Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
An exploration of factors characterising unusual spatial clusters of COVID-19 cases in the East Midlands region, UK: A geospatial analysis of ambulance 999 data

Harriet Elizabeth Moore a, Bartholomew Hill b,c, Niro Siriwardena b, Graham Law b, Chris Thomas b, Mark Gussy b, Robert Spaight b, Frank Tanser b

a DIRE Research Group Lead, UK
b EDGE Consortium Affiliates, UK
c Loughborourgh University Water Engineering and Development Centre, UK

HIGHLIGHTS

- Air quality and features of urban landscapes are risk factors for COVID-19.
- Deprived areas face different challenges for mitigating contagion compared to affluent areas.
- Identifying clusters of COVID-19 transmission could be used to inform 'isolate, test, trace'.
- Ambulance calls reflect acute cases of COVID-19 and could be used for pre-hospital triage.
- Factors that are associated with clusters are highly location specific.

ARTICLE INFO

Keywords: COVID-19, Vulnerability, Bioecological model, Exposure, Underlying susceptibility, Built environments

ABSTRACT

Complex interactions between physical landscapes and social factors increase vulnerability to emerging infections and their sequelae. Relative vulnerability to severe illness and/or death (VSID) depends on risk and extent of exposure to a virus and underlying health susceptibility. Identifying vulnerable communities and the regions they inhabit in real time is essential for effective rapid response to a new pandemic, such as COVID-19. In the period between first confirmed cases and the introduction of widespread community testing, ambulance records of suspected severe illness from COVID-19 could be used to identify vulnerable communities and regions and rapidly appraise factors that may explain VSID. We analyse the spatial distribution of more than 10,000 suspected severe COVID-19 cases using records of provisional diagnoses made by trained paramedics attending medical emergencies. We identify 13 clusters of severe illness likely related to COVID-19 occurring in the East Midlands of the UK and present an in-depth analysis of those clusters, including urban and rural dynamics, the physical characteristics of landscapes, and socio-economic conditions. Our findings suggest that the dynamics of VSID vary depending on wider geographic location. Vulnerable communities and regions occur in more deprived urban centres as well as more affluent peri-urban and rural areas. This methodology could contribute to the development of a rapid national response to support vulnerable communities during emerging pandemics in real time to save lives.

1. Introduction

There is growing recognition in the fields of epidemiology (Miller et al., 2012; Norris et al., 2020; Viegi et al., 2006), and urban planning (Durand et al., 2011; Seo et al., 2019) that complex interactions between physical landscapes and social factors drive vulnerability to contagion and severe illness, and thus, understanding these mechanisms should be the focus of disease prevention and management. Urban landscapes can simultaneously influence chronic conditions associated with susceptibility to severe pathogenic illness, such as obesity by facilitating access to fast food vendors (Duras et al., 2018), and increase exposure to disease through high density urbanism (Goryakin et al., 2017; Wu et al., 2017).

On March 11th, 2020 the World Health Organization declared the novel coronavirus disease 2019 (COVID-19) a global pandemic. Over the subsequent months, the research community has prioritised understanding contagion and transmission pathways (Park, Thwaites, & Openshaw, 2020) as well as identifying and supporting vulnerable
communities and regions (Daras et al., 2021; Khalatbari-Soltani et al., 2020; Patel et al., 2020). Vulnerable communities include those with pre-existing chronic conditions that are known to increase susceptibility to severe COVID-19 illness, such as diabetes (Peric & Stulnig, 2020) and overweight or obesity (Steinberg et al., 2020). Medical and public health research has led to widespread appreciation that the impact of the pandemic has, and continues to be, heterogenous; some communities and regions appear to be more vulnerable to severe illness than others (Marmot et al., 2020; Patel et al., 2020).

One common observation is that urban and peri-urban areas tend to be sites of high rates of infection and mortality compared to more dispersed rural areas (Stier et al., 2020). High rates of infection suggest that urban landscapes are more exposed to transmission, and, given that many cases of infection are asymptomatic or involve very mild symptoms (Kim et al., 2020), high rates of mortality may indicate that communities within those landscapes are also more susceptible to severe illness (Guilmoto, 2020). Characteristics of urban landscapes that are typically associated with the transmission of infectious diseases include population and employment density (Hu et al., 2015), and housing crowdedness (Low et al., 2013; Neiderud, 2015). However, these relationships are rapidly changing and vary depending on region and specific location within urban areas. For example, ‘extended urbanisation’ is shifting the dynamics of vulnerability; in some cases, communities on urban peripheries may be more vulnerable than those in denser urban centres with greater access to healthcare and social support (Connolly et al., 2021). Indeed, in the case of COVID-19, typical relationships between urban space and infectious disease do not consistently explain mortality, with high rates of severe cases occurring in less dense urban areas (Frank & Wali, 2021). Thus, there is a need to consider how urban landscapes influence the underlying susceptibility of communities to severe illness as well as exposure to infectious diseases.

Surprisingly, of more than 40,000 papers using clinical diagnoses of COVID-19 published in 2020, <150 have considered the implications for urban planning, and those that do tend to focus on the impact that ‘lockdown’ has had on urban landscapes, such as pollutant rates, rather than the impact of urban landscapes on health outcomes (Sharifi & Khavarian-Garmsir, 2020). In this study, we consider the relationship between built environments, exposure to emerging infectious diseases (EIDs), and susceptibility to severe COVID-19 illness provisionally diagnosed by medically trained professionals. In this context, ‘severe illness’ refers to patients presenting with severe symptoms that require the attendance of emergency medical services.

Efforts to shield the most vulnerable communities and regions in society are more likely to be effective if they happen rapidly and in real-time (Kasda et al., 2020). Compared to other common disasters like flooding, obtaining geographically accurate data to evaluate the spatial dimensions of a pandemic, and to support vulnerable communities and regions, faces unique challenges. In the UK, the use of contact tracing and testing to identify community cases of COVID-19 commenced after the first confirmed case on January 31st, 2020. However, community testing ceased in early March as cases rose rapidly and the virus was classified as a category 3 pathogen, confining testing to level 3 laboratories. On April 2nd the UK government outlined a five-pillar strategy for expanding testing capacity including the introduction of community testing. On April 2nd the UK government outlined a five-pillar strategy for expanding testing capacity including the introduction of community (Connolly et al., 2021). Indeed, in the case of COVID-19, typical relationships between urban and infection and mortality vary depending on region and specific location within urban areas. For example, ‘extended urbanisation’ is shifting the dynamics of vulnerability; in some cases, communities on urban peripheries may be more vulnerable than those in denser urban centres with greater access to healthcare and social support (Connolly et al., 2021). Indeed, in the case of COVID-19, typical relationships between urban space and infectious disease do not consistently explain mortality, with high rates of severe cases occurring in less dense urban areas (Frank & Wali, 2021). Thus, there is a need to consider how urban landscapes influence the underlying susceptibility of communities to severe illness as well as exposure to infectious diseases.

Surprisingly, of more than 40,000 papers using clinical diagnoses of COVID-19 published in 2020, <150 have considered the implications for urban planning, and those that do tend to focus on the impact that ‘lockdown’ has had on urban landscapes, such as pollutant rates, rather than the impact of urban landscapes on health outcomes (Sharifi & Khavarian-Garmsir, 2020). In this study, we consider the relationship between built environments, exposure to emerging infectious diseases (EIDs), and susceptibility to severe COVID-19 illness provisionally diagnosed by medically trained professionals. In this context, ‘severe illness’ refers to patients presenting with severe symptoms that require the attendance of emergency medical services.

Efforts to shield the most vulnerable communities and regions in society are more likely to be effective if they happen rapidly and in real-time (Kasda et al., 2020). Compared to other common disasters like flooding, obtaining geographically accurate data to evaluate the spatial dimensions of a pandemic, and to support vulnerable communities and regions, faces unique challenges. In the UK, the use of contact tracing and testing to identify community cases of COVID-19 commenced after the first confirmed case on January 31st, 2020. However, community testing ceased in early March as cases rose rapidly and the virus was classified as a category 3 pathogen, confining testing to level 3 laboratories. On April 2nd the UK government outlined a five-pillar strategy for expanding testing capacity including the introduction of community testing. On April 2nd the UK government outlined a five-pillar strategy for expanding testing capacity including the introduction of community testing.

Thus, ambulance records may be a reliable measure of severe COVID-19 related illness in real-time. To our knowledge, ambulance data have not previously been utilised to identify communities and regions that may be vulnerable to severe illness from COVID-19 or to investigate the social and environmental factors that may influence vulnerability. Our novel methodology presents an opportunity for health professionals to identify and support vulnerable communities who are likely to experience severe illness from a new virus in the early phase of a pandemic.

Ambulance records of provisional diagnosis hold several important advantages over hospital admittance records and laboratory records of confirmed cases for identifying vulnerable communities affected by severe illness. Hospitals and laboratories are required under General Data Protection Regulation (GDPR) to aggregate patient information for reporting. Others have explored socio-economic predictors of aggregate confirmed cases in the UK at less granular scales (eg, Daras et al., 2021). However, without costly and time-consuming data linkage via NHS Digital, aggregated patient data does not allow for meaningful analysis of spatial patterns or characteristics of physical and social environments that explain geographical trends of severe illness. Ambulance data, including postcode region, allows a more granular analysis of factors that predict severe illness from COVID-19 infection in real-time.

