8.4) and age-standardised mortality was 1.6 per 100 000 people higher (95% CIs=-0.8 to 4.0) compared with non-coastal LSOAs.

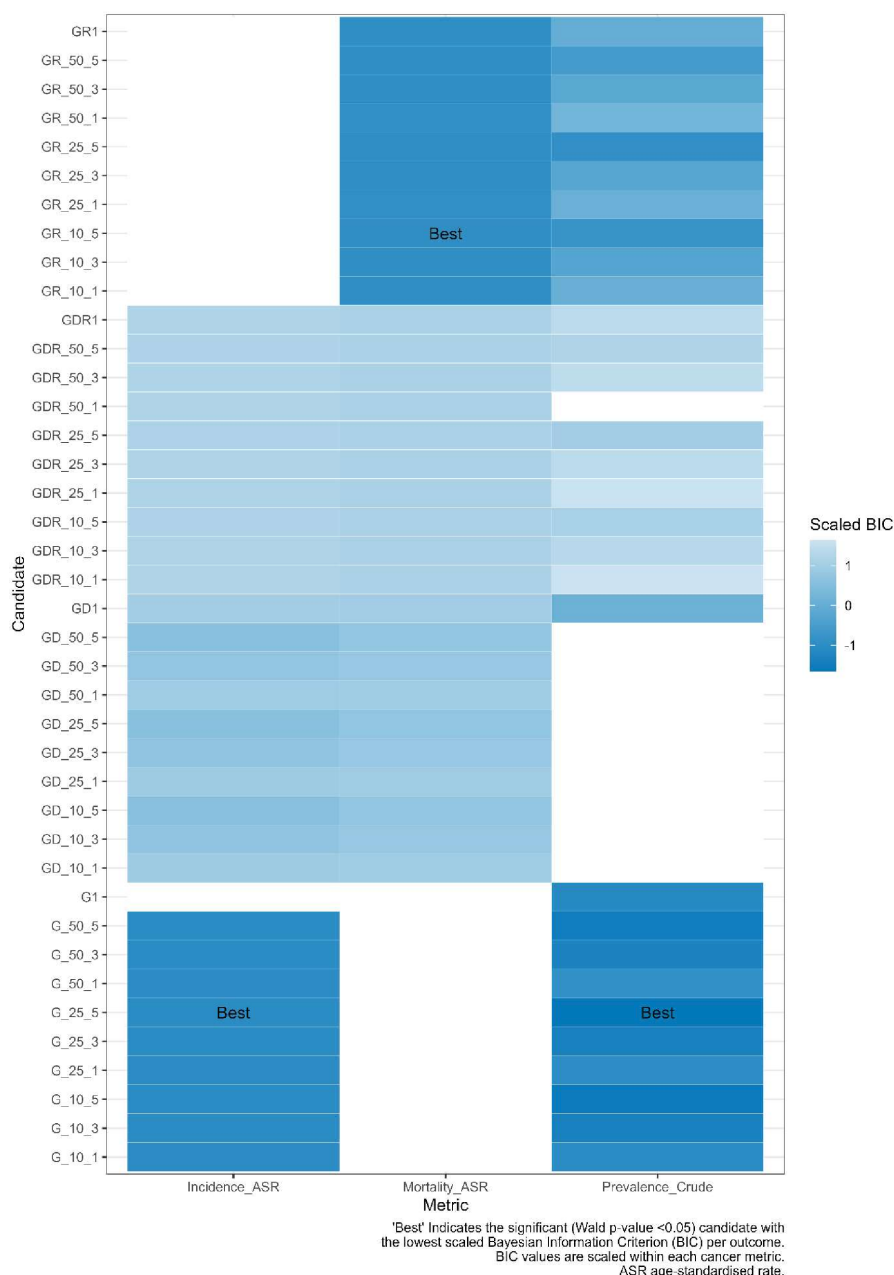
Our first sensitivity analysis sought to augment our adjustment set to account for a wider range of potential confounders that may be associated with coastal health inequality. As such we added variables describing LSOA-level average distance to hospital, and to GPs, alongside a measure of air quality and one of mental health, to our existing adjustment. These additions did not alter the choice of 'best' coastal variable per outcome and made no significant difference to the adjusted  $R^2$  values from the original models, suggesting overfitting (data not shown). We opted, therefore, to keep our adjustment set as originally described.

As a secondary sensitivity analysis to ensure that the chosen coastal variable was not only suitable for reporting of coastal health inequalities in cancer, we used LSOA-level CHD prevalence as an additional health outcome. The 'G\_25\_5' model was also a significant addition to the fully adjusted model of CHD (table 2).