Our study presents a novel methodology for identifying communities and regions that are vulnerable to severe illness during the early phase of a pandemic before laboratory testing is widespread. This involves considering both risk of exposure to a contagious disease as well as underlying susceptibility to severe illness. We identify unusual clusters of provisionally diagnosed severe COVID-19 cases in real-time using medical data collated by EMAS. Provisional diagnostic of suspected COVID-19 is determined by paramedics based on observed signs, such as patient acuity, self-reported symptoms, such as OTD, and objective medical measures, such as blood oxygen levels. Our analysis explores the characteristics of communities and regions within built environments where unusual clusters occur, including landscape features and socio-economic dynamics. In this context, ‘unusual clusters’ refers to high numbers of cases occurring within spatial proximity that are unlikely to have occurred by chance. Taken together, these analyses offer a real-time approach for identifying and protecting vulnerable communities in the critical early stages between the first confirmed case of a new EID and widespread community testing, as well as for identifying the characteristics of those communities most affected by severe illness over the course of a pandemic.

1 UK Postcodes include two components, for example LN6 7TS. Region postcodes include the first component and the first letter of the second component; LN6 7.
2. Conceptualising the relationship between severe suspected COVID-19 cases and built environments:

Theories about individual health and wider environments emerged in the 1980s as a critical response to medical and epidemiological paradigms; traditional medical models conceptualise health in terms of the presence or absence of biological disease, and the outcome of exposures that occur entirely at the level of the individual (Barbour, 1997). More contemporary ‘social’ perspectives suggest that health transcends the individual and recognize the important role that social networks (Smith et al., 2018) and wider social environments, including deprivation, play in health outcomes (Marmot, 1998).

Efforts to consider interactions between individual or biological factors and social factors often draw on Bronfenbrenner’s bioecological theories (Eriksson & Hammarstrom, 2018), including his original Ecological Social Model, as well as the more recent Process-Person-Context-Time model (Repetti, Taylor, & Seeman, 2002). These frameworks advanced the field of public health by introducing a way to conceptualise the multi-level social interactions that influence health and wellbeing. Bronfenbrenner’s models divide the social world of an individual into four ‘systems’; the Microsystem, including the most immediate elements of the social world, such as family, the Mesosystem, including extended social networks, the Exosystem, including wider community services, and the Macrosystem including commonly shared cultural and social beliefs and values (Bronfenbrenner, 1979).

While progressive, bioecological theories focus almost exclusively on the social world and social vulnerabilities and rarely consider landscape features that are also understood to influence health outcomes (Campbell & Wiesen, 2009; Cervero & Duncan, 2003; Williams, 2016). In contrast, in the field of urban planning, vulnerability is often conceptualised in relation to hazards and risks in the landscape, such as exposure to direct communicable disease, as well as more distal relationships, or ‘teleconnections’ (Seto et al., 2012) between landscape features, like access to green space (De Vries, Van Dillen, Groenewegen, & Spreeuwenberg, 2013; Markveych et al., 2017) and underlying health conditions, such as obesity (Daras et al., 2018). Importantly, exposure to a virus does not necessarily precipitate a medical emergency, rather, severe symptoms requiring emergency medical attention reflect the cumulative effect of exposure and underlying susceptibility. Thus, vulnerability is multifaceted, incorporating components of the physical landscape and components of the social world. While both bioecological and urban risk theories have advanced ways of thinking about health outcomes and pathways of vulnerability, a holistic approach is needed that considers the range of factors in built environments that precipitate severe illness and death from COVID-19.

In the case of COVID-19, the relationship between severe illness and characteristics of the built environment involves both direct and indirect pathways. On the one hand, environments can influence the direct exposure of individuals to communicable disease. On the other hand, landscape features can indirectly affect the underlying susceptibility of communities to severe symptoms, compared to experiencing mild symptoms or presenting as asymptomatic, by supporting or preventing healthy lifestyles. Features of neighbourhoods that can influence health behaviours like exercise include access to green space for passive recreation such as walking, and facilities for active exercise, such as sports grounds or leisure centres (Hartig et al., 2020). Further distance from these healthy landscape features is associated with lower levels of activity and a higher risk of cardiovascular disease (Shen & Lung, 2016) and obesity (Lachowycz & Jones, 2011). However, landscape features can also reflect the social characteristics of wider living environments; high crime rates co-occur with poor physical infrastructure like housing in deprived communities. Crime can deter access to nearby outdoor spaces (Gomez et al., 2004) while poor housing indicates lower incomes and a greater likelihood of underlying chronic health conditions (Krieger & Higgins, 2002).

While deprivation broadly is associated with susceptibility to severe illness, the socio-economic characteristics of patients with severe symptoms of COVID-19 have often been overlooked. Thus, Khalatbari-Soltani et al. (2020) call for the systematic recording of these dynamics for identifying vulnerable groups in the early stages of a pandemic. Patel et al. (2020) suggest that deprivation is likely to be associated with increased VSID from COVID-19 in three ways. Firstly, more deprived neighbourhoods often experience overcrowding which results in increased risk of infection compared to less densely populated areas. Secondly, poorer people are more likely to be employed in roles without opportunities to work from home which also increases risk of exposure. Finally, poverty is a risk factor for chronic comorbidities that in turn predict severe illness and hospitalisation from COVID-19, such as cardiovascular disease (Mehra et al., 2020), diabetes (Peric & Stulnig, 2020) and obesity (Steinberg et al., 2020).

Others have investigated the relationship between severe COVID-19 related illness and individual features of social worlds, such as deprivation (e.g., Patel et al., 2020) and physical landscapes, such as air pollution (Travaglio et al., 2021). We consider the cumulative impact of factors across demographic, socio-economic and environmental domains to explore the characteristics of vulnerable communities (Klaghadi et al., 2020) identified spatially by unusual clusters. This approach resounds well with the underlying philosophy of bioecological modelling; health outcomes are the culmination of interactions between and within domains that make up the built environment, and across individual and neighbourhood scales. Thus, in addition to social interactions, we include features of physical landscapes in our analysis to consider vulnerability across scales in Bronfenbrenner’s socio-ecological landscape.

3. Methods:

3.1. Site and location

The East Midlands is located in the Central Eastern part of England and spans an area of 15,627 km² (Fig. 1). The estimated total population of the region is 4.8 million including the most populous urban areas of Derby, Leicester, Lincoln, Northampton and Nottingham (Office of National Statistics, 2020a). The proportion of the population identifying as White UK in the East Midlands is low (14.6%) compared to the national average (20.2%) (Office of National Statistics, 2020b), although some regions, including Leicester, have a much higher proportion of non-white population. In 2016, 18.5% of people in the region lived in the most deprived quintile (Public Health England, 2018). Nottingham, Derby and Leicester are the economic core of this region, with around 48% of businesses, and 50% of the population located in these cities (European Commission, 2020). The East Midlands is also the 3rd most rural region in England (European Commission, 2020).

3.2. Research aims and questions

The first aim of the research was to identify unusual clusters of suspected COVID-19 cases in the East Midlands of the UK using more than 10,000 records of provisional diagnoses for severe COVID-19 collated by EMAS during the first ‘wave’ of the pandemic between March 2nd and May 11th (Kontopantelis et al., 2021). This was achieved using a Kulldorff spatial scan statistic implemented in the geospatial software SatScan™ which compares the actual distribution of cases to the predicted distribution based on population density. The null hypothesis tested is that cases are randomly distributed rather than occurring in unusual clusters. The second aim of the research was to explore factors that predict cluster membership. This analysis involved computing a binary logistic regression with variables including measures of patient demographics, deprivation, and landscape features (Section 3.3). The third aim was to elucidate the individual characteristics of each unusual COVID-19 cluster, using geospatial analysis and mapping to determine the strongest predictors of cluster membership.
3.3. Measures

Table 1 summarizes the datasets and measures included in the research. Data collated by and obtained from EMAS includes provisional diagnoses of suspected COVID-19 by medically trained clinicians, age, and sex. More severe COVID-19 symptoms tend to be associated with older age (Romero Starke et al., 2020), and mortality is nearly twice as high in males compared to females (Ortolan et al., 2020). While ethnicity is also commonly associated with severe symptoms (Sze et al., 2020), reliable data was unavailable in real-time. The diagnostic algorithm employed by medically trained clinicians is guided by Public Health England’s case definition criteria (Public Health England, 2018), including observations of illness, self-reported symptoms like OTD, and objective medical measures like blood oxygen levels.

On this basis, we could anticipate that unusual clusters of suspected underlying health conditions like diabetes and obesity (Green et al., 2018), hazardous features of landscapes, and closer proximity to retail vendors like fast food and sex. More severe COVID-19 symptoms tend to be associated with high scores of IMD, AHAHI, and RUC values. Thus, the final dataset contained 10,345 geospatial points. Only call outs for provisionally diagnosed COVID-19 were included in the dataset.