To allow for swift uptake of our coastal variable to health data analyses, we have included in the supplement the list of 2011 coastal LSOAs (defined by 'G\_25\_5', see online supplemental file 3.1). Although data were not available at the time of writing to revalidate our candidate using more recent data, we have provided the list of coastal LSOAs from the 2021 census by applying our 'G\_25\_5' candidate method to these more recent boundaries (see online supplemental file 3.2).

## DISCUSSION

The National Disease Registration Service (NDRS) is committed to fulfilling the NHS's Public Sector Equality Duty by publishing National and Official statistics broken down by health inequality dimensions, as standard, wherever this is feasible and compliant with data privacy rules. Currently, most regular reports of cancer epidemiology in England are provided broken down by age group, gender, ethnic background, IMD quintile and geography. Despite its inclusion in NHS England's 'Core20PLUS5' framework, publishing by coastal status has not been implemented in these reports yet, because there is no agreed definition of 'coastal communities'. This study has sought to rectify this deficit by nominating a simple, parsimonious definition that can be replicated throughout cancer data reporting. Our coastal definition



**Figure 2** Candidate variable significance and model scaled BIC for each cancer statistic (all cancers, combined years 2016–2020). All models were adjusted for mean population age, proportion female residents and proportion white ethnicity residents, plus IMD quintile and rural urban classification index where relevant. White cells indicate non-significance. BIC, Bayesian Information Criterion; IMD, Index of Multiple Deprivation.

is at LSOA-level because this is a widely used census geography in the UK, and because geographies of lower granularity mask important intercommunity variation in health outcomes. We have chosen a candidate variable that best explained differences in cancer outcomes in the presence of socioeconomic adjustment, but as the definition of ‘coastal’ chosen excluded these factors, our candidate variable still enables adjustment or stratification by accepted measures of deprivation and rurality, thus allowing multidimensional exploration of geographical variation in cancer outcomes.

Despite a range of evidence for health inequalities for coastal communities,<sup>3 5 6 8 25</sup> data and insight into this

area have been hampered by the lack of a coherent definition of ‘coastal’ that allows for the inclusion of both rural and urban coastal communities and is sufficiently granular to avoid the ‘masking’ of pockets of deprivation by more wealthy nearby areas. A report by Atterton *et al* (2006)<sup>5</sup> employed the Vickers classification of Coastal Britain, which uses a local authority-level classification derived from a principal component analysis and k-means clustering algorithm, using initially 129 different geographical and demographical variables.<sup>26</sup> A report by Public Health England in 2020 stated that those older people living in coastal areas may be at higher risk of social isolation and loneliness.<sup>25</sup> This

**Table 1** LSOA-level characteristics for coastal and non-coastal areas, as defined by the 'G\_25\_5' candidate model

	Not coastal (n=25 466)	Coastal (n=7369*)	Overall (N=32 835)
Index of Multiple Deprivation quintile (1=most deprived)			
Mean (SD)	3.11 (1.42)	2.61 (1.33)	3.00 (1.41)
Median (Min, Max)	3.00 (1.00, 5.00)	2.00 (1.00, 5.00)	3.00 (1.00, 5.00)
Mean population age in 2016 (years)			
Mean (SD)	40.1 (5.23)	40.9 (6.30)	40.3 (5.50)
Median (Min, Max)	40.2 (20.8, 62.6)	40.5 (23.2, 64.4)	40.3 (20.8, 64.4)
Rural			
Not rural (RUC 2011 codes A-C)	20 780 (81.6%)	6458 (87.6%)	27 238 (83.0%)
Rural (RUC 2011 codes D, E or F)	4686 (18.4%)	911 (12.4%)	5597 (17.0%)
Proportion of female residents in 2016			
Mean (SD)	0.507 (0.0231)	0.507 (0.0269)	0.507 (0.0240)
Median (Min, Max)	0.509 (0.147, 0.655)	0.510 (0.250, 0.607)	0.509 (0.147, 0.655)
Proportion of white ethnicity residents in 2016			
Mean (SD)	0.862 (0.142)	0.891 (0.150)	0.869 (0.145)
Median (Min, Max)	0.916 (0.381, 1.00)	0.960 (0.381, 1.00)	0.932 (0.381, 1.00)
Age-standardised incidence rate of all cancers			
Mean (SD)	618 (138)	643 (136)	623 (138)
Median (Min, Max)	606 (8.46, 3820)	631 (140, 2010)	612 (8.46, 3820)
Age-standardised mortality rate of all cancers			
Mean (SD)	275 (99.8)	295 (101)	279 (100)
Median (Min, Max)	261 (0, 1700)	282 (0, 1670)	266 (0, 1700)
Crude prevalence of all cancers			
Mean (SD)	341 (122)	359 (144)	345 (127)
Median (Min, Max)	335 (5.20, 964)	348 (17.0, 986)	337 (5.20, 986)
RUC Rural Urban Classification Index 2011. *N.B. Fewer LSOAs are assigned to the 'coastal' group here than are mapped (see online supplemental file 2) due to modelling-specific exclusions. LSOA, Lower Super Output Area.			

report comprised a literature review of studies that referenced coastal health/inequalities. Their working definition of 'coastal' was 'any coastal settlement within a local authority area whose boundaries include the UK foreshore, including local authorities whose boundaries

only include estuarine foreshore. Coastal settlements include seaside towns, ports and other areas which have a clear connection to the coastal economy'. Unfortunately, 'clear' in this context, was not defined. This report also noted that there was a paucity of data on

**Table 2** Coastal ('G\_25\_5') variable coefficients and model diagnostics from fully adjusted linear models of each cancer outcome (age-standardised incidence rate, crude prevalence rate, age-standardised mortality rate) at LSOA level

	Coefficient (p value) (95% CI)			
	Age-standardised incidence rate	Crude prevalence rate	Age-standardised mortality rate	Coronary heart disease prevalence
G_25_5	9.11* (5.62, 12.61)	6.39* (4.44, 8.35)	1.59 (0.19) (-0.79, 3.97)	0.10* (0.08, 0.12)
n	32 835	32 835	32 835	32 835
R <sup>2</sup>	0.10	0.67	0.21	0.54

Rates are calculated based on total cancer cases or deaths from cancer between 2016 and 2020 in England, as recorded by the National Disease Registration Service and Office of National Statistics (for deaths data).

Models adjusted for IMD quintile, rural urban classification index 2011, mean population age, proportion female residents and proportion white ethnicity residents.

\*P<0.001.

IMD, Index of Multiple Deprivation; LSOA, Lower Super Output Area; ONS, Office of National Statistics.

coastal health in England, in comparison to studies of rural health inequalities.<sup>25</sup>

Asthana and Gibson (2022) mapped several health outcomes, including cancer, by Lower and Middle Super Output areas in England. Cardiovascular disease showed a distinct core/periphery distinction, as did several other health conditions. The working definition of 'coastal' in this study was LSOAs which include or overlap a built-up area of any size which lies within 500 m of the 'Mean High Water Mark' coastline. Again, it was noted that most health data reporting is made available at local authority or higher geographies, thus masking this core/periphery distinction. At the time of writing, several research groups are working to identify other definitions of 'coastal' for varied uses. This includes an ESRC funded project at the University of Plymouth and the ONS's Built-Up Areas classification. These have been produced or are aimed at seeking definitions to support policy research, whereas our focus was purely associated with cancer data reporting. We anticipate that in future, a range of novel disease-specific empirically derived variables may be produced, which may share similarities with those reported here. At that point, it may be that a more harmonised, cross-disciplinary definition can be ascertained collaboratively, but this work must start with the development of more subject-specific solutions, as we provide here. It would also be interesting to compare how our 'simpler' methodology compares to these more complex approaches at capturing coastal inequalities.

Using an holistic definition of cancer, we sought to identify a coastal variable that would be broadly applicable to all cancers, but we recognise that, just as IMD quintile may not be the most appropriate measure of deprivation as it relates to all diseases, having a single definition of a concept enables acceleration in research and understanding of that concept. We anticipate that our selected candidate variable will not be the statistically superior version to capture coastal health variation across all cancer types and metrics, but statistical precision must be balanced against pragmatic utility. A single, easily computed variable can be widely implemented quickly across all analyses, providing comparability, consistency, and clarity, which is preferable to a suite of variables whose use varies per disease and which cannot then be directly compared.

Although our focus with this study was to produce a method to report cancer data by coastal status, it is unclear whether this description would be applicable in other disease settings. We report a simple sensitivity analysis showing that our coastal variable captures a significant amount of residual variation in CHD diagnoses above our adjustment set as an illustration, but further exploration of other diseases was beyond the scope of this work. Understanding coastal variation in other diseases represents an interesting future direction for research. Additionally, future research should apply this work to understand the causal processes through which coastal

inequalities materialise to ensure that any measure of coastal status is capturing the correct information.