3.4. Data handling and cleaning

The database of suspected COVID-19 cases was obtained from EMAS, including the date 999 calls were received, partial postcodes (rather than full addresses) of ambulance attendance locations, sex, and age. In total, 10,438 records were received, however, 93 records were removed because they contained errors, were missing geospatial information, or were unable to link to postcode population. All records of suspected COVID-19 cases were successfully linked to IMD, AHAHI, and RUC values. Thus, the final dataset contained 10,345 geospatial points. Only call outs for provisionally diagnosed COVID-19 were included in the dataset.

3.5. Statistical and spatial data analysis

Data analysis was conducted in three steps. Step one involved identifying unusual clusters of suspected severe COVID-19 cases by using population data as a baseline for the expected distribution of cases. For this analysis data were represented at the postcode region scale. The output included the location of statistically significant clusters, and a binary dataset distinguishing all cases that fell within clusters from all cases that fell outside of clusters. Step two involved converting the postcode region data to Lower Super Output Area scale for the purpose of linking the COVID-19 dataset with existing national datasets, including IMD, the AHAHI, and RUC. The output was a unique linked database combining clinical and landscape scale data. In step three, statistical analyses were conducted to identify demographic, socio-economic, and environmental factors that predicted cluster and non-cluster membership, and geospatial analysis was used to characterize each individual cluster.

3.5.1. Identifying unusually high clusters of suspected COVID-19 cases

We applied a Kulldorff spatial scan statistic (Discrete Poisson model) implemented in SatScan™ software version 9.6.1 to perform the spatial analysis scanning to detect unusual clusters of COVID-19 cases across the surveillance area. A spatial scan statistic is a cluster detection test that detects the location of clusters and evaluates their statistical significance (Kulldorff et al., 2005; Kulldorff, 1997). This was done by gradually scanning a window across the study area, noting the number of observed and expected observations, based on population (Office of National Statistics, 2011), inside the window at each location using a Discrete Poisson model. For any given position of the centre, the radius of the circle changes continuously so that it can take any value. For each circle, the spatial scan statistic calculates the likelihood of the observed number of cases occurring inside and outside of the circle. The circle with the maximum likelihood is the most likely cluster, and thus the least likely to have occurred by chance. This method tests the null hypothesis that clusters are randomly distributed. Statistically significant suggests that unusual spatial clustering is unlikely to have occurred by chance. The isotopic circular scan method employed by the software has previously been validated for identifying clusters of other infectious disease, such as malaria (Coleman et al., 2009), HIV (Namokha et al., 2013; Tanser et al., 2018), tuberculosis (Smith et al., 2018), and various chronic diseases (Cuadros et al., 2019; Tomita et al., 2020).

The Poisson Model was purely spatial. The model parameters included unconstrained spatial cluster size, and the criteria for reporting hierarchical clusters was set to ‘no cluster centres in other clusters’.

3.5.2. Data conversion to LSOA and database compilation

To compile the LSOA dataset, IMD, RUC and AHAHI scores were merged using the join tool in ArcGIS Pro 2.6.0. The join used Lower Super Output Area codes (LSOA11CD) as these identifiers are consistent between the EMAS COVID-19 database and the remaining datasets. These processes are visualised in Fig. 2.

Geospatial analysis was also used to identify which cases fall into specific clusters compared to cases that are randomly distributed. In one instance, two clusters were found to overlap. However, for the purpose of characterizing clusters it was necessary to assign all cases to a single cluster. Thus, these cases (N = 54) were assigned to clusters based on their location from a centre line of intersection between the overlapping
clusters. The output, a novel database, was used for regression analysis to identify factors that predict cluster membership (Section 5.2), and for geospatial analysis to produce maps representing teleconnections (Section 5.3).

3.5.3. Statistical analysis and spatial representation of significant clusters

Binary logistic regression analysis was used to identify factors that predict whether individual cases of suspected severe COVID-19 occur in unusual clusters. All measures reported in Table 3 and Table 4 were included in the regression model. While the IMD was included in the binary regression, we also conducted an ANOVA to determine whether mean differences in deprivation and affluence occur between areas with clusters and areas characterized by random distribution. In the UK, deprivation is often associated with early transmission patterns (Balsegaram et al., 2012), and high rates of contagion (Rushton et al., 2007). Thus, it is possible that deprivation is a common denominator for all areas with suspected cases of COVID-19, rather than a distinguishing feature of cluster membership. ANOVA was computed to explore more nuanced spatial differences between each cluster, and areas with cases that do not occur in clusters.

Regression output and cluster output from SatScan™ was used to display the relationship between determinants of clusters visually. The cluster output from SatScan™ was converted to a layer (‘cluster shapefile’) within ArcGIS Pro 2.6.0. Of 41 clusters identified, 13 were statistically significant (P < .05). All non-significant clusters were removed from the dataset. A polygon representing the East Midlands was extracted from the UK Counties 2017 shapefile (‘UK shapefile’) to create a background in ArcScene. Relative Risk values were assigned to each cluster within the cluster shapefile. Both the cluster shapefile and UK shapefile were converted to rasters and combined to create a unique raster displaying the Relative Risk of clusters in 3-D. The unique raster was then converted to a TIN in order to be represented clearly in ArcScene. This step addressed a display problem due to the rasters and polygons merging and warping the slope of elevation in the image. Displaying clusters involved using a scale of graduated colours from green to red that were manually selected based off the spread of the data.

4. Results:

4.1. Identifying unusually high clusters of suspected severe COVID-19 cases

SatScan™ Poisson Modelling identified 13 statistically significant (P < .05) unusually high clusters of suspected COVID-19 cases, displayed in Fig. 3. Per 100,000 population, the number of observed cases range from 951 West of Skegness to 3,417 East of Rugby. By comparison, the range of cases occurring per 100,000 outside of clusters was between 8 and 660. Fig. 4 demonstrates the relative risk of each cluster, meaning the likelihood of contracting severe illness in an area compared to regions where cases are randomly distributed. The spatial characteristics of each cluster, including approximate location, radius, expected and observed number of cases, P-Values, specific relative risk ratios, and the number of cases in each cluster per 100,000 population are reported in Table 2.

4.2. Factors that predict cases of COVID-19 falling into unusual clusters compared to randomly distributed cases

4.2.1. Descriptive statistics

In total, 10,345 cases of suspected severe COVID-19 with sufficient information to include in analysis were reported and recorded by EMAS between March 2nd and May 11th, 2020. Of all cases, 1,123 fell into unusual clusters compared to population, while the remaining 9,222 cases were distributed randomly. The mean (M) and standard deviation (SD) for measures of IMD and AHAHI included in our analysis are
presented in Table 3. The proportion of cases in unusual clusters compared to randomly distributed by sex and RUC categories are presented in Table 4.

4.2.2. Regression analysis

A binary logistic regression analysis was conducted to investigate factors that are associated with cluster membership. Given the highly unequal distribution of cases by binary categories, the probability cutoff was set to 0.6, as distinct from the usual cutoff of 0.5 for randomly distributed binary data (Calabrese, 2014), and the model parameters were set to predict the log-odds of membership in the major category (randomly distributed cases) compared to the minor category (unusual clusters). We also performed bootstrap sampling to account for dependencies between cases in clusters. This analysis did not change the P-values or significant predictors in the regression model.

The results indicate that 12 of 16 variables that input to the Access to Healthy Assets and Hazardous Index (AHAHI), and 4 rural and urban categories are significant predictors of whether cases are distributed randomly or appear in unusual clusters by population (Chi-square = 2028.36, df = 26, P = .00). Age, sex, Index of Multiple Deprivation

---

Table 1

Datasets, measures and sources.

| Dataset* | Measure | Source |
|----------|---------|--------|
| EMAS COVID-19 2020 | Suspected cases of COVID-19 (March 2nd-May 11th), sex, age | East Midlands Ambulance NHS Trust |
| IMD 2019 | IMD Decile | https://hub.arcgis.com/datasets/communities::lower-super-output-area-lsoa-imd-2019-ogb1936 |
| RUC 2011 | Categorical scale 1 (most urban) to 10 (most rural)** | https://hub.arcgis.com/datasets/ons::rural-urban-classification-2011-of-lower-layer-super-output-areas-in-england-and-wales |
| AHAHI 2019 | Retail Environment (distance in km) | Gambling, fast food, pubs/clubs/bars, off license, tobacconists |
| | Health services (distance in km) | GPs, A&E, dentists, pharmacies, leisure |
| | Physical environment (distance in km)
| | Air pollution (levels)

*All data scales at Lower Super Output Area.
**only 8 categories were present in the East Midlands dataset; provisional diagnoses of COVID-19 requiring ambulance attendance in the East Midlands were not recorded in Urban-Major Conurbations, Villages, and Small Town and Fringe areas.
1Passive Green Space includes parks, gardens, golf courses, and allotments. Active Green Space includes sporting areas such as playing fields and tennis courts.
2PM, NO₂ and SO₂ measures are annual μg m⁻³, micrograms per cubic meter of air.