### Strengths and limitations

We employed LSOA-level cancer statistics derived from the NDRS databases, covering all of England over multiple years. This constitutes gold-standard cancer data at a very granular level, which is a major strength of this study. We also back-up our cancer-specific findings with corroboration of the usefulness of our candidate variable in a non-cancer setting, namely in the modelling of CHD, to illustrate the potential cross-disciplinary utility of our definition.

The cancer data used in this study comprised the combined years 2016–2020 to obtain enough case numbers across all cancer metrics for statistical analysis, and to avoid most of the COVID-19 era data which may have differed meaningfully in its characteristics.<sup>27</sup> Single-year analyses were beyond the scope of this study but given the relative stability of the presumed causal patterns at play, we feel that using this combined period was a valid approach.

Not all cancer types are associated with deprivation and demography in the same ways, for example, breast cancer incidence shows a seemingly paradoxical relationship with IMD.<sup>28</sup> When considering the relationship of coastal geography with varied cancer metrics as we have done here, we cannot guarantee that the chosen variable would be the 'best performing' candidate for all cancer types. This cancer-specific work was beyond the scope of this study but given that the principal aim of this work was to derive a single, maximally 'useful' coastal variable, we posit that differences in results for specific cancers would be of academic interest only at this stage and would not alter our recommendations.

Our cancer and confounder datasets were not available for the 2021 boundary definitions of LSOAs at the time of writing, meaning that we could not revalidate our model using more recent data to match the 2021 boundary changes. This may limit uptake of their usage. To minimise this issue, we have included the 2021 LSOA codes and their 'G\_25\_5' coastal designation in our online supplemental file 3.2. These analyses will be updated following the release of the relevant datasets.

X Claire Welsh @DrClaireEWelsh and Mark Alan Green @markalangreen

**Acknowledgements** This work uses data that has been provided by patients and collected by the NHS as part of their care and support. The data are collated, maintained and quality assured by the National Disease Registration Service, which is part of NHS England.

**Contributors** CW designed the study, carried out all analyses and drafted the manuscript, and is the guarantor of the study. MG contributed to the methodological approach, CQ, CatherineW and JL extracted the cancer data and drafted the Supplementary File S1 section. JA, SV and LP contributed to study design and critically revised the manuscript. All authors have read and approved the final manuscript for submission. ClaireW accepts full responsibility for the finished work and/or the conduct of the study, had access to the data and controlled the decision to publish.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.



**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Ethics approval** Not applicable. Specific ethical approval for this study was not required per the definition of research according to the UK Policy Framework for Health and Social Care Research.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. Cancer data used in this study are available upon request and subject to data protection limitations, from NDRS via the Data Access Request Service (<https://digital.nhs.uk/services/data-access-request-dars>). Other data sources are publicly available at the cited locations.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Claire Welsh <http://orcid.org/0000-0001-9477-0775>