---

Fig. 2. Schematic of database compilation and spatial analysis including data joining, and data display as 2-D and 3-D maps using ArcGIS Pro 2.6.0.
Fig. 3. The geographic location of 13 statistically significant ($P < 0.05$) clusters of COVID-19, identified using a Kulldorff spatial scan statistic. Further details of clusters are given in Table 2.
Fig. 4. Spatial representation of relative risk of suspected cases of COVID-19 in the East Midlands of the UK between March 2nd and May 11th 2020. Taller clusters, and clusters closer to red on the colour gradient reflect greater risk of contracting COVID-19. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Spatial characteristics of unusual clusters of suspected COVID-19 cases presented in Map 1, extracted from SatScan output, including population, number of cases, expected cases, log likelihood, P-value, relative risk, cases per 100,000 population and approximate location of clusters. Population has been determined at the regional postcode scale.

| Cluster | Radius (km) | Population | Number of Cases | Expected Cases | Log Likelihood Ratio | P-Value | Relative Risk | Cases per 100,000 population | Location               |
|---------|-------------|------------|-----------------|----------------|----------------------|---------|---------------|-------------------------------|------------------------|
| 1       | 49.21       | 82,653     | 911             | 652.93         | 48.82                | <0.00   | 1.43          | 1102                          | Nottingham             |
| 2       | 20.67       | 14,120     | 210             | 111.55         | 34.88                | <0.00   | 1.90          | 1487                          | Leicester              |
| 3       | 2.78        | 32,220     | 379             | 254.53         | 27.19                | <0.00   | 1.51          | 1176                          | Derby                  |
| 4       | 33.98       | 18,430     | 233             | 145.59         | 22.53                | <0.00   | 1.61          | 1264                          | West Peak District     |
| 5       | 0.84        | 907        | 31              | 7.16           | 21.61                | <0.00   | 4.34          | 3417                          | East of Rugby          |
| 6       | 1.08        | 4,331      | 77              | 34.22          | 19.76                | <0.00   | 2.26          | 1777                          | East Peak District     |
| 7       | 9.92        | 3,690      | 65              | 29.15          | 16.34                | <0.00   | 2.24          | 1761                          | West Grimsby           |
| 8       | 42.15       | 87,897     | 836             | 694.36         | 14.60                | 0.00    | 1.22          | 951                           | West of Skegness       |
| 9       | 11.16       | 9,235      | 121             | 72.96          | 13.29                | 0.00    | 1.67          | 1310                          | Southwest of Leicester |
| 10      | 1.17        | 1,543      | 32              | 12.19          | 11.10                | 0.00    | 2.63          | 2073                          | Southwest of Derby     |
| 11      | 1.3         | 12,443     | 148             | 98.3           | 10.98                | 0.00    | 1.51          | 1369                          | Northampton            |
| 12      | 4.99        | 5,285      | 74              | 41.75          | 10.15                | 0.01    | 1.78          | 1400                          | East Grimsby           |
| 13      | 13.33       | 6,029      | 81              | 47.63          | 9.70                 | 0.02    | 1.71          | 1343                          | North of Chesterfield  |
appear in clusters, giving an overall percentage correct prediction rate of 91.4%. Rural town and fringe are not significant. The model correctly predicted urban categories (Urban major conurbation, Urban city and town and centres, gambling districts, Green Space and dentists) and 3 rural and categories (RUC).

Table 4 displays the binary logistic regression results for the independent variables that were found to be associated with cluster membership. Compared to randomly distributed cases, cases in clusters are more likely to be located closer to pubs/clubs/boozers, Blue Space, off licenses, Passive Green Space, as well as in areas with higher levels of Nitrogen Dioxide, and in RUC categories ‘urban minor conurbation’, ‘urban city and town in sparse’, ‘rural town and fringe’, and ‘rural village and dispersed’. Randomly distributed cases that do not occur in clusters are located closer to tobacconists, GP practices, A&E hospitals, and pharmacies, as well as in areas with higher levels of particulate matter and Sulphur Dioxide.

Table 3
Descriptive statistics for measures of Index of Multiple Deprivation (IMD), Access to Healthy Assets and Hazardous Index (AHAH) and age for cases of severe COVID-19 in unusual clusters (M_IN, SD_IN) compared to cases randomly distributed outside clusters (M_OUT, SD_OUT). Measures of IMD are decile values. Measures of AHAH include four domains: distance (km) from retail environments, health services, physical environments, and air quality.

| Domain          | Factor                  | M_IN  | SD_IN  | M_Out | SD_Out |
|-----------------|-------------------------|-------|--------|-------|--------|
| Retail environments | Gambling              | 2.02  | 2.63   | 2.50  | 2.87   |
|                  | Fast food               | 1.85  | 2.65   | 2.18  | 2.48   |
|                  | Pubs/clubs/bar          | 1.40  | 1.91   | 1.87  | 2.22   |
|                  | Off License             | 4.00  | 5.50   | 4.87  | 6.62   |
|                  | Tobaccoists             | 3.26  | 3.861  | 3.63  | 3.41   |
| Health services  | GP's                    | 1.44  | 1.47   | 1.67  | 1.55   |
|                  | A&E                     | 16.76 | 16.40  | 15.52 | 10.30  |
|                  | Dentists                | 1.65  | 1.97   | 2.10  | 2.28   |
|                  | Pharmacies              | 1.21  | 1.50   | 1.39  | 1.62   |
|                  | Leisure                 | 3.12  | 3.95   | 3.95  | 4.31    |
| Physical environment | Green Space (passive)   | 0.34  | 0.25   | 0.36  | 0.48   |
|                  | Green Space (active)     | 0.54  | 0.59   | 0.58  | 0.55   |
|                  | Blue Space              | 2.24  | 1.79   | 2.57  | 2.13   |
|                  | Nitrogen Dioxide        | 12.59 | 2.31   | 11.77 | 1.81   |
| Air pollution    | Particulate Matter      | 13.64 | 1.60   | 14.30 | 0.80   |
|                  | Sulphur Dioxide         | 1.40  | 0.29   | 1.24  | 0.23   |
|                  | IMD decile             | 4.38  | 2.84   | 5.04  | 2.87    |

Table 4
Proportion of cases in unusual clusters (IN( %)) compared to randomly distributed cases outside clusters (OUT( %)) by sex and Rural Urban Classification Categories (RUC).

| RUC              | Sex | In (%) | Out (%) |
|------------------|-----|--------|---------|
| Urban major conurbation | Female | <1     | <1      |
| Urban minor conurbation     |       | 34.9   | 16.5    |
| Urban city and town   | Male | 49.1   | 62      |
| Urban city and town in sparse setting | <1 | 0.6    |
| Rural town and fringe |       | 10.3   | 12.8    |
| Rural town and fringe in sparse setting | <1 | <1     |
| Rural village and dispersed | | 4.3    | 7.3     |
| Rural village and dispersed in sparse setting | 1.3 | 0.2    |
| Sex               |     | 54     | 53      |
| Female            |     | 45     | 46      |
| Male              |     | <1     | <1      |

4.2.3. Index of Multiple deprivation ANOVA
Regression analysis revealed that IMD deciles was not a significant predictor of whether provisionally diagnosed severe COVID-19 cases were distributed randomly or occurred in unusual clusters. ANOVA was also computed to identify whether IMD scores varied between each individual cluster, and areas with randomly distributed cases. There was a significant difference for IMD decile scores between Cluster 1 (M = 2.93, SD = 2), Cluster 2 (M = 2.42, SD = 1.21), Cluster 3 (M = 3.22, SD = 2), Cluster 4 (M = 6, SD = 2.33), Cluster 5 (M = 7.4, SD = 2), Cluster 6 (M = 9, SD = 0.6), Cluster 7 (M = 3.9, SD = 2.83), Cluster 8 (M = 2.86, SD = 0.83), Cluster 9 (M = 8.06, SD = 1.86), Cluster 10 (M = 7.84, SD = 1.9), Cluster 11 (M = 2.49, SD = 1.04), Cluster 12 (M = 4.68, SD = 1.18), Cluster 13 (M = 5.79, SD = 2.48), and cases that are evenly distributed (M = 5.04, SD = 2.88), F(13, 10331) = 40.96, p = .00.

4.3. Characteristics of unusually high severe COVID-19 clusters
The statistical analysis presented in section 5.2 considers factors related to all the clusters of suspected severe COVID-19 cases in the East Midlands. The geospatial analysis presented below considers the characteristics of individual clusters. The following series of maps (Figs. 5–9) display the distribution of factors related to retail environments, health services, physical environments (including RUC), air pollution, and IMD. With the exception of RUC, all other factors are represented as deciles values.

Importantly, clusters displayed on the maps reflect the radius within which individual cases of suspected COVID-19 occur. To preserve the anonymity of EMAS patients, we have not displayed the specific location of cases within clusters. Table 6 synthesizes the characteristics of each individual cluster compared to areas with cases that are randomly distributed. Average scores for RUC, IMD and AHAH have also been deidentified . Rather than exact values, Table 6 displays the distribution of factors related to retail environments, health services, physical environments (including RUC), air pollution, and IMD. With the exception of RUC, all other factors are represented as deciles values.

In some instances, the visual characteristics of a cluster may vary from the characteristics reported in Table 6. For example, Fig. 9 displays the distribution of IMD scores within clusters. A

---

3 The asymmetry in predictive accuracy for cases appearing in clusters compared to cases not appearing in clusters is a common phenomenon of highly unequal datasets (Calabrese, 2014) and reflects the true rarity of cases appearing in clusters.

4 Patient anonymity is a requirement of the approved IRAS. It may be possible to triangulate cluster information, such as radius, and specific values, such as IMD, to identify more specific locations. Our approach maintains anonymity and complies with the terms of ethical approval.
Table 5: Binary logistic regression analysis predicting cluster membership. Positive B values indicate an increased likelihood of random distribution and a decreased likelihood of cases occurring in clusters. Negative B values indicate a decreased likelihood of random distribution and an increased likelihood of cases occurring in clusters.

|                                | B    | SE    | Wald  | df | Exp(B)       | 95% CI     |
|--------------------------------|------|-------|-------|----|--------------|------------|
| AHAHII Accessibility to fast food outlets | -0.16| 0.04  | 15.63 | 1  | 0.85**       | 0.78, 0.92 |
| Accessibility to pubs/bars/nightclubs | 0.2  | 0.04  | 21.43 | 1  | 1.22*        | 1.12, 1.33 |
| Accessibility to Blue Space | 0.09 | 0.02  | 14.5  | 1  | 1.1*         | 1.04, 1.14 |
| Accessibility to Off Licenses | 0.02 | 0.01  | 5.18  | 1  | 1.02**       | 1.04       |
| Accessibility to tobacconists | -0.1 | 0.02  | 17.73 | 1  | 0.91*        | 0.87, 0.95 |
| Passive Green Space (within 900 m buffer) | 0.56 | 0.1   | 33.26 | 1  | 1.75*        | 1.45, 2.11 |
| Accessibility to GP practices | -0.14| 0.045 | 10.28 | 1  | 0.87*        | 0.92, 1.2  |
| Accessibility to A&E hospitals | -0.12| 0.005 | 5.18  | 1  | 0.9**        | 0.89, 0.91 |
| Accessibility to pharmacies | -0.11| 0.05  | 3.89  | 1  | 0.9**        | 0.81, 1.01 |
| Level of Nitrogen Dioxide (NO$_2$) | -1.12| 0.05  | 591.83| 1  | 1.75*        | 0.3, 0.4   |
| Level of Particulate Matter (PM10) | 1.51 | 0.06  | 662.64| 1  | 4.53*        | 4.04, 5.9  |
| Level of Sulphur Dioxide (SO$_2$) | 1.98 | 0.28  | 48.26 | 1  | 7.22*        | 4.13, 12.6 |
| RUC Urban minor conurbation | -0.92| 0.09  | 103.03| 1  | 0.4*         | 0.33, 0.48 |
| Urban city and town in a sparse setting | -0.54| 0.17  | 10.23 | 1  | 0.58*        | 0.48, 0.81 |
| Rural town and fringe | -3.01| 1.26  | 5.77  | 1  | 0.05**       | 0.00, 0.58 |
| Rural village and dispersed | -3.9 | 0.7   | 34.76 | 1  | 0.02*        | 0.00, 0.07 |

*Statistically significant at $P < .01$.
**Statistically significant at $P < .05$.

Fig. 5. Maps depicting distance (km) from ‘harmful’ retail environments derived from the Access to Healthy Assets and Hazardous Index (AHAH) that are associated with cluster membership, including off licenses, pubs/ bar/clubs, fast food outlets and tobacconists. The green spectrum indicates areas that are further away and the red spectrum indicates areas that are closer. The 13 clusters of high numbers of suspected COVID-19 cases (identified using a Kulldorff spatial scan statistic) are superimposed as black circles and numbered consistent with Table 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. 6. Maps depicting distance (km) from ‘healthy’ services derived from the Access to Healthy Assets and Hazardous Index (AHAHI) that are associated with cluster membership, including A&E hospitals, GPs, and pharmacies. The green spectrum indicates areas that are closer and the red spectrum indicates areas that are further away. The 13 clusters of high numbers of suspected COVID-19 cases (identified using a Kulldorff spatial scan statistic) are superimposed as black circles and numbered consistent with Table 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. 7. Maps depicting distance (km) from physical environments derived from the Access to Healthy Assets Hazardous Index (AHAHI) and degree of urbanization/rurality, that are associated with cluster membership, including Green Space (passive), Blue Space, and RUC categories. The 13 clusters of high numbers of suspected COVID-19 cases (identified using a Kulldorff spatial scan statistic) are superimposed as white circles and numbered consistent with Table 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. 8. Maps depicting the level of pollutants derived from the Access to Healthy Assets Hazardous Index (AHAHI) that are associated with cluster membership, including Particulate Matter (PM10), Sulphur Dioxide (SO\textsubscript{2}) and Nitrogen Dioxide (NO\textsubscript{2}). The green spectrum indicates lower levels of pollutants and the red spectrum indicates higher levels. The 13 clusters of high numbers of suspected COVID-19 cases (identified using a Kulldorff spatial scan statistic) are superimposed as black circles and numbered consistent with Table 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. 9. Map of Index of Multiple Deprivation (IMD) distribution and unusual clusters of suspected COVID-19 cases. The green spectrum indicates greater affluence and the red spectrum indicates greater deprivation. The 13 clusters of high numbers of suspected COVID-19 cases (identified using a Kulldorff spatial scan statistic) are superimposed as black circles and numbered consistent with Table 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Table 6
Characteristic of individual clusters of unusually high suspected cases of COVID-19 compared to randomly distributed cases, including the proportion of cases in urban (U) and rural (R) areas (RUC), Index of Multiple Deprivation (IMD) Decile, and Access to Healthy Assets and Hazardous Index (AHAHI) indicators (average distance (km) from retail environments, health services, and physical environments, as well as average level of air pollution).

| Cluster | IMD Decile | RUC (%)* | Retail Environment | Health Services | Physical Environment | Air Pollution |
|---------|------------|----------|--------------------|----------------|---------------------|--------------|
|         | Central    | Peripheral | Entirely or higher | Entirely or higher | Entirely or higher | Entirely or higher |
|         | Inland    | Coastal | urban % | More affluent | Coastal | urban | % | More affluent | More affluent | More affluent |
|         | More affluent | Coastal | urban | % | More affluent | Coastal | urban | % | More affluent | More affluent |
|         | More deprived | Coastal | urban | % | More deprived | Coastal | urban | % | More deprived | More deprived |

*Scores indicate % of sites in more urban and more rural areas.

- Retail Environment: Fast food (FF), Pubs/bars/clubs (PBC), Off license (OL), Tobacconists (T).
- Health Services: General Practitioners (GPs), A&E Hospitals (A&E), Pharmacies (P).
- Physical Environment: Blue Space (B), Green Space (passive) (G).
- Air Pollution: Particulate Matter 10 (PM), Nitrous Oxide (NO), Sulphur Dioxide (SO).

Table 7
Characteristic of clusters categorized as ‘Inland Urban’, ‘Rural and Mosaic’, and ‘Coastal Urban’, including Index of Multiple Deprivation (IMD), geographic location (inland or coastal), urban and rural dynamics, and Access to Healthy Assets and Hazardous Index (AHAHI).

| Characteristic | Inland Urban | Rural & Mosaic | Coastal Urban |
|---------------|--------------|----------------|---------------|
| IMD Decile    | More deprived | More affluent  | More deprived |
| Geographic location | Inland | Coastal | Inland | Coastal |
| Urban/rural    | Entirely or higher | Entirely or higher | Entirely or higher |
| AHAHI         | Retail    | Closer       | More distant  |
| Health        | Closer    | More distant  | More distant  |
| Physical      | Closer    | More distant  | More distant  |
| Air pollution | Worse     | Better       | Variable*    |

*Skegness cluster has better quality; Grimsby clusters have poorer quality.

Fig. 10. Schematic showing the social-environmental Mesosphere demonstrating the multi-level factors associated with severe illness from COVID-19. The dotted arrow indicates the interaction between socio-economic factors and physical landscape factors within the Mesosphere.

5. Discussion

One year into the COVID-19 pandemic research related to public health, epidemiology, and urban planning has advanced knowledge and understanding about the groups in society that are particularly vulnerable to severe illness from COVID-19. Most research about vulnerable communities and regions considers the association between COVID-19 cases and individual domains, such as deprivation or urbanisation. To our knowledge, the only prior study to use an approach similar to the methodology presented here is Klaghadi et al. (2020) who examined the relationship between confirmed COVID-19 cases in Harris County, Texas and 46 variables across five domains, including access to health services, and environmental exposures. However, the study did not distinguish
between severe illness and asymptomatic or mild cases. Further, their research used aggregate measures collated from a census to estimate the demographic characteristics of patients rather than individual records. Thus, while granular, the approach is limited for identifying communities vulnerable to severe illness and/or death. The trends identified may more accurately reflect transmission, rather than underlying susceptibility.