Mark Alan Green <http://orcid.org/0000-0002-0942-6628>

Lizz Paley <http://orcid.org/0000-0002-3879-5377>

#### REFERENCES

- McCartney G, Popham F, McMaster R, *et al*. Defining health and health inequalities. *Pub Health (Fairfax)* 2019;172:22–30.
- NHS England. Core20PLUS5 (adults) – an approach to reducing healthcare inequalities. 2023. Available: <https://www.england.nhs.uk/about/equality/equality-hub/national-healthcare-inequalities-improvement-programme/core20plus5/>
- Chief Medical Officer. Chief medical officer's annual report 2021 – health in coastal communities – summary and recommendations. 2021. Available: <https://www.gov.uk/government/publications/chief-medical-officers-annual-report-2021-health-in-coastal-communities>
- Ministry of Justice UG. Public sector equality duty. UK. 2012. Available: <https://www.gov.uk/government/publications/public-sector-equality-duty>
- Atterton J. Ageing and coastal communities centre for rural economy research report final report for the coastal action zone partnership. 2006.
- Asthana S, Gibson A. Averting a public health crisis in England's coastal communities: a call for public health research and policy. *J Public Health (Bangkok)* 2022;44:642–50.
- Xiao-Dong Z, Shao-Zhao Z, Xun H, *et al*. Association of Residential Proximity to the Coast With Incident Myocardial Infarction: A Prospective Cohort Study. *Front Cardiovasc Med* 2022;9:752964.
- Eurostat. Eurostat regional yearbook 2011: coastal regions. 2011. Available: <https://ec.europa.eu/eurostat/documents/3217494/>
- 5728589/KS-HA-11-001-13-EN.PDF.pdf/c0dd33ed-0db2-4d8b-ae03-26d9bf3e57fc?t=1414775982000
- Hooyberg A, Roose H, Grellier J, *et al*. General health and residential proximity to the coast in Belgium: Results from a cross-sectional health survey. *Environ Res* 2020;184:109225.
- Ministry of Housing Communities & Local Government. English indices of deprivation. 2020. Available: <https://www.gov.uk/government/collections/english-indices-of-deprivation>
- Cancer Research UK. Cancer in the UK 2020: socio-economic deprivation. 2020. Available: [https://www.cancerresearchuk.org/sites/default/files/cancer\\_inequalities\\_in\\_the\\_uk.pdf?\\_gl=1\\*16j0s4f\\*\\_ga\\*MjAzOTY3MDU2My4xNjQ1MjAyMjg1\\*\\_ga\\_58736Z2GN-N\\*MTY1NjY3NjM0MC40LjEuMTY1NjY3NjM2Ni4zNA..&\\_ga=2.208526097.269631211.1656672614-2039670563.1645202285](https://www.cancerresearchuk.org/sites/default/files/cancer_inequalities_in_the_uk.pdf?_gl=1*16j0s4f*_ga*MjAzOTY3MDU2My4xNjQ1MjAyMjg1*_ga_58736Z2GN-N*MTY1NjY3NjM0MC40LjEuMTY1NjY3NjM2Ni4zNA..&_ga=2.208526097.269631211.1656672614-2039670563.1645202285)
- National Cancer Registration and Analysis Service. Cancer registration statistics, England 2020. 2022. Available: <https://digital.nhs.uk/data-and-information/publications/statistical/cancer-registration-statistics/england-2020>
- Department for Environment Food and Rural Affairs. Rural urban classification. 2021. Available: <https://www.gov.uk/government/statistics/2011-rural-urban-classification>
- NDRS. National disease registration service. 2023. Available: <https://digital.nhs.uk/ndrs>
- Henson KE, Elliss-Brookes L, Coupland VH, *et al*. Data Resource Profile: National Cancer Registration Dataset in England. *Int J Epidemiol* 2020;49:16.
- Office for National Statistics. Output areas: introduction to output areas – the building blocks of census geography. 2016. Available: <https://www.ons.gov.uk/census/2001censusandearlier/dataandproducts/outputgeography/outputareas>
- Bibby P, Brindley P. The 2011 rural-urban classification for small area geographies: a user guide and frequently asked questions (v1.0). 2013. Available: [www.nationalarchives.gov.uk/doc/open-government-licence/](http://www.nationalarchives.gov.uk/doc/open-government-licence/)
- Public Health England. Health inequalities in ageing in rural and coastal areas. 2019. Available: <https://www.gov.uk/government/publications/health-inequalities-in-ageing-in-rural-and-coastal-areasfv>
- Laconi E, Marongiu F, DeGregori J. Cancer as a disease of old age: changing mutational and microenvironmental landscapes. *Br J Cancer* 2020;122:943–52.
- Van Herck Y, Feysaerts A, Alibhai S, *et al*. Is cancer biology different in older patients? *Lancet Healthy Longevity* 2021;2:e663–77.
- Lopes-Ramos CM, Quackenbush J, DeMeo DL. Genome-Wide Sex and Gender Differences in Cancer. *Front Oncol* 2020;10:597788.
- Office for National Statistics. Nomis – official census and labour market statistics. 2023. Available: <https://www.nomisweb.co.uk/>
- Green MA, Daras K, Davies A, *et al*. Developing an openly accessible multi-dimensional small area index of 'Access to Healthy Assets and Hazards' for Great Britain, 2016. *Health Place* 2018;54:11–9.
- University of Liverpool. Place-based longitudinal data resource. 2023. Available: <https://pldr.org/about-us/>
- Haighton C, Dalkin S, Brittain K. An evidence summary of health inequalities in older populations in coastal and rural areas. London. 2019. Available: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/824723/Health\\_inequalities\\_in\\_Ageing\\_in\\_Rural\\_and\\_Coastal\\_Areas-Full\\_report.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/824723/Health_inequalities_in_Ageing_in_Rural_and_Coastal_Areas-Full_report.pdf)
- Vickers D, Rees P, Birkin M. A new classification of UK local authorities using 2001 census key statistics. n.d. Available: <http://eprints.whiterose.ac.uk/5005/><https://eprints.whiterose.ac.uk/>
- Maringe C, Spicer J, Morris M, *et al*. The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a national, population-based, modelling study. *Lancet Oncol* 2020;21:1023–34.
- Smith D, Thomson K, Bamba C, *et al*. The breast cancer paradox: A systematic review of the association between area-level deprivation and breast cancer screening uptake in Europe. *Cancer Epidemiol* 2019;60:77–85.