Severe illness from COVID-19 requiring emergency medical services reflects the intersection of exposure and underlying susceptibility. Vulnerability to severe symptoms is the outcome of complex interactions between individual demographic characteristics and community-scale socio-economic and environmental factors. Our approach, identifying and interrogating unusual clusters of severe illness from COVID-19, and investigating associations between unusual clusters and social and environmental features of landscapes offers a methodology for further supporting vulnerable communities and regions in real-time.

5.1. Identifying unusual clusters and predicting cluster membership

Our spatial analysis revealed 13 statistically significant clusters of suspected COVID-19 cases (Fig. 2) with rates of severe illness ranging from 951 to 3,417 per 100,000 population. Regression analysis identified 13 factors that predict cluster membership. Overall, the predictive accuracy of our regression model is high, with lower specific accuracy for cases occurring in clusters compared to cases occurring outside of clusters. However, the proportion of cases predicted in clusters and outside of clusters are both acceptable and suggest good model fit.

Compared to the reference condition (urban towns and cities), clusters of severe illness are more likely to occur in urban minor conurbations, urban cities and towns in sparse areas, rural towns and fringe areas, and rural villages and dispersed areas. Clusters occur closer to pubs/bars/clubs, off license stores, and Passive Green Space, and further away from fast food, tobacconists, GP practices, A&E hospitals, and pharmacies. The strongest predictors of cluster membership are closer location to Passive Green Space (such as commons and wilderness areas) and high levels of NO₂. Strong associations were also found to PM10 and SO₂ levels.

Some landscape scale trends are consistent with wider literature. For example, NO₂ concentrations are associated with respiratory hospital admissions more generally (Pannullo et al., 2017) as well as COVID-19 related mortality (Kiaghadi et al., 2020; Travaglio et al., 2021). Our results also provide support for other research demonstrating increased vulnerability to disease in urban areas compared to rural areas (Paul et al., 2020), and those further away from health services (Daras et al., 2019). Clusters 1 to 3 are entirely urban, while 10, 12, 13 and 7 are located in areas with higher than average proportions of sites in urban areas compared to non-clusters sites, and to national proportions (Department of Environment, Food, and Rural Affairs, 2020).

Other findings are less consistent with assumptions about landscape teleconnections and health outcomes. The likelihood of cluster membership simultaneously increases at locations closer to off license stores and pubs/bars/clubs, but more distant from fast food venues and tobacconists. The AHAHI, and associated literature, assumes that access to ‘healthy’ assets promotes better health condition while access to ‘hazardous’ assets facilitates poorer health condition (Green et al., 2018). These trends may be related to the nature of amenities and services in rural areas compared to urban areas, and may explain why some retail environments increase likelihood of cluster membership while others do not. Rural towns often contain local pubs while tobacconists and fast food venues are less common. Thus, distance from both retail outlets (more commonly found in more densely populated areas) and health services may reflect poorer access to services more generally, and thus greater vulnerability to illness (Jordan et al., 2004).

Similarly, health literature suggests that closer proximity to green space is associated with better health outcomes (Daras et al., 2018). We found that clusters are more likely to occur closer to, rather than more distant from Passive Green Space, like commons or conservation areas. This may be related to the nature of Passive, as opposed to Active green spaces. Passive Green Space like commons is likely to reflect urban periphery or rurality while Active Green Space like gymnasiuums tend to be located in urban centres. The varied relationships between landscape features and vulnerability in urban compared to more regional areas, deserve more detailed consideration. For example, it is possible that closeness to Passive Green Space reflects social behaviour during the pandemic. In a perspective piece published in this Special Edition (Moore, Hill, Siriwardena, Tanser, & Spaght, 2021) we examine the relationship between landscape features and the implications for COVID-19 exposure and underlying susceptibility in more depth. During extended phases of lockdown parks and arboretums became social hubs that were poorly monitored by local authorities. News reports documented continual violations of social distancing rules in public spaces like beaches and common green areas. Thus, improving the monitoring and enforcement of social distancing in these spaces may be a future avenue for mitigating high rates of severe COVID-19 cases.

Below we suggest that the balance of expected and unexpected associations between unusual clusters and landscape features reflect differences in the individual characteristics of clusters, and the nature of vulnerability between more rural and more urban landscapes.

5.2. Characteristics of individual clusters

The characteristics of clusters vary in two ways. Firstly, the degree of relative risk, and secondly in relation to wider geographic context. In order, clusters with the highest relative risk compared to the mean value were east of Rugby (5), east of the Peak District (6), south west of Derby (10), west of Grimsby (7), Leicester (2), and East Grimsby (12) (Fig. 3). This analysis gives some indication of regions where communities may be particularly vulnerable.

Spatial analysis (Figs. 5–9) revealed several important geographic distinctions between clusters. On this basis we classify clusters in the following categories: inland urban, rural or rural–urban mosaic, and coastal urban (Table 7). Category One, ‘Inland Urban’ including Nottingham, Leicester, Derby, Northampton, and Chesterfield, are predominantly or entirely urban and characterized by closeness to healthy and hazardous services and beneficial physical environments. Clusters located closer to city centres (central Urban Inland: 1, 2, 3, 11) are more deprived, while clusters located in the periphery (peripheral Inland Urban: 13, 10) are more affluent. Category Two, ‘Rural and Mosaic’ clusters in the Peak District, near Rugby, and south west of Leicester, are either entirely rural or display a rural–urban mosaic with a higher proportion of cases in rural areas compared to areas with randomly distributed cases. Rural and Mosaic clusters are characterized by further distance from healthy and hazardous services, closer proximity to beneficial physical environments, and greater affluence. Category Three, ‘Coastal Urban’, includes clusters in predominately urban areas near Skegness and Grimsby. These clusters are characterized by deprivation, and further distance from all services and beneficial physical environments. Importantly, while each category includes clusters with higher levels of NO₂, the sources are likely to vary; traffic contributes to poor air quality in large urban centres, such as Nottingham, while the operation of power plants effects air quality in more regional areas, such
as the coastal Grimsby clusters.

5.3. Understanding vulnerability in the social-environmental landscape

Bioecological models suggest that health outcomes are the cumulative result of complex interactions between individual demographic and biological factors, and the social characteristics of wider environments (Bronfenbrenner, 1979; Eriksson & Hammarström, 2018). Social factors may reflect both exposure to transmission of a contagious virus, (e.g., poorly designed housing estates), and underlying susceptibility related to pre-existing health conditions (Patel et al., 2020). In addition to social dynamics, our analysis included physical characteristics of the built environment that may explain vulnerability to severe symptoms of infectious disease, such as distance from green space (Green et al., 2018).

Our analysis suggests that unusual clusters occur at the nexus of individual susceptibility and exposures in the built environment. However, the dynamics of vulnerability vary between geographic locations. For example, Inland Urban clusters are located closer to all services while Coastal Urban clusters are located further from all services. Except Peripheral Inland Clusters, these regions are more deprived than areas with cases occurring randomly. Thus, the cumulative effect of exposure in high density urban areas and susceptibility associated with deprivation may be more important determinants of vulnerability than distance from specific healthy and hazardous features of built environments.

The characteristics of clusters in more affluent areas, including two with very high relative risk (10, 6) suggests that the dynamics of vulnerability vary markedly from clusters in poorer regions. Firstly, affluent clusters tend to be in more regional locations including urban peripheries and rural areas which are typically occupied by older communities (Office of National Statistics, 2020c). In the U.S., rural communities with high rates of severe COVID-19 symptoms are characterised by aging populations and greater distance from health services (Lakhani et al., 2020). Similar characteristics may explain high relative risk in more peripheral and rural clusters in the East Midlands. With one exception (Skegness), the average age of patients located in the Peripheral Inner Urban clusters is higher than all other clusters, and Rural and Mosaic clusters are located further from health services. These dynamics indicate a ‘rural paradox’; lower risk of transmission, greater susceptibility to severe symptoms, and less access to the medical services required to meet the needs of susceptible communities. Taken together, these observations suggest that the relative contribution of demographic, socio-economic, and environmental factors to vulnerability varies depending on wider geographic location. Factors that influence underlying health susceptibility, like older age and distance from health services, may be stronger predictors of severe illness than socio-economic status in regional locations that are less exposed to transmission risks. In contrast, deprivation and high-density urbanism may outweigh the benefits of closeness to, or distance from, physical features of the built environment. In these cases, it is likely that susceptibility related to deprivation, and exposure related to urbanization, are more powerful drivers of overall vulnerability than access to health services or retail outlets.

Similar to closeness to passive green spaces discussed above, our findings about vulnerability in rural areas suggest some policy responses for future pandemics and phases of lockdown. News reporting during the first national phase of lockdown suggests that the public viewed rural areas as less vulnerable to contagion and mortality related to COVID-19 compared to urban areas (e.g., McCarthy, 2020). Further, rural communities reported the phenomenon of people from urban and peri-urban areas ‘flocking’ to rural regions for recreation during phases of lockdown when only essential travel was legally permitted (Asquith, 2020). In the event of future phases of lockdown, mitigating high rates of severe illness in rural areas with aging populations may require more stringent policing of travel between urban and rural areas.

In summary, factors and processes that explain vulnerability to severe COVID-19 illness and or death are complex and highly location-specific. Bioecological models traditionally focus on the interaction between complex social systems while urban theories emphasise distal associations within physical environmental landscapes. We suggest that social and physical landscape factors rightly belong in a theoretical space akin to Bronfenbrenner’s Mesosystem, which includes processes and interactions that occur within homes, communities, and neighbourhoods (Bronfenbrenner, 1979). Fig. 10 visualises what this Mesosystem might look like.

The granularity of our analysis, facilitated by the high resolution of the data, offers some important insights for supporting the most vulnerable communities in real-time during the early phase of a pandemic when laboratory testing is limited and public policy is informed by cases in the community. In the case of COVID-19, those vulnerable communities include deprived urban neighbourhoods and more affluent regional neighbourhoods.

5.4. Strengths and limitations

There are three limitations of the research. Firstly, big data does not capture individual behaviour; distance from green space and other amenities does not reflect use. Secondly, factors beyond the scope and scale of this research may affect ambulance use. People with close proximity to hospitals with A&E services are more likely to access those services directly rather than calling an ambulance. Similarly, willingness to call an ambulance may vary between communities. Poor health literacy, including ability to recognize symptoms of illness, is often associated with deprivation (Niksic et al., 2015). As a result, it is likely that our data does not represent all severe cases of suspected COVID-19 in the study region. Thus, qualitative community scale research is needed to ground truth the trends and associations reported here. Finally, without data linkage, suspected COVID-19 cases cannot be confirmed. However, the preliminary diagnosis of suspected COVID-19 is based on the assessment of trained medical professionals following the guidelines and algorithms that were widely employed by medical services in the early phases of the pandemic before rapid testing was available. Further, measures taken by ambulance paramedics, including blood oxygen levels (Soltan et al., 2021) and self-reported ODT (Wee et al., 2020; Patterson et al., 2020; Printza & Constantinidis, 2020), have been demonstrated to predict positive cases with a high degree of accuracy. The need for rapid response is paramount. The spatial accuracy of our approach, using a novel routinely collated dataset demonstrates a methodology for identifying vulnerable communities in real-time, as well as understanding the demographic, socio-economic, and environmental characteristics of vulnerability across dynamic geographic landscapes.

6. Conclusions

Vulnerability to severe illness from contagious disease occurs at the intersection of exposure and underlying susceptibility. The effect of biological, social, and environmental risk factors is cumulative. Thus, single characteristics of built environments like deprivation or air pollution do not explain severe symptoms that require emergency medical attention. Our analysis builds on advancements in public health, epidemiology and urban planning by integrating features of the built environment with more traditional bioecological frameworks that tend to focus on complex social interactions.

The analysis of ambulance attendance data for monitoring the progress of the pandemic is a novel approach in the UK, and to our knowledge, has not been used to identify clusters of COVID-19 elsewhere. We acknowledge that analysing suspected COVID-19 cases is an imperfect science. However, we offer some insights that may be of benefit for rapid response as well as longer-term urban planning:

- Joining ambulance data to publicly available big datasets like the IMD and AHAI could identify vulnerable communities in real-time;
• Understanding the social and environmental characteristics of vulnerability may help policy makers to mitigate the impact of a new EID on communities;
• Identifying vulnerable communities in real-time could inform earlier localised lockdowns to mitigate transmission and reduce rates of severe illness. Targeting areas where contagion is likely to result in high rates of hospitalisation would also reduce burden on emergency medical services;
• Opportunities for mitigating transmission also include more effective monitoring and enforcement of social distancing rules in Passive Green Space, including parks, commons and arboretums, as well as for urban-rural travel during lockdown;
• The dynamics of vulnerability vary between urban centres and more peripheral or rural regions, and between more deprived compared to more affluent communities. The opportunities for minimising the impacts of a pandemic include reducing the underlying susceptibility of communities as well as minimising transmission. In part, this involves urban planning to enhance opportunities for health behaviours. Improving the safety of green spaces for cost-free exercise, and increasing infrastructural and financial access to healthy food would promote healthier lifestyles in deprived communities. Further improving access to health services in more affluent and isolated communities may help to mitigate the most severe outcomes of a pandemic. However, in both cases this requires top-down financial investment to encourage healthy retail outlets to locate in deprived neighbourhoods, and health services to locate in low-density neighbourhoods.

At the time of writing, twelve months has elapsed since the declaration of the COVID-19 pandemic and the introduction of national responses to contain transmission. Some approaches have proven more successful than others. Identifying unusual clusters of suspected COVID-19 cases and the factors that predict the location of clusters offers a way forward for the UK to adopt more targeted physical distancing approaches that have been effective for preventing further outbreaks and reducing the economic burden of nation-wide lockdown elsewhere. Unequal health outcomes and severe illness in the UK reflects decades of systemic disadvantage and accumulated vulnerability (Marmot et al., 2020). Addressing underlying susceptibility will require long-term investment in areas including neighbourhood quality, educational attainment, and closing income gaps. Mitigating the impact of future pandemics necessarily involves ‘levelling up’ health across the UK, including between rural and urban spaces, coastal and inland spaces, and deprived and affluent communities.

As a global society, we have entered an indeterminate phase of uncertainty and trial-and-error in combating the pandemic. In the wake of the most immediate threat to human life, policy makers face the challenge of redefining the relationship between societies and their urban habitats. Utilizing big-data to identify hot-spots of vulnerability could be used as a method to inform current mitigation policy, as well as longer-term transitions towards healthier urban landscapes.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.landurbplan.2021.104299.

References

Asquith, J., 2020. People have been flocking to rural areas during COVID-19 lockdowns, viewed 24 June 2021, https://www.forbes.com/sites/jamesasquith/2020/03/29/people-have-been-flocking-to-rural-areas-during-covid-19-lockdowns/?sh=69382a576578.
Balseira, S., Ogilvie, F., Glasswell, A., Anderson, C., Cleary, V., Turbutt, D., & McCloskey, B. (2012). Patterns of early transmission of pandemic influenza in London—link with deprivation. Influenza and Other Respiratory Viruses, 6(3), e35–e41. Barbour, A. (1997). Caring for patients: A critique of the medical model. Stanford University Press.
Beever, S. D., Westmoreland, E., de Jong, M. C., Williams, M. L., & Carslaw, D. C. (2012). Trends in NOx and NO2 emissions from road traffic in Great Britain. Atmospheric Environment, 54, 107–116. Bronfenbrenner, U. (1979). The ecology of human development: Experiences by nature and human design. Cambridge, MA: Harvard University Press. Byrd, T., Stack, M., & Furey, A. (2010). The assessment of the presence and main constituents of particulate matter ten micrometres (PM10) in Irish, rural and urban air. Atmospheric Environment, 44(1), 75–87. Calabrese, R. (2014). Optimal cut-off for rare events and unbalanced classification costs. Journal of Applied Statistics, 41(8), 1678–1692. Campbell, L. K., & Wiesen, A. (2009). Restorative commons: Creating health and well-being through urban landscapes. USDA Forest Service. Cervero, R., & Duncan, M. (2003). American Journal of Public Health, 93(9).
Chowdhury, M., Coleman, M., Mahbuba, A. M., Kok, G., Marmot, M., & Darbhum, D. N. (2009). Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. Malaria Journal, 8(1), 68.
European Commission., 2020. East Midlands - Internal Market, Industry, Entrepreneurship And SMEs - European Commission. eurostat.ec.europa.eu/growth/tools-databases/european-innovation-monitor/base-profile/east-midlands, Accessed 21 September 2020.
Connolly, C., Keil, R., & Ali, S. H. (2021). Extended urbanisation and the spatialities of infectious disease: Demographic change, infrastructure and governance. Urban Studies, 58(2), 245–263. Cuadros, D. F., Tomita, A., Vandommel, A., Slotow, R., Burns, J. K., & Tanner, F. (2019). Spatial structure of depression in South Africa: A longitudinal panel survey of a nationally representative household sample. Scientific Reports, 9, 1–10. Daras, K., Davies, A., Green, M. A., & Singleton, A. (2018). In Consumer Data Research (pp. 166–177). UCL Press. https://doi.org/10.2307/j.ctvqhsn6.15. Daras, Konstantinos., Green, Mark A., Davies, Alec., Singleton, Alex., Barr, Benjamin., 2019, Access to Healthy Assets and Hazards (AHAI). https://doi.org/10.6084/m9.figshare.8295842.v1, viewed 5 August 2020.
Daras, K., Alexiou, A., Rose, T.C., Buchan, I., Taylor-Robinson, D, and Barr, B., 2021. How does vulnerability to COVID-19 vary between communities in England? Developing a Small Area Vulnerability Index (SAVI). J Epidemic Environment, Department of Environment, Food, and Rural Affairs., 2020. Rural Population 2014/2015. www.gov.uk/government/publications/rural-population-and-migration/rural-population-201415#:~:text=ht...2015%201%20ראשי%20גופי%20 in 20%20客家%20 אנשים%20andr%20/population,viewed 6 September 2020.
De Vries, S., Van Dillen, S. M., Groenewegen, P. P., & Spreeuwers, P. (2013). Social Science & Medicine.
Durand, C. P., Andalib, M., Dunton, G. F., Wolch, J., & Pentz, M. A. (2011). A systematic review of built environment factors related to physical activity and obesity risk implications for smart growth urban planning. Obesity Reviews, 12(S5), e173–e182.
The Health Foundation. 2020. NHS Test and Trace: the journey so far. Published 23 September 2020, https://www.health.org.uk/publications/long-reads/nhs-test-and-trace-the-journey-so-far, accessed 28 February 2021.
Erikson, M., & Hammarström, A. (2018). Social Theory & Health.
Frank, L. D., & Wall, B. (2021). Treating two pandemics for the price of one: Chronic and infectious disease impacts of the built and natural environment. Sustainable Cities and Society, 73, 103089. https://doi.org/10.1016/j.scs.2021.103089.
Gomez, E. J., Johnson, B. A., Selva, M., & Sallis, J. P. (2004). Violent crime and outdoor physical activity among inner-city youth. Preventive Medicine, 39(5), 876–881. Goytia, C. Y., Rocco, L., & Simpson, M. (1972). The contribution of urbanization to non- communicable diseases: Evidence from 173 countries from 1980 to 2008. Economics & Human Biology, 26, 151–163.
Green, M. A., Daras, K., Davies, A., Barr, B., & Singleton, A. (2018). Developing an openly accessible multi-dimensional and small area index of Access to Healthy Assets and Hazards’ for Great Britain, 2016. Health & Place, 54, 11–19.
Guilmoto, C. Z. (2020). COVID-19 death rates by age and sex and the resulting mortality vulnerability of countries and regions in the world. Med Rev.
Hartig, T., Antell-Burt, T., Bergsten, Z., Amcoff, J., Mitchell, R., & Feng, X. (2020). Associations between greenspace and mortality vary across contexts of community change: A longitudinal ecological study. Journal of Epidemiology and Community Health, 74(6), 534–540.
Hu, H., Nigmatulina, K., & Eckhoff, P. (2013). The scaling of contact rates with population density for the infectious disease models. Mathematical Biosciences, 244 (2), 125–134.
Jordan, H., Roderick, P., Martin, D., & Barnett, S. (2004). Distance, rurality and the need for primary care: Access to health services in South West England. International Journal of Health Geographics, 3(1), 21.
Kasda, E., Robson, C., Saunders, J., Addady, A., Ford, C., Sinha, N., … Paine, L. (2020). Using event reports in real-time to identify and mitigate patient safety concerns during the COVID-19 pandemic. Journal of Patient Safety and Risk Management, 25 (4), 156–158.
Khalatbari-Soltani, S., Cumming, R. C., Delpierre, C., & Kelly-Irving, M. (2020). Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. Journal of Epidemiology and Community Health, 74(8), 620–623.
Khiagahi, A., Rifai, H.S. and Liaw, W., 2020. Assessing COVID-19 risk, vulnerability and infection prevalence in communities in Xinjiang, China, 15(10), p.e024116. Kim, G.U., Kim, M.J., Ra, S.H., Lee, B., Bar, S., Jung, J. and Kim, S.H., 2020. Clinical characteristics of asymptomatic and symptomatic patients with mild COVID-19. Clinical Microbiology and Infection, 26(7), pp.948-e11.
Kontopantelis, E., Mamas, M., Milfield, J., Anslow, M., & Doran, T. (2021). Excess mortality in England and Wales during the first wave of the COVID-19 pandemic. Journal of Epidemiol Community Health, 75(3), 213–223.
Paul, R., Arif, A. A., Adeyemi, O., Ghosh, S., & Han, D. (2020). Progression of COVID-19 from urban to rural areas in the United States: A spatiotemporal analysis of prevalence rates. The Journal of Rural Health, 36(4), 591-601.

Peric, S., & Slunig, T. M. (2020). Diabetes and COVID-19. Wiener Klinische Wochenschrift, 132(13-14), 356-361.

Printza, A., & Constantinidis, J. (2020). The role of self-reported small and taste disorders in suspected COVID-19. European Archives of Otology-Laryngology.

Public Health England, 2018. East Midlands Profile A Summary Of Public Health In The Region 2018. PHE publications, p.7.

Repetto, R. L., Taylor, S. E., & Seaman, T. E. (2002). Risky families: Family social environments and the mental and physical health of offspring. Psychological Bulletin, 128(2), 330.

Romero Starke, K., Peterite-Haack, G., Schubert, M., Kampf, D., Schleibner, A., Hegewald, J., & Seidler, A. (2020). The age-related risk of severe outcomes due to COVID-19 Infection: A rapid review, meta-analysis, and meta-regression. International Journal of Environmental Research and Public Health, 17(16), 5974.

Rusthon, S. P., Goodfellow, M., O’Donnell, A. G., & Magee, J. G. (2007). The epidemiology of atypical mycobacterial diseases in northern England: A space-time clustering and Generalized Linear Modelling approach. Epidemiology & Infection, 135(5), 765-774.

Seo, S., Choi, S., Kim, K., Kim, S. M., & Park, S. M. (2019). Association between urban green space and the risk of cardiovascular disease: A longitudinal study in seven Korean metropolitan areas. Environment international, 125, 51-57.

Seto, K. C., Remberg, A., Boone, C. G., Fragkias, M., Haase, D., Langanke, T., … Simon, D. (2012). Urban land telecommunications and sustainability. Proceedings of the National Academy of Sciences, 109(20), 7687-7692.

Sharifi, A., & Khavarian-Garmisi, A. R. (2020). The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. Science of the Total Environment.

Shen, Y. S., & Lung, S. C. (2016). Can green structure reduce the mortality of cardiovascular diseases? Science of The Total Environment, 566-567, 1159-1167.

Smith, C. M., Lessells, R., Grant, A. D., Herbst, K., & Tanser, F. (2018). Spatial clustering of drug-resistant tuberculosis in Hlabisa subdistrict, KwaZulu-Natal, 2011–2015. The International Journal of Tuberculosis and Lung Disease, 22(3), 287-293.

Soltan, A. A., Kouchaki, S., Zha, T., Kiyasseh, D., Taylor, T., Hussain, Z. B., … Clifton, D. A. (2021). Rapid triage for COVID-19 using routine clinical data for patients attending hospital: Development and prospective validation of an artificial intelligence screening test. The Lancet Digital Health, 3(2), e78-e87.

Steinberg, E., Wright, E., & Kushner, B. (2020). In young adults with COVID-19, obesity is associated with adverse outcomes. Western Journal of Emergency Medicine, 21(4), 1895-1901.

Stier, A., Berman, M., & Bettencourt, L. (2020). COVID-19 attack rate increases with city size. Mansueto Institute for Urban Innovation Research Paper Forthcoming.

Sze, S., Pan, D., Nevill, C.R., Martin, C.A., Nazareth, J., Minhas, J.S., Divall, P., Khunti, K., Abrams, K.R. and Nellums, L.B. (2020). Ethnicity and clinical outcomes in COVID-19: a systematic Review and Meta-analysis. EClinicalMedicine, p.100630.

Tanser, F., Barnighausen, T., Dobra, A., & Sartorius, B. (2018). Identifying ‘corridors of transmission’ in a severely affected rural African population: A case for a shift toward targeted prevention strategies. International Journal of Epidemiology, 47(2), 537-549.

Tomita, A., Cuadros, D. F., Mbabuadi, T., Sartorius, B., Ncama, B. P., Dangour, A. D., … Burns, J. K. (2020). Spatial clustering of food insecurity and its association with COVID-19: A geospatial analysis of nationally representative South African data, 2008–2015. Scientific Reports, 10(1), https://doi.org/10.1038/s41598-020-70647-1

Travaglio, M., Yu, Y., Popovic, R., Selly, L., Leal, N. S., & Martins, L. M. (2021). Links between air pollution and COVID-19 in England. Environmental Pollution, 268, 115859, https://doi.org/10.1016/j.envpol.2021.115859.

Viegi, G., Maio, S., Pistelli, F., Baldacci, S., & Carrozzu, L. (2006). Epidemiology of chronic obstructive pulmonary disease: Health effects of air pollution. Respirology, 11(5), 523-532.

Woo, I. E., Chan, Y. F. Z., Teo, N. W. Y., Cheong, B. P. Z., Thiem, S. Y., Wong, H. M., … Tan, T. T. (2020). The role of self-reported olfactory and gustatory dysfunction as a screening criterion for suspected COVID-19. European Archives of Otorhinolaryngology, 277(8), 2389-2396.

Williams, A. M. (2016). Therapeutic landscapes. International encyclopedia of geography: People, the earth, environment and technology. John Wiley & Sons, World Health Organization., 2020. Media Statement: Knowing the risks of COVID-19. https://www.who.int/intهما/news/detail/08-03-2020-knowing-the-risk-for-covid-19--trans-Most%20people%20are%20less%20at%20risk. Viewed, 1 October 2020.

Wu, T., Perrings, C., Kinzig, A., Collins, J. P., Minter, B. A., & Daszak, P. (2017). Economic growth, urbanization, globalization, and the risks of emerging infectious diseases in China: A review. Ambo, 46(1), 18-29